

Machine-learning methods for weak lensing analysis of the ESA Euclid sky survey

THÈSE N° 8276 (2018)

PRÉSENTÉE LE 2 MARS 2018
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE D'ASTROPHYSIQUE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Thibault Adrien KUNTZER

acceptée sur proposition du jury:

Prof. F. Mila, président du jury
Dr F. Courbin, directeur de thèse
Prof. J.-L. Starck, rapporteur
Prof. H. Hoekstra, rapporteur
Prof. M. Jaggi, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

A mon grand-papa Raymond.

Acknowledgements

This work has been made possible with the support of many people. The first person I would like to praise for his unwavering support is my supervisor Frédéric Courbin. Frédéric found the time to listen to many of my half-cooked ideas, helped me to develop them and guided me on the winding path that lead me here. I collaborated intensively with Malte Tewes during these past few years, I would like to thank him for challenging me to always aim high. Many others in the EPFL Laboratory of astrophysics (Lastro) staff contributed to my successes: Olivier Geneva and Yves Revaz whom I bugged too many times, Jean-Paul Kneib, Carol Maury and Sophie Oblette for creating the right environment, and Georges Meylan for welcoming me into his staff.

I owe much to those with whom I shared so many good moments. Discussions are key to a sound approach to solving problems. Many times our serious and focussed exchanges strayed from science to goofy and sometimes weird subjects. I feel privileged to have argued with Vivien Bonvin, Markus Rexroth, Rémy Joseph, Olga Tihhonova, Aurélien Wyttenbach, Sébastien Peretti, Cyril Gregory, Claude Guidi, and many more. I would also thank Romain Meyer for his hard work in helping me explore some of my wildest ideas.

From time to time, it is feels good to lay down one's burden, – only to pick up another. I was involved in many different projects, sometimes very weakly linked to the main subject of this thesis. I thank Andrea Fortier, Christoph Schaefer, Claude Nicollier, Camille Pirat, Christophe Paccolat, and Ludovic Monnerat for allowing me to channel my energy into different fields.

Je veux remercier les nombreux amis et membres de famille qui m'ont soutenu, sans toujours comprendre mes explications, certes un peu alambiquées. Je tiens à souligner le soutien sans limite, la gentillesse et la compréhension de Tiffany sans laquelle cette thèse aurait été autrement plus pénible et ces quatre ans ternes.

Thibault

Abstract

A clear picture has emerged from the last three decades of research: our Universe is expanding at an accelerated rate. The cause of this expansion remains elusive, but in essence acts as a repulsive force. This so-called dark energy represents about 69% of the energy content in the Universe. A further 26% of the energy is contained in dark matter, a form of matter that is invisible electromagnetically. Understanding the nature of these two major components of the Universe is at the top of the list of unsolved problems. To unveil answers, ambitious experiments are devised to survey an ever larger and deeper fraction of the sky. One such project is the European Space Agency (ESA) telescope *Euclid*, which will probe dark matter and infer desperately needed information about dark energy.

Because light bundles follow null geodesics, their trajectories are affected by the mass distribution along the line of sight, which includes dark matter. This is gravitational lensing. In the vast majority of cases, deformations of the source objects are weak, and profiles are slightly sheared. The nature of the dark components can be fathomed by measuring the shear over a large fraction of the sky. The shear can be recovered by a statistical analysis of a large number of objects.

In this thesis, we take on the development of the necessary tools to measure the shear. Shear measurement techniques have been developed and improved for more than two decades. Their performance, however, do not meet the unprecedented requirements imposed by future surveys. Requirements trickle down from the targeted determination of the cosmological parameters. We aim at preparing novel and innovative methods. These methods are tested against the *Euclid* requirements. Contributions can be classified into two major themes. A key step in the processing of weak gravitational lensing data is the correction of image deformations generated by the instrument itself. This point spread function (PSF) correction is the first theme. The second is the shear measurement itself, and in particular, producing accurate measurements.

We explore machine-learning methods, and notably artificial neural networks. These methods are, for the most part, data-driven. Schemes must first be trained against a representative sample of data. Crafting optimal training sets and choosing the method parameters can be crucial for the performance. We dedicate an important fraction of this dissertation to describing simulations behind the datasets and motivating our parameter choices.

We propose schemes to build a clean selection of stars and model the PSF to the *Euclid* requirements in the first part of this thesis. Shear measurements are notoriously biased because of their small size and their low intensity. We introduce an approach that produces

Acknowledgements

unbiased estimates of shear. This is achieved by processing data from any shape measurement technique with artificial neural networks, and predicting corrected estimates of the shape of the galaxies, or directly the shear. We demonstrate that simple networks with simple trainings are sufficient to reach the *Euclid* requirements on shear measurements.

Key words: Space-borne observatory *Euclid* – Weak gravitational lensing – Shear measurement – Point spread function (PSF) – Artificial neural networks (ANNs).

Résumé

Un constat sans appel s'est dégagé des trois dernières décennies de recherche : notre Univers est en expansion accélérée. La cause de cette accélération reste à ce jour mystérieuse, mais se traduit par une force répulsive agissant à l'échelle cosmologique. Appelée énergie sombre, ce composant est majoritaire dans le bilan énergétique de l'Univers : 69% du total. Un autre composant, 26%, est la matière sombre, qui n'interagit pas électromagnétiquement avec la matière baryonique. Faire lumière sur ces composantes est de première priorité. Pour y parvenir, des expériences ambitieuses sont imaginées. Un de ces projets est l'observatoire *Euclid*, de l'agence spatiale européenne (ESA), qui va explorer la matière noire et permettre une meilleure compréhension de l'énergie sombre.

Les faisceaux lumineux suivent des géodésiques nulles. Leur trajectoire sera impactée par la distribution de masse, y compris de masse sombre. Cet effet est appelé lentille gravitationnelle. Dans la plus grande partie des cas, ces déformations de l'image des objets sources sont faibles. Les déformations se manifestent par un léger cisaillement des profils. La nature des composantes sombres peut être déduite en observant le cisaillement sur une large fraction du ciel. Le cisaillement peut être calculé statistiquement des formes d'un grand nombre d'objets. Cette thèse s'intéresse au développement des outils nécessaires à cette mesure. Des méthodes de mesure ont été développées et améliorées depuis plus de vingt ans, mais leur performance n'atteint pas les exigences imposées par les prochaines grandes expériences. Ces exigences découlent des ambitions sur la détermination des paramètres cosmologiques. Notre objectif est d'implémenter de nouvelles méthodes qui sont testées par rapport aux exigences d'*Euclid*. Les contributions sont classifiées en deux thèmes. Un processus clef dans le traitement des données de lentille faible est la correction de la fonction d'étalement du point (PSF). Cette correction est le premier thème de recherche. Le second est la mesure du cisaillement et, en particulier, comment produire des mesures exactes.

Nous utilisons des techniques d'apprentissage automatique. Ces techniques, pour la plupart, apprennent des modèles depuis les données elles-mêmes. Les entraînements se font depuis un échantillon représentatif. La préparation de ces données d'apprentissage et le choix des paramètres peuvent se révéler cruciaux. Une large place sera faite à la discussion des simulations et des paramètres.

Nous établissons des techniques pour modéliser la PSF en se basant sur une sélection propre d'étoiles selon les critères euclidiens. Avec un modèle de PSF, on peut procéder à la mesure de la forme des profils de galaxies. Ces mesures sont connues pour être particulièrement biaisées du fait de la petite taille et faible intensité des galaxies observées. Nous introduisons

Acknowledgements

une approche qui engendre des estimations exactes du cisaillement. Ce résultat est atteint par le traitement de données venant de quelque algorithme de mesure de cisaillement par des réseaux de neurones artificiels. Ces réseaux prédisent des estimations corrigées de la forme, ou directement du cisaillement. Nous démontrons que des réseaux très simples, avec des entraînements simples, sont suffisants.

Mots-clefs : Observatoire spatial *Euclid* – Effet de lentille gravitationnel faible – Mesure de cisaillement – fonction d'étalement du point (PSF) – réseaux de neurones artificiels.

Contents

Acknowledgements	i
Abstract	iii
Résumé	v
List of acronyms and abbreviations	xi
List of figures	xv
List of tables	xvii
1 Introduction	1
2 Cosmology and gravitational lensing	3
2.1 Foundations of cosmology	4
2.1.1 Einstein field equations of General Relativity	4
2.1.2 The cosmological principle	4
2.1.3 The Friedmann-Lemaître-Robertson-Walker metric	6
2.1.4 Cosmological redshift	6
2.1.5 The Hubble constant and the deceleration parameter	7
2.1.6 The Friedmann-Lemaître equations	8
2.1.7 The content of the universe	9
2.2 Observational cosmology	12
2.2.1 Type Ia Supernovæ	12
2.2.2 The image of last scattering: Cosmic Microwave Background	13
2.2.3 Large-scale structures	14
2.2.4 A cosmological concordance model	15
2.2.5 Dark matter	16
2.2.6 Baryonic acoustic oscillations	17
2.2.7 Going beyond Λ CDM to explain dark energy?	18
2.3 Gravitational lensing effect	19
2.3.1 Light travelling in a gravitational field	19
2.3.2 The eclipse experiments	21
2.3.3 The dawn of gravitational lensing	21
	vii

2.3.4	The lens equation	22
2.3.5	Magnifying and distorting the source image	25
2.4	Weak gravitational lensing and its applications	26
2.4.1	The weak regime	26
2.4.2	The deflection angle for weak lensing	28
2.4.3	Reduced shear estimator from galaxy shapes	29
2.4.4	Mass map reconstruction	30
2.4.5	Decomposition in E- and B-modes	31
2.4.6	Lensing by large-scale structures	31
2.4.7	Contamination of the lensing signal by intrinsic alignment	35
2.4.8	Cluster lensing	36
2.4.9	Galaxy-galaxy lensing	36
2.4.10	Lensing by voids and filaments	37
2.4.11	CMB lensing	38
2.4.12	Flexion: powerful but infeasible?	38
2.5	Other gravitational lensing regimes	39
2.5.1	Strong lensing	39
2.5.2	Microlensing	41
3	Surveys and data analysis for weak gravitational lensing	43
3.1	The forward process	43
3.2	The inverse process	45
3.2.1	Estimators of shapes	45
3.2.2	Adversarial effects	48
3.2.3	Computational challenges	52
3.3	Weak lensing statistics	53
3.3.1	Second-order statistics	53
3.3.2	Higher order statistics	55
3.4	Weak lensing analysis pipeline	56
3.5	Weak lensing surveys	57
3.5.1	Challenges	57
3.5.2	Ground-based surveys	60
3.5.3	Space-based surveys	61
3.6	The <i>Euclid</i> mission	62
3.6.1	Scientific goals	62
3.6.2	Implementation of <i>Euclid</i>	64
3.6.3	Weak lensing survey	66
4	Machine-learning techniques	67
4.1	The turbulent rise of machine learning	68
4.1.1	Gradual evolution of the machine-learning approach	68
4.1.2	The early days	69
4.1.3	Modern machine learning	70

4.1.4	A golden age in the making?	71
4.2	A feel for machine learning	71
4.2.1	Least-square fitting	71
4.2.2	The curse of dimensionality	72
4.2.3	Principal Component Analysis	73
4.3	Artificial neural networks	77
4.3.1	Fundamentals	77
4.3.2	Optimisation of model parameters	79
4.3.3	Building an effective ANN	82
4.3.4	Convolutional neural networks	85
4.4	Random forests	87
4.4.1	Decision trees	87
4.4.2	Growing forests	87
4.5	ANNs for dimensionality reduction: auto-encoders	88
4.5.1	Learning the identity function	88
4.5.2	Application to MNIST	90
4.5.3	State-of-the-art improvements	93
4.6	Evaluating performances	95
4.6.1	Classification metrics	96
4.6.2	Regression metrics	97
5	Weak lensing measurements	99
5.1	The effect of unresolved binaries on PSF determination	100
5.1.1	Multiple stellar systems	100
5.1.2	Unresolved binaries-induced bias in PSF determination	102
5.1.3	Unresolved-binary bias in <i>Euclid</i> observations	107
5.1.4	<i>Paper I</i> : Evaluating the effect of stellar multiplicity on the PSF of space-based weak lensing surveys	109
5.1.5	Strategies for reducing the effect of unresolved binary bias	120
5.1.6	<i>Paper II</i> : Detecting unresolved binary stars in <i>Euclid</i> VIS images	130
5.2	Stellar spectral classification based on VIS images	141
5.2.1	Spectral type determination using single-band imaging	141
5.2.2	<i>Paper III</i> : Stellar classification from single-band imaging using machine learning	147
5.2.3	Predicting a colour-magnitude diagram from HST single-band data	159
5.3	PSF measurement and reconstruction	167
5.3.1	<i>Euclid</i> typical PSF and stellar population	167
5.3.2	Description of the approach	169
5.3.3	Quality metric evaluation	171
5.3.4	Auto-encoder architectures and PSF reconstruction	172
5.3.5	Development outlooks	177
5.4	A machine-learning technique for shape measurements	178

Contents

5.4.1	Early phases of the technique	178
5.4.2	Overview of the measurement concept	179
5.4.3	Designing an accurate regression estimator in the presence of feature noise	181
5.4.4	Input features and shape measurement	182
5.4.5	Methods for shear estimations	184
5.4.6	Galaxy image and PSF simulations	187
5.4.7	Summary of the results and lessons learned	189
5.4.8	<i>Paper IV: Weak-lensing shear measurement with machine learning: Teaching artificial neural networks about feature noise</i>	191
5.4.9	Outlooks	218
6	Finding galaxy-scale strong lenses	219
6.1	A data avalanche	219
6.2	Automated search methods	221
6.2.1	Early methods	221
6.2.2	Detecting strong lenses in upcoming surveys	222
6.3	Deep convolutional networks for SL detection	223
6.3.1	Strong lensing finding challenge	223
6.3.2	CNN architectures	224
6.3.3	Performance of the models and outlooks on future development	229
7	Conclusions and outlooks	231
	Bibliography	235
	Curriculum vitae	260

List of acronyms and abbreviations

Λ CDM cosmological constant as dark energy plus cold dark matter model.

2PCF two-point correlation function.

3PCF three-point correlation function.

aCAE asymmetric contractive AE.

ACF auto-correlation function.

AE auto-encoder.

AI artificial intelligence.

ANN artificial neural network.

AUC area under the curve.

BAO baryonic acoustic oscillations.

BFGS Broyden-Fletcher-Goldfarb-Shanno algorithm.

BMG Besançon model of the Galaxy.

CCD charged coupled device.

CFHTLens CFHT Lensing Survey.

CFHTLS Canada-France-Hawaii Telescope Legacy Survey.

CHFT Canada Hawaii France Telescope.

CMB cosmic microwave background.

CMD colour-magnitude diagram.

CNN convolutional neural network.

COSMOGRAIL COSmological MONitoring of GRAVItational Lenses.

COSMOS COSmological Evolution Survey.

List of acronyms and abbreviations

CPU central processing unit.

CTI charge transfer inefficiency.

DES Dark Energy Survey.

ELU exponential linear units.

EOS equation of state.

ESA European Space Agency.

fc fully-connected.

FLRW Friedmann-Lemaître-Robertson-Walker.

FPR false-positive rate.

FWHM full width at half maximum.

GEMS Galaxy Evolution from Morphologies and SEDs.

GGSLC Galaxy-Galaxy Strong Lensing challenge.

GPU graphical processing unit.

GR general relativity.

GREAT GRavitational lEsing Accuracy Testing.

HOLICOW H_0 Lenses in COSMOGRAIL's Wellspring.

HSC Hyper Suprime-Cam survey.

HST *Hubble* Space Telescope.

IA intrinsic alignment.

KIDS Kilo Degree Survey.

Lastro EPFL Laboratory of astrophysics.

LRG luminous red galaxy.

LSST Large Synoptic Survey Telescope.

LUT look-up table.

MACHO MAssive Compact Halo Object.

mas	milli-arcseconds.
MK	Morgan-Keenan.
MLE	maximum likelihood estimator.
MLP	multilayer perceptron.
MNIST	Modified National Institute of Standards and Technology.
MS	main sequence.
MSB	mean square bias.
MSE	mean square error.
MSWB	mean square weighted bias.
NISP-P	Near-Infrared Spectrometer and Photometer near-IR photometer.
NISP-S	NISP slitless spectrograph.
PCA	principal component analysis.
PDF	probability distribution function.
photo-z	photometric redshift.
PSF	point spread function.
ReLU	rectified linear unit.
RF	random forest.
RMB	restricted Boltzmann machine.
RMS	root-mean-square.
ROC	receiver operating characteristic.
S/N	signal-to-noise.
SDSS	Sloane Digital Sky Survey.
SED	spectral energy distribution.
SKA	Square Kilometre Array telescope.
SL	strong lens.
SLACS	Sloan Lens ACS Survey.

List of acronyms and abbreviations

SNC shape noise cancellation.

SNla supernovæ type Ia.

spec-z spectroscopic redshift.

STEP Shear TEsting Programme.

SVD singular value decomposition.

SWG Science Working Group (*Euclid* Consortium).

TPR true-positive rate.

VAE variational auto-encoder.

VIS VISible Image.

WD white dwarf.

WFIRST Wide Field Infrared Space Telescope.

WL weak gravitational lensing or weak lensing.

WMAP3 Wilkinson Microwave Anisotropy Probe.

List of Figures

2.1	Galaxy distribution in the 2dF Galaxy Redshift Survey and SDSS	5
2.2	Dynamical behaviour of FLRW universes	11
2.3	Anisotropies in the CMB	13
2.4	Full sky simulation of the large-scale structures using the PKDGRAV3 code . . .	15
2.5	Sketch of gravitational lens system	23
2.6	Sketch of effect of convergence	25
2.7	Sketch of effect of shear	26
2.8	E-/B-modes decomposition	31
2.9	Weak gravitational lensing by large-scale structures	32
3.1	Forward process for a sheared galaxy and a star	44
3.2	Illustration of the effect of PSFs on galaxy shapes	49
3.3	Typical weak lensing analysis pipeline	56
3.4	Overview of the <i>Euclid</i> spacecraft	65
4.1	PCA applied to a multivariate Gaussian example	74
4.2	PCA transforms applied to MNIST	75
4.3	PCA with two components transforms applied to MNIST	76
4.4	principal component analysis (PCA) eigenvectors for MNIST	77
4.5	Picture of a perceptron neuron	77
4.6	Sketch of a layer of neurons	79
4.7	Example of a CNN architecture	86
4.8	Example of an AE architecture	89
4.9	AE transforms applied to MNIST	91
4.10	AE with two components transforms applied to MNIST	92
4.11	AE filters	93
5.1	Sketch of binary star formation	101
5.2	Effect of null-contrast binaries on the measured profile	103
5.3	Reconstruction errors for a single binary star	104
5.4	Distortions of stacked PSF profiles due to binaries	105
5.5	Numerical model of the binary star density as a function of distance and contrast	120
5.6	Numerical model of the binary star density as a function of distance and mass ratio	121

List of Figures

5.7	Relative errors in the PSF shape for bins of binary systems	122
5.8	Relative errors in the PSF shape for given selection cuts of binary stars	123
5.9	Sketch of the unresolved binary measurement principle	125
5.10	Concept of the iterative PSF field reconstruction and unresolved binaries elimination	126
5.11	Sketch of the ACF technique for unresolved binaries identification	127
5.12	Position configurations for the measurement of binary-induced shape bias	129
5.13	Examples of stellar spectra	142
5.14	Illustration of the chromatic effect on the PSF	143
5.15	Simulated <i>Euclid</i> -stellar images of different spectral types	144
5.16	Concept of the stellar single-band classifier	145
5.17	CMD for the cluster used in the HST star colour predictions	160
5.18	Results of the same cluster experiment	163
5.19	Results of the different clusters but same extinction experiment	164
5.20	Committee trained with ACS GGC images, and tested against NGC 6397	164
5.21	Result of training the stellar classifier with simulation	165
5.22	<i>Euclid</i> VIS-like PSF and <i>GalSim</i> similar-characteristics star	168
5.23	Typical distributions of <i>Euclid</i> stars observed by VIS	168
5.24	Concept of the AE PSF reconstruction algorithm	170
5.25	Concept of the learned-performance metric	171
5.26	aCAE component maps on the PSF field for dataset A	174
5.27	aCAE component maps on the PSF field for dataset D	176
5.28	Conceptual structure of the early MegaLUT scheme	179
5.29	Concept of the current shape measurement approach	180
5.30	Comparison of different S/N estimates	184
5.31	Training sets for networks predicting shear and weight	185
5.32	Training sets for networks predicting ellipticities	187
5.33	<i>Euclid</i> -like galaxies at different magnitude and Sérsic indices	188
5.34	Visualisation of shape measurement networks	189
6.1	Baseline and residual architectures of the CNN lensfinder	225
6.2	Invariant and views architectures of the CNN lensfinder	227
6.3	Dihedral equivariant architecture	228
6.4	ROC curves of the different architectures	229

List of Tables

3.1	<i>Euclid</i> -derived requirements for the knowledge and reconstruction of the PSF .	66
4.1	Confusion matrix for a binary classification and its terminology.	96
5.1	List of clusters used in the stellar classifier application to real data	159
5.2	List of the different datasets used in the PSF reconstruction algorithm study . .	169
5.3	Results of the PSF reconstruction on dataset A	174
5.4	Results on dataset B	175
5.5	Results on dataset C	175

1 Introduction

Cosmology, the study of the Universe as a whole, is an ambitious, but demanding science. It aims at answering questions like “what are the components of the Universe?”, and “how do they evolve with (cosmological) time?” Currently, the knowledge in cosmology is limited: we estimate the age of the Universe to $\sim 13.8 \cdot 10^9$ years, and understand $\sim 5\%$ of its content, which, likely, accounts for (almost) all known particles. The missing 95% is referred to as dark ([Planck Collaboration 2016a](#)). One component is dark matter, a form of matter which manifests itself only through gravitational interactions. Luminous matter alone would not explain rotation curves of galaxies, massive galaxy clusters, or the evolution and state of large-scale structures. Dark matter is invoked to explain this missing matter and makes up $\sim 26\%$ of the Universe. The second component is dark energy, accounting for $\sim 69\%$ of the Universe. This component is responsible for the accelerated cosmic expansion. Models developed over decades have somewhat converged towards a concordance model. Even though there remain non-negligible issues, the concordance model is remarkably successful and economical, as only six parameters are sufficient to describe it fully. The time has now come to stress-test this model, and determine with precision and accuracy its parameters.

The motivation of this thesis are to be found in the very ambitious *Euclid* mission ([Laureijs et al. 2011](#)). This ESA space telescope is a mission dedicated to cosmology. It will fathom the Universe and its history with unprecedented precision and accuracy. To that end, *Euclid* will survey a third of the sky, image about a billion galaxies and spectroscopically determine at least fifty million redshifts. The constraint forecasts on the cosmological parameters show that *Euclid* will represent a significant leap forward in the understanding of the Universe. Such a bold mission comes with major challenges in many domains. This thesis proposes solutions to some of the challenges, in particular to those linked to the gravitational weak lensing probe.

The propagation of light in the Universe is affected by the dark and luminous mass content, because mass locally bends space. Lensing, being sensitive to the total mass, enables the survey and characterisation of dark matter. Imprints of dark energy can also be revealed, unveiling some answers to the many questions about its nature. Weak gravitational lensing is a powerful, but taxing probe (see e.g. [Hoekstra and Jain 2008](#)). Its observational effect is to change the

shapes of distant galaxies by up to $\sim 1\%$. A signal must therefore be extracted statistically from a very large number of shape measurements and corrected for several instrumental effects. Measurements and calibrations must be exquisite, an exercise proved to be extremely difficult. The goals of the work behind this thesis were to explore innovative solutions to address weak lensing challenges in this field.

We contribute novel data processing techniques in two main areas. First, the characterisation and reconstruction of the PSF, which is a major source of systematics in weak gravitational studies. Second, we propose a shape measurement scheme that is accurate, fast and which can virtually correct for any effect. Most of the algorithms that will be presented are based on data-driven approaches and take advantage of a field that knows tremendous success: machine learning, and, in particular, artificial neural network (ANN).

This thesis is structured as followed. In Chap. 2, we present the fundamentals of modern cosmology, review key aspects of observational cosmology and present the field of gravitational lensing, in particular weak lensing. Chapter 3 is dedicated to data analysis and surveys for weak gravitational lensing, including a description of the future *Euclid* mission. We then turn, in Chap. 4, to introducing important machine-learning concepts and algorithms. Chapter 5 is devoted to the presentation of the contributions towards solving the challenges of the *Euclid* mission. We will describe approaches that are, of course, not limited to the *Euclid* survey. In Chap. 6, we introduce a possible solution to another challenge: automatically finding objects affected by the strong gravitational lensing regime. Finally, we summarise the work done and offer a few perspectives in Chap. 7.

2 Cosmology and gravitational lensing

Overview

We live a fascinating time both for cosmology and gravitational lensing. Cosmology underwent multiple and profound revolutions. The earliest cosmological theories were not rooted in science, but in religion. But over the last hundred years or so, cosmology has tremendously changed our view of the Universe based on rigorous mathematical treatments and astronomical observations. In the last two decades, there have been an extraordinary amount of work in cosmology that lead to a remarkably successful model, even if a lot of imperfections and mysteries remain. The most striking of them is that we simply do not know what 95% of the content of the Universe is. Gravitational lensing is a powerful tool to study cosmology. This tool is a direct consequence of the theory of general relativity, emerges at the same time and provides the first observational proof of general relativity. After a long standstill, gravitational lensing came back at the forefront of research. Now, it is considered as one of the most important cosmological probe and prompted numerous surveys, in particular the *Euclid* mission.

In this chapter, we give an incomplete overview of cosmology and gravitational lensing to provide to the reader with the theoretical keys to understand the work achieved for this PhD thesis. We briefly highlight the foundations of modern cosmology in Sect. 2.1. We then turn in Sect. 2.2 to observational cosmology, the concordance model inferred from it and its two biggest enigma: dark matter and dark energy. We continue by exposing the basics of gravitational lensing in Sect. 2.3. We review the powerful weak gravitational lensing technique in Sect. 2.4 and finish by describing the scientific interest of other lensing regimes (Sect. 2.5).

2.1 Foundations of cosmology

In the following section, we will briefly introduce important ideas and results of modern cosmology. We will highlight some of the key results, which are needed to grasp the nature and the scientific importance of gravitational lensing. There are many textbooks on the subject of general relativity and cosmology. Two of them have been particularly useful during the preparation of this manuscript: [Weinberg \(1972\)](#) and [Hobson et al. \(2006\)](#).

2.1.1 Einstein field equations of General Relativity

The gravitational mass of a body and its inertial mass are equivalent. This is because (locally) gravity equals acceleration. This principle is the equivalence principle. Building on this fundamental principle, Einstein introduced his work on general relativity (GR) in November 1915 ([Einstein 1915](#)). One of the important result of GR is that the motion of bodies is determined by the curvature of space-time. The mathematical description of space-time, i.e. the metric, is a key quantity in the derivation of the equations of movement. Another crucial conclusion of GR is the so-called Einstein field equations,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (2.1)$$

where $g_{\mu\nu}$ is the metric, which is a rank-4 tensor; $R_{\mu\nu}$ the Ricci tensor and R is the scalar curvature which are both derived from the metric; G is the (usual) gravitational constant, c the speed of light, $T_{\mu\nu}$ is the energy-momentum tensor and Λ a cosmological constant. The left-hand terms of eq. (2.1) express the geometry of the space-time and the right-hand term describes the energy content. Hence, the geometry of space-time is directly linked to the content in the Universe. Solving this equation is mathematically extremely difficult. In the next section, we explore a fundamental hypothesis to simplify the problem.

2.1.2 The cosmological principle

An important assumption in modern cosmology is the so-called cosmological principle. This principle states that the Universe is isotropic and homogeneous on large scales. The isotropic condition means that the surface number density of galaxies is the same independently of the direction to the observed patch of sky. The homogeneity is more difficult to measure as we cannot observe from different regions of the Universe, i.e. we cannot measure the invariance under translation. To observe homogeneity, one should measure the distances to mass clumps and reconstruct a 3D map. At first glance, homogeneity seems like a bold claim. The stars we observe concentrate in the Milky Way, which belong to the Local Group, a part of a supercluster of galaxies. The goal of the 2dF Galaxy Redshift Survey¹ was to infer a 3D reconstruction of the position of more than 230,000 galaxies ([Colless et al. 2003](#)). Figure 2.1 shows the distribution of the galaxies in 3D at, cosmologically speaking, intermediate to large

¹<http://magnum.anu.edu.au/~TDFgg/>

scales of two datasets: the 2dF and another survey, the Sloane Digital Sky Survey (SDSS)². The sources are distributed in clusters and filaments, not randomly. It is only when the distribution of mass is smeared out on scales of the order of 300 Mpc (1 parsec = $3.086 \cdot 10^{16}$ m) that the Universe appears homogeneous (Wu et al. 1999). There is, at least, another way to verify homogeneity: the cosmic microwave background (CMB). This black-body radiation is extraordinarily isotropic and homogeneous. We will cover the CMB in more details in Sect. 2.2.2. In other words, isotropy translates the fact that the Universe looks the same in every direction and homogeneity that the Universe has the same properties whatever the location. A key consequence follows from the cosmological principle: there is no preferred direction in the Universe.

We can now return to the energy-momentum tensor $T_{\mu\nu}$. Using the cosmological principle, we can assume it to be of the form

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2} \right) U_\mu U_\nu - p g_{\mu\nu}, \quad (2.2)$$

with p the pressure and ρ the energy density of the fluid contained in the Universe and U the 4-velocity of the medium. This is the perfect fluid hypothesis.

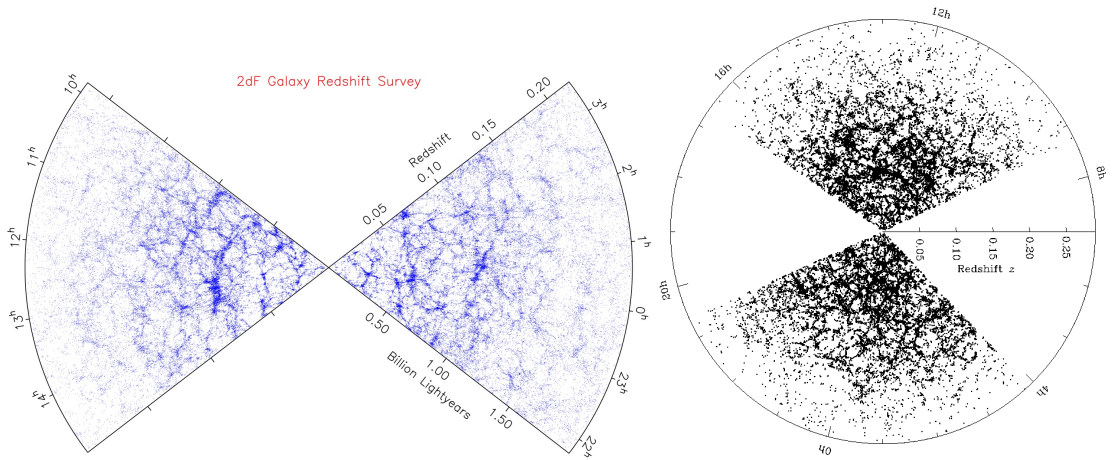


Figure 2.1 – Galaxy distribution in our Universe as per the final data release of the 2dF Galaxy Redshift Survey (*left*, Colless et al. 2003) and the SDSS (*right*, Zehavi et al. 2011). The galaxies (blue and black dots) are not randomly distributed in distance (or redshift – Sect. 2.1.4), but tend to cluster in so-called large-scale structures (Sect. 2.2.3).

²<http://www.sdss.org/>

2.1.3 The Friedmann-Lemaître-Robertson-Walker metric

A mathematical consequence of the cosmological principle is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric ([Friedmann 1922](#); [Lemaître 1927](#); [Walker 1935](#))

$$g_{\mu\nu}dx^\mu dx^\nu = ds^2 = c^2 dt^2 - R^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right]. \quad (2.3)$$

In the above equation, r is the radial coordinate, θ, φ two angular coordinates (making up the three spatial dimensions) and t is the time dimension. The constant k takes the values $+1, 0$ or -1 and describes the geometry of the Universe, respectively closed (positive curvature), flat (zero curvature) and open (negative curvature). Finally, $R(t)$ is the cosmic scale factor, and its normalisation to its current value R_0 : $a(t) = R(t)/R_0$. In closed geometry, the normalised scale factor a can be interpreted as the radius of the universe. The coordinates (r, θ, φ) in the FLRW are comoving, i.e. they are adjusted for an contraction or an expansion of the universe.

Studying the FLRW metric under its different curvatures leads to the re-writing of the metric for a more convenient form

$$ds^2 = c^2 dt^2 - R^2(t) [d\omega^2 + f_k^2(\omega) (d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (2.4)$$

where we have introduced a radial coordinate ω , which is comoving too. The factor f_k depends on the curvature of the universe by the following relation

$$f_k(\omega) = \begin{cases} k^{-1/2} \sin(k^{1/2} \omega) & k > 0 \quad (\text{closed}) \\ \omega & k = 0 \quad (\text{flat}) \\ |k|^{-1/2} \sinh(|k|^{1/2} \omega) & k < 0 \quad (\text{open}). \end{cases} \quad (2.5)$$

2.1.4 Cosmological redshift

A photon emitted by a distant source located a comoving radial coordinate ω_E and at a time t_E will be received by a comoving observer at a time t_R at the origin of the spatial coordinate system (without loss of generality because of the cosmological principle). By looking at the geodesics that the photon follow, we can see that $d\theta = d\varphi = 0$ and since the photon has no momentum, we have $g_{\mu\nu}dx^\mu = 0$: the photon will follow a null geodesics $ds = 0$. The metric reduces, for an incoming photon, to

$$\int_{t_E}^{t_R} \frac{cdt}{R(t)} = \int_0^{\omega_E} d\omega. \quad (2.6)$$

If another photon is emitted at $t_E + \delta t_E$, it will be received by the observer at $t_R + \delta t_R$:

$$\int_{t_E}^{t_R} \frac{cdt}{R(t)} = \int_0^{\omega_E} d\omega = \int_{t_E + \delta t_E}^{t_R + \delta t_R} \frac{cdt}{R(t)} \iff \int_{t_E}^{t_E + \delta t_E} \frac{cdt}{R(t)} = \int_{t_R}^{t_R + \delta t_R} \frac{cdt}{R(t)}. \quad (2.7)$$

If the lag between emissions, δ_E , and the lag between receptions, δ_R , are small, $R(t)$ is a constant and the above simplifies to

$$\frac{\delta t_R}{R(t_R)} = \frac{\delta t_E}{R(t_E)}. \quad (2.8)$$

In other words, the photons suffered a (cosmological) Doppler effect. We define the redshift z , often used as a proxy for a distance (we will see later why), in terms of wavelength of the photon, λ , as

$$z = \frac{\lambda_R - \lambda_E}{\lambda_E} = \frac{R(t_R)}{R(t_E)} - 1 = \frac{a(t_R)}{a(t_E)} - 1. \quad (2.9)$$

This quantity is null for a photon emitted very close to its observer. If z is positive, the photon is redshifted, i.e. appears reddened to the observer and was emitted in an expanding universe (in that case $a(t)$ is increasing with t).

Velocities of distant objects were measured through redshift in early years of the 20th century, first by [Slipher \(1913\)](#) who measured the Andromeda galaxy, and then of other known galaxies ([Slipher 1915](#)). Cosmological redshifts were measured off twenty-two objects by Edwin Hubble. Both authors concluded that most objects were receding. Hubble established an empirical relationship between the observed velocity and the distance of the object ([Hubble 1929a,b](#)). [Lemaître \(1927\)](#) first proposed that the recession of the galaxies was due to an expanding Universe.

2.1.5 The Hubble constant and the deceleration parameter

The subscript 0 denotes the present time. The scale factor $R(t)$ can be expanded as a Taylor series around the present time. We introduce two quantities: the Hubble parameter $H(t)$ and the deceleration parameter $q(t)$. The Hubble parameter is $H(t) = \dot{a}(t)/a(t)$ and the deceleration $q(t) = -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)}$. The curious reader will have noticed the rather strange formulation of this deceleration parameter. When it was first introduced, researchers were convinced that the rate of expansion of the Universe was slowing down, thus q would be positive for a deceleration of the expansion of the Universe. The redshift can be written in terms of look-back time $t_0 - t$ (and assuming $t_0 - t \ll t_0$) as

$$z = (t_0 - t)H_0 + (t_0 - t)^2 \left(1 + \frac{q_0}{2}\right) H_0^2 + \dots \quad (2.10)$$

where the redshift only depends on present values of the parameters. For nearby galaxies, $z \simeq (t_0 - t)H_0$ and the proper distance is $d \simeq c(t_0 - t)$. The recession velocity of a galaxy is given by the Doppler shift,

$$v \simeq cz = H_0 d, \quad (2.11)$$

linking redshift to velocity and distance. This is the famous Hubble law. Hubble first measured a value of ~ 500 (km/s)/Mpc for H_0 , vastly superior to today's value of 68-72 (km/s)/Mpc (Planck Collaboration 2016a; Riess et al. 2016; Bonvin et al. 2017). Note that the dimensionless reduced Hubble parameter h is often used as defined by $H_0 = 100 \cdot h$ km/s/Mpc. This amazingly simple result is valid for nearby objects (i.e. low z). Writing the general equation of the look-back time $t_0 - t = \int_t^{t_0} dt$ and of the distance to the emitter is more complicated. The expressions for the general case also depend on the history of the Universe, i.e. on the evolution of the scale factor $R(t)$.

In the above, we used the notion of distance to the emitting galaxy to derive the Hubble law. However, the concept of distance is confusing: the Universe was smaller when the photon was emitted. There can be different distances in the FLRW metric, which do not agree with each other. The coordinate ω is sometimes called the coordinate distance or comoving radial distance. The distance used in the above to infer Hubble's law is the proper distance $d = R(t)\omega$. This definition of the distance cannot be measured in practice. In the gravitational lensing formalism, we will use the angular-diameter distance D_{ang} , which is

$$D_{\text{ang}}(z) = x/\theta = \frac{R_0 f_k(\omega)}{1+z}, \quad (2.12)$$

or, straightforwardly, the ratio between the physical diameter of an object x and its observed angular size θ . There are other distance measures that are defined in terms of observables, notably the luminosity distance and the redshift distance, which both depend on details of the cosmological model.

2.1.6 The Friedmann-Lemaître equations

With the FLRW metric, an isotropic and homogeneous Universe is described by only two equations, known as the Friedmann equations. The first Friedmann equation can be inferred by looking at the 00 component of the field equations, i.e. the time component, and leads to

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{kc^2}{R^2} - \frac{\Lambda c^2}{3} = \frac{8\pi G}{3}\rho. \quad (2.13)$$

By studying the spatial dimensions and rearranging, the second equation is obtained,

$$\frac{\ddot{R}}{R} = \frac{\Lambda c^2}{3} - \frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right). \quad (2.14)$$

These equations describe the expansion of the universe. In the second equation, it is interesting to see that the constant k , which describes the geometry does not appear. In this model, a universe is static when the two last terms are equal. The Λ constant counteracts gravity, hence providing a static solution, satisfying Einstein's beliefs in a static universe (for a short review, see e.g., Ferrara 2017). The term containing the pressure and density decelerates the rate of expansion of the universe (the second derivative of the radius is negative), whereas

the effect of the cosmological constant Λ is to accelerate the expansion. The conservation of energy-momentum ($\nabla_\mu T^{\mu\nu} = 0$) and a combination of the two Friedmann-Lemaître equations (2.13 and 2.14) yield the energy conservation equation

$$\dot{\rho} + H\left(\rho + \frac{p}{c^2}\right) = 0. \quad (2.15)$$

A common interpretation of a cosmological constant is a repulsive force. If we interpret Λ as a fluid, we can write its density as $\rho_\Lambda = \Lambda c^2 / 8\pi G$. Using its equation of state (EOS), $p = w\rho c^2$, we see that, to get an acceleration ($\ddot{R} > 0$) we need $w < -1/3$, i.e. a negative pressure. In fact, $w = -1$ for a cosmological constant. The name of this bizarre, and as yet undetected and elusive, fluid? Dark energy (or sometimes vacuum energy in the literature³).

2.1.7 The content of the universe

The FLRW and Friedmann-Lemaître equations follow from GR. This is however not sufficient to describe the expansion history $a(t)$ as there is an equation missing. This missing equation is the EOS. From injecting the EOS ($p = w\rho c^2$) into equations 2.15 and 2.13, we get a description of the evolution of the scale factor. We can explore the energy components of the universe and their temporal evolution. We assume that the cosmologically perfect fluid consists of three component: radiation, matter and dark energy.

For the radiation (i.e. massless photons at c or very fast, relativistic!, particles), the EOS parameter reads $w = 1/3$. The evolution is given by (remembering that the subscript 0 denotes the current time)

$$\rho_r(t) = \frac{\rho_{r,0}}{(a(t))^4} \quad \text{or} \quad \rho_r(z) = \rho_{r,0} (1+z)^4. \quad (2.16)$$

It is interesting to see that a blackbody radiation behaves as $\rho_r(T) \propto T^4$. The evolution of the temperature of this blackbody radiation is thus linear with redshift

$$T(t) = \frac{T_0}{(a(t))} \quad \text{or} \quad T(z) = T_0 (1+z). \quad (2.17)$$

Constituents for which velocity is negligible compared to c (like, e.g., stars) exert almost no pressure at all. We can safely set $p = 0$ which yields for the evolution of the matter density

$$\rho_m(t) = \frac{\rho_{m,0}}{(a(t))^3} \quad \text{or} \quad \rho_m(z) = \rho_{m,0} (1+z)^3. \quad (2.18)$$

³There is a significant problem in naming this a vacuum energy. There is a discrepancy of $\sim 50 - 120$ orders of magnitude between particle physics predictions and cosmology (e.g., [Bernard and LeClair 2013](#)). To reconcile the two predictions, a constant should be extremely finely tuned (to many decimal points).

Finally, a cosmological constant imposes $w = -1$. The dark matter density can be written as

$$\rho_{\Lambda}(t) = \rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi G}. \quad (2.19)$$

The density of Λ remains constant throughout the history of the universe.

The matter density in the universe is simply the summation of the three densities,

$$\rho = \frac{\rho_{r,0}}{a^4} + \frac{\rho_{m,0}}{a^3} + \rho_{\Lambda}. \quad (2.20)$$

Note that, to simplify the notation, we have dropped the time dependence. The history of this universe is thus modelled by only four parameters: the densities of the three components at present time and the Hubble constant H_0 . If the universe is flat, $k = 0$, and there is no cosmological constant, $\Lambda = 0$, the density is linked to the Hubble parameter; this density is referred to as critical,

$$\rho_{\text{cr}} = \frac{3H^2(t)}{8\pi G}. \quad (2.21)$$

This critical density is the equivalent density required to have a flat universe. We can now define dimensionless density parameters,

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_{\text{cr}}}, \quad \text{with } i = m, r, \Lambda, \quad (2.22)$$

which are convenient when re-writing the second Friedmann-Lemaître equation (eq. 2.14),

$$1 = \Omega_r + \Omega_m + \Omega_{\Lambda} + \Omega_k. \quad (2.23)$$

The last term of the above equation is the curvature density parameter $\Omega_k = -c^2 k / H^2 R^2$. This relation is remarkable: the sum of the dimensionless density parameters of the components and the curvature is equal to one. This is true throughout the history of the universe. The fate of the Universe can be predicted from the values of the density parameters, as shown in Fig. 2.2.

The *Planck* space mission⁴ aimed at measuring these cosmological parameters (Planck Collaboration 2016a). They can be summarised as: the Hubble constant $H_0 \simeq 70$ (km/s)/Mpc, the matter density $\Omega_{m,0} \simeq 0.3$ and the dark energy density parameter $\Omega_{\Lambda} \simeq 0.7$. In the current epoch, the density of radiation is completely negligible. These parameters are usually analysed jointly with other probes, like the study of supernovæ Ia to infer H_0 with high precision (Riess et al. 2016). There are a few surprises in the measured cosmological parameters:

1. There is a cosmological constant (or at least something very similar to one). The Universe we live in is actually dominated by its dark energy content, of which we know

⁴<http://sci.esa.int/planck/>

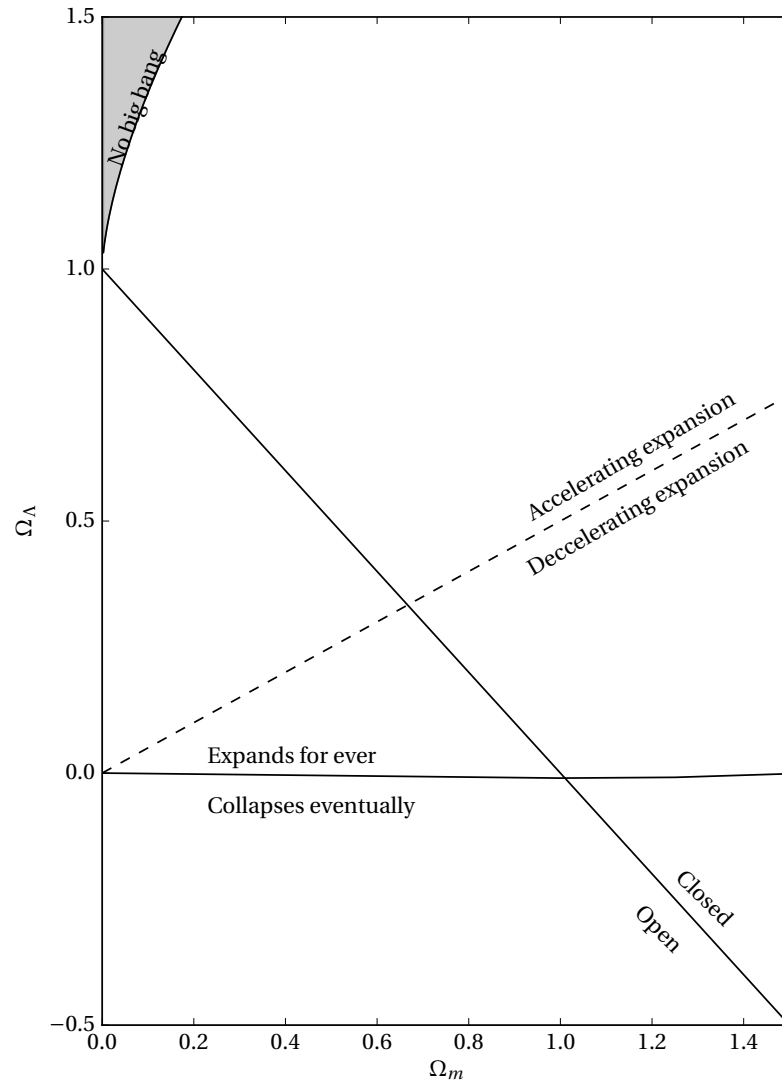


Figure 2.2 – Fate and behaviour of FLRW universes as function of the parameters $(\Omega_m, \Omega_\Lambda)$.

practically nothing. It was completely unexpected when it was discovered at the turn of the millennia by [Riess et al. \(1998\)](#) and [Perlmutter et al. \(1999\)](#). The data thus favours a ever-expanding universe, at a rate which increases with time (i.e. $q_0 < 0$)!

2. It is worth noticing that the matter content includes baryonic matter (i.e. common matter: stars, planets, dust, ...) and dark matter (see Sect. 2.2.5). One sixth of the total matter density in the Universe is in the form of baryonic matter. This suggests that we somewhat understand about 5% of the total energy density in the Universe.
3. The measured curvature is very close to flatness, thus $\Omega_k \simeq 0$; the equivalent total density of the Universe is thus approximately equal to the critical density, $\rho_{\text{cr}} \simeq 9.2 \cdot 10^{-27} \text{ kg m}^{-3} \simeq 5.5 \text{ protons m}^{-3}$.

We live in a dark-energy dominated Universe. It was not always the dominant component. To study the evolution of the different components, we can once again re-write the Friedmann equation, in terms of the Hubble parameter and the present-time value of the density parameters,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = H_0^2 (\Omega_{m,0} a^{-3} + \Omega_{r,0} a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0} a^{-2}). \quad (2.24)$$

Setting all of the densities to zero, except one can help draw a picture of the evolution of the Universe. During the first era, short after the big bang, our Universe consisted of a gas of relativistic particles. This era was dominated by radiation. The scale factor grew as the square root of the time. Then, because of the expansion, the temperature of this gas was cooling down. The energy of the gas decreasing, the particles stopped being relativistic and we transitioned into the second era, which was matter dominated. The scale factor evolved proportionally to $t^{2/3}$. The third era is the current one, dark energy is dominant and the scale factor grows exponentially with time.

2.2 Observational cosmology

The physical understanding of cosmology as, partly exposed in the above, prompted numerous attempts to (i) test its predictions and (ii) measure its parameters. We will now review some important observational aspects and describe the current consensus of the community of the model of the Universe. We select topics which are particularly relevant to gravitational lensing (like the large-scale structure, dark matter and dark energy) or to *Euclid* science and shamelessly ignore other important subjects like the Big Bang nucleosynthesis.

2.2.1 Type Ia Supernovæ

As already mentioned above, this technique was key in the discovery of the acceleration of the expansion (i.e. the famous $q_0 < 0$) of the Universe (Riess et al. 1998; Perlmutter et al. 1999). If a white dwarf in a binary system (we will expand on multiple star systems in Sect. 5.1) accretes material from its companion, the core density and temperature increase. If and when such a white dwarf reaches the Chandrasekhar mass limit, carbon fusion ignites provoking a catastrophic supernovæ type Ia (SNIa) explosion. The stars undergoing this scenario have the same mass at the explosion, hence the explosions are similar (the rate of SNIa explosion changes as a function of the evolution of host galaxies, see e.g., Maoz and Mannucci 2012). They produce the same light curves which are exploited as standard candles, and one can measure the luminosity distance to the host galaxy. Calibrations of the method are necessary. They are based on the so-called distance ladder principle. Simply put, one distance determination calibrate other methods. Say we observe a galaxy in which it is possible to measure the distance to the observer with two different technique. A well constrained method (sometimes referred to as anchor), but working at a smaller range is used to calibrate the further-reaching

method. Combined with determination of the redshift of the galaxy, one can determine H_0 in the Hubble law. Supernovæ-monitoring programs continue to this day and provide tight constraints on H_0 (e.g., [Riess et al. 2016](#)).

2.2.2 The image of last scattering: Cosmic Microwave Background

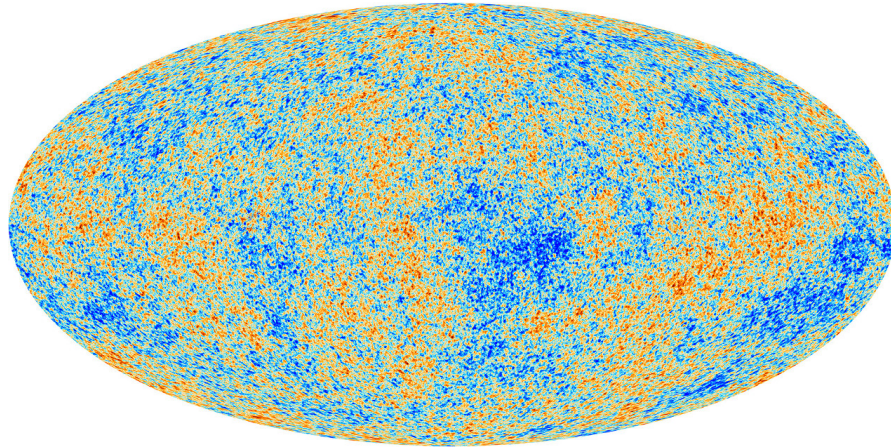


Figure 2.3 – Fluctuation of the CMB temperature as observed by the *Planck* mission. Credits: ESA and the *Planck* Collaboration.

In the period after the Big Bang, the temperature was too high for electrons and protons to pair and form hydrogen atoms. At that time, photons were scattered by the then free-flying charged particles. An effect of the expanding Universe was to cool down its content and about 380 000 years after the Big Bang, the temperature passed the level that allowed to form stable neutral hydrogen. The Universe ceased to be opaque to photons and became transparent. The Cosmic Microwave Background radiation (CMB) is the relic of this last scattering of photons. As the Universe expands, the energy of the CMB photons gets diluted in a larger volume, and the CMB photons cool down. In first approximation, the temperature evolves linearly with redshift (remember eq. 2.17). The photons were emitted at a temperature nearing $T \simeq 3000$ K, but can be observed today at a temperature of $T_0 \simeq 2.73$ K. In the late 1940s, predictions of the CMB began to emerge out of the work of Ralph Alpher, Robert Herman and George Gamov. The signal was first serendipitously observed by two engineers working on perfecting antennas in 1964: Arno Penzias and Robert Wilson and reported in [Penzias and Wilson \(1965\)](#).

Extracting cosmological information out of this signal is demanding. First, the motion of the Earth imprints a characteristic Doppler shift and foreground emissions of our Milky Way contaminates the signal. Once the pre-processing of the data finished, a remarkably isotropic picture is unveiled. The remaining temperature fluctuations are of the order of $\Delta T/T \sim 10^{-5}$! Discovering and characterising the CMB was a great success for modern cosmology as it confirmed the predictions of (i) the black-body spectrum (so precise and accurate that the error bars on plots comparing measurements and observations must be enlarged by a factor of a few hundreds to be visible) and of (ii) the power spectrum of the anisotropies.

The CMB signal provides a wealth of information, and we will not review all what we can learn from it. CMB measurements mainly place constraints on the geometry (i.e. Ω_k), matter and radiation contents. Both $\Omega_m h^2$ and $\Omega_b h^2$ can be determined, i.e. the total density of matter and the baryonic content (such that $\Omega_m = \Omega_b + \Omega_{\text{DM}}$, with DM standing for dark matter). The dark energy content Ω_Λ is also constrained, although weakly if only CMB data is used. Joint analyses of multiple probes lift their intrinsic degeneracies and allow a better determination of the parameters. The observations by the *Planck* satellite revealed unprecedented details in the CMB temperature fluctuations compared to past experiment (Fig. 2.3) and constrained cosmology further.

2.2.3 Large-scale structures

As presented in Fig. 2.1, galaxies in our Universe are not located randomly. This fact is known since the mid 1930s (Hubble 1934). They congregate in huge clusters linked by filaments as in a cosmic web and separated by huge empty regions. These so-called large-scale structures are very interesting objects to study. Indeed, their evolution and statistical properties are imprinted by the primordial density fluctuations from which they have grown. Tracing the large-scale structures with surveys like 2dF and SDSS requires the assumption that the galaxy distribution follow the distribution of dark matter. We will see later (Sect. 2.4) that this condition is relaxed in the case of weak gravitational lensing studies. The growth of structures in the radiation and matter era can be described using perturbation theory (e.g., Peacock 2003). Later, other tools, such as hydro- or N -body simulations, are needed. Such a simulation is the 2-trillion particle *Euclid* Flagship simulation of the universe, with the goal of providing theoretical predictions to be compared with the data. The full-sky projection at $z = 0$ is shown on Fig. 2.4.

The seeds of the large-scale structures are the small density fluctuations in an early universe. An important quantity to describe structure formation is the relative density contrast δ , i.e. the difference of the local density to the mean density normalised by the mean density. An overdense region has $\delta > 0$ while an underdense region has $\delta < 0$. As suggested by the very small fluctuations of the CMB temperature, those fluctuations must have been small in the early ages. The overdensities or underdensities grow over time due to their own gravity. The local expansion of space-time depends on the local density. A overdense (underdense) region will expand more slowly (faster) than a mean-density region. This gravitational instability leads to an increase of $|\delta|$, and the evolution of the initial fluctuations into clusters of galaxies and voids. The details of the growth of the structure depend on dark energy and dark matter (e.g., Frieman et al. 2008; Huterer et al. 2015). Large-scale structures are thus an excellent tool for cosmology (for a review in the context of *Euclid*, see Amendola et al. 2016) and a very active field of research (Libeskind et al. 2018). We will develop more on how large-scale structures are used in weak lensing surveys in Sect. 2.4.6, 2.4.8 and 2.4.10.

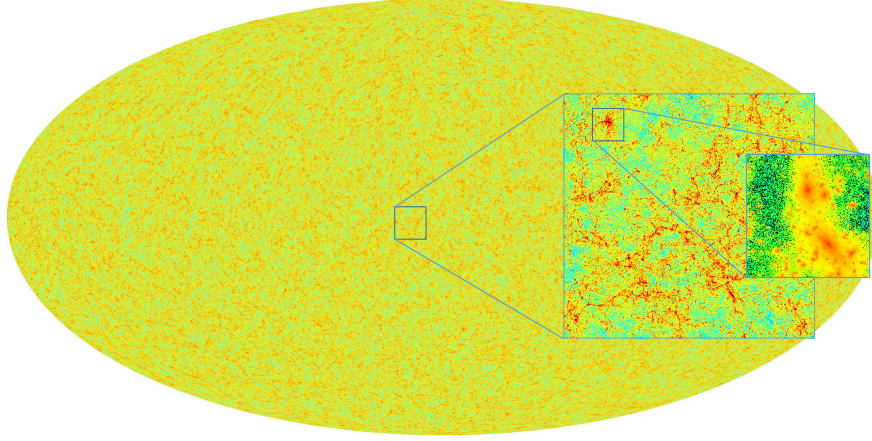


Figure 2.4 – A full sky projection of the dark matter distribution from two-trillion particle simulation for the *Euclid* mission at $z = 0$. The insets in the N -body simulation show the clumpy and filamentary structure. Credit: [Potter et al. \(2017\)](#) and the *Euclid* Consortium Cosmological Simulations Science Working Group (*Euclid* Consortium) (SWG).

2.2.4 A cosmological concordance model

Many cosmological observations made over the past two decades would suggest that our Universe is in agreement with the so-called cosmological constant as dark energy plus cold dark matter model (Λ CDM) model. This model encapsulates a cosmological constant as dark energy plus cold⁵ dark matter. This Λ CDM model is extremely successful, most notably in the prediction of the CMB and models the observed large-scale structures very well. It represents a flat universe with $\Omega_k = 0$. In their present-day value ([Planck Collaboration 2016a](#), and according to the *Planck* 2015 results), dark energy makes up $\Omega_\Lambda \simeq 67\%$ (for cosmological constant Λ , the EOS parameter is $w = -1$), the dark matter $\Omega_{\text{dm}} \simeq 26\%$ and the baryonic matter only $\Omega_b \simeq 5\%$. The *Planck* collaboration inferred a Hubble constant of $H_0 \simeq 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$. In addition, it is a particularly thrifty model: the universe is fully described by only six parameters. Despite its enormous success, there remain issues and shortcomings (e.g., [López-Corredoira 2017](#)). The uniformity of the CMB and the remarkable flatness of the Universe however cannot be explained by this model alone. Those two problems, respectively known as the horizon and flatness problems, can be solved by extending it and including a phase of extremely vast and fast inflation. Inflation does not solve everything, there still open issues with this standard model of cosmology. For example, clear tensions between the *Planck* data and other probes were unveiled in the past years (e.g. [Freedman 2017](#)). The community as a whole is currently working to understand these tensions. Are they the result of yet unknown systematic biases or hint of some new physics? On the smaller scales of galaxy formation and evolution, there are also issues. Results of cosmological simulations based on Λ CDM models deviate from the observations on the number of galaxy satellites (the “missing satellite problem”), core densities of halos (the “too big to fail” problem), and on the profile

⁵Cold as in non-relativistic.

of mass halos (“the cusp-core” conundrum), just to name a few. The solutions could lie in the better understanding of the baryonic and dark physics. I.e. a better understanding of the different processes to be modelled and of dark matter. For a short review, we refer the reader to [Weinberg et al. \(2015\)](#).

The six parameters are the (i) baryonic and (ii) dark matter densities, (iii) the age of the Universe, (iv) the scalar spectral index, (v) the curvature fluctuation amplitude and (vi) the reionisation optical depth. Those parameters may not be directly observable by the different probes. Many parameters can be derived from the six usual parameters of the Λ CDM model. One of them is the normalisation of the matter power spectrum, denoted as σ_8 , which is often determined observationally. The mouthful power spectrum of the matter density perturbations is usually shortened to matter power spectrum. The parameter σ_8 is also referred to the density perturbation amplitude. This σ_8 parameter is the root-mean-square (RMS) linear fluctuation in mass distributions on scales of $8 h^{-1}$ Mpc and has a value of the order of unity. This density perturbation amplitude is computed in spheres of radii $R = 8 h^{-1}$ Mpc,

$$\sigma_8 = \frac{1}{b} \sqrt{\left\langle \left(\frac{\Delta n}{\bar{n}} \right)^2 \right\rangle} \approx 0.8. \quad (2.25)$$

In the above, n is the local number of galaxies, \bar{n} is the average density of the galaxy population, $\Delta n = n - \bar{n}$ and b the linear bias factor, which is the ratio between galaxy overdensities and dark matter, i.e.,

$$\left(\frac{\delta \rho}{\rho} \right)_{\text{galaxies}} = b \left(\frac{\delta \rho}{\rho} \right)_{\text{mass}}. \quad (2.26)$$

For a gentle introduction to this parameter, see [Schneider \(2006\)](#). Determining the value of σ_8 is a degenerate problem in gravitational lensing: only the combination $\sigma_8 \Omega_m^\alpha$ (where $0.2 \lesssim \alpha \lesssim 0.8$) can be measured. Joint analysis of different probe can lift the degeneracy (e.g., for weak gravitational lensing surveys, [Kilbinger et al. 2013](#); [Alsing et al. 2017](#)). Current measurements of $S_8 = \sigma_8 \sqrt{\Omega_m}/0.3$ exhibit tensions with the *Planck* 2015 results (see discussions in [Abbott et al. 2016](#); [Hildebrandt et al. 2017](#)).

2.2.5 Dark matter

Observations of dark matter are not particularly recent (for a historical review, see [Bertone and Hooper 2016](#)). Indeed, Fritz Zwicky and Knut Lundmark measured the radial velocity of galaxies and reported that the galaxies were significantly more massive than expected ([Lundmark 1930](#); [Zwicky 1933](#)). Another very influential paper was published by [Rubin and Ford \(1970\)](#). They measured the rotation curve of galaxies and observed a flattening of the rotation speed at high radius rather than the Kepler profile which confirms the presence of a large amount of unseen mass ([Roberts and Rots 1973](#)). Those two types of observations lead, to the surprise of the authors, to the conclusion that the mass-to-light ratio M/L was

more than one on average and sometimes significantly more than unity. We know now that the mean $M/L \sim 5$, indeed the Universe consists of $\sim 5\%$ of ordinary matter and $\sim 26\%$ of dark matter. These early observations alone could be explained by an extra unknown and unseen mass or by a modification of the laws of gravity at large scales. A clear detection of dark matter was made in a merger of two galaxies dubbed the Bullet Cluster (Clowe et al. 2004, 2006). The luminous and dark matter distributions were located at different positions. This is evidence that dark matter cannot be simply explained away by modified gravity theories. Additional observations of dark matter are obtained by the CMB power spectrum and structure formation. Nowadays, and after an enormous amount of research, the evidence of dark matter are numerous, but its nature is still elusive. Cosmological constraints favour a cold dark matter. A number of dark matter candidates are proposed, but the observational constraints, both coming from cosmology and particle physics, fail to distinguish between them. Perhaps a minor fraction of dark matter is of baryonic matter (like e.g. white dwarfs), but it is likely to be an exotic particle (Freese 2017). Dark matter searches are under way both in particle physics and cosmology, but as yet, there is no definitive answer.

2.2.6 Baryonic acoustic oscillations

Baryonic acoustic oscillations (BAO) are imprints in the galaxy clustering of early universe sound waves. We present the fundamentals of BAO; for an in-depth introduction, see Weinberg et al. (2013). This effect was first studied as it impacted CMB data, but soon after it was also understood to affect the matter power spectrum (see, e.g., Peebles and Yu 1970). In the hot, dense, and young Universe (for $z \lesssim 1000$), the gas is ionised. At that point, the photons are coupled to baryons. The consequence is that the radiation pressure of photons is large with respect to the gravitational forces. Baryonic perturbations will thus oscillate as sound waves. Photon diffusion damps the oscillations for comoving scales smaller than $\sim 8h^{-1}$ Mpc. At recombination, the mean free path of the photons drastically increases, and they decouple from the baryons. The perturbations are no longer subject to the large radiation pressure, but only to gravitational forces. These perturbations coalesce to form galaxies. Contrary to dark matter perturbations, which develop in place, baryon-photon perturbations migrate away from the point of origin. So at recombination, the baryon waves are left frozen in a shell. Galaxy correlation functions show an excess of pairs at the acoustic scale of ~ 150 Mpc in the current Universe (Eisenstein 2005). The field is still young: the first detections of this excess in the matter power spectrum were published in 2005 (Cole et al. 2005; Eisenstein et al. 2005).

BAOs provide a cosmological ruler that can be used to infer the angular-diameter distance $D_{\text{ang}}(z)$, a measure of the expansion history, and the Hubble parameter $H(z)$. Combining CMB results with BAO yields the matter density Ω_m , and lifts a degeneracy in supernovae inference, such that the EOS w is accessible as well. The BAO measurement technique require a very large sample of galaxies and accurate redshift determination. To properly measure the BAO standard ruler at different redshifts, the data should be of similar quality, irrespective of the redshift of the source. Instrumental and astrophysical biases (such as photometric errors,

and non-linear structure formations) should also be controlled to determine the most precise and accurate cosmological parameters possible. BAO surveys require so much data that it is foreseen that future surveys, like *Euclid*, might still be dominated by statistical errors (Laureijs et al. 2011).

2.2.7 Going beyond Λ CDM to explain dark energy?

The addition of the cosmological constant to Einstein's field equations (eq. 2.1) is consistent with observations. Another great advantage of the cosmological model is that it is particularly economical. The *Planck* observations favour a constant value for the EOS parameter w in a flat universe, and this is consistent with the standard Λ CDM model. However, there remain issues which were identified even before its surprising come-back in the early 21st century (for an interesting discussion about the cosmological constant problem, see Weinberg 1989). Alternative theories that fit observations have been developed (for recent reviews see, e.g., Joyce et al. 2016; Wang et al. 2016). There are two broad classes of ideas: (i) dark energy models and (ii) modified gravity. The former deals with additional components to the EOS (e.g. the parameter w depends on the normalised scale factor a : $w(t) = w_0 + w_a[(1 - a(t))]$) while the latter re-writes general relativity. This is to say, they alter one, the other or both sides of the field equations: there is some new component in the universe or gravity has as yet unknown properties, or a mix of the two. General relativity has been extraordinarily successful in its predictions and has become essential even for day-to-day applications like navigation. However, GR breaks down at the smallest scales. This observation should prompt us to be vigilant at the largest scales too. To be consistent with the observations, a cosmological constant must be added to GR, which might be a hint that something else is going on. However, GR has been tested in a vast range of distance, without failure so far. The nature of dark energy is not constrained in the concordance model. Let's take one example of an alternative approach. The cosmological principle is the basis of the current understanding of our Universe. There are ways not to invoke dark energy if we suppose inhomogeneity (for a another recent review, see Debono and Smoot 2016). The mysterious origin of 95% of the content of the Universe is deeply troubling, in particular because of the large contribution of this extremely poorly known dark energy.

Even if some of the alternative models proposed are viable, they also suffer from limitations. Despite all of its shortcomings, Λ CDM remains the most trusted model, due partly to the lack of convincing alternative (for a survey of the community, see Bull et al. 2016). The cosmologically-oriented surveys, like *Euclid*, also evaluate their results in terms of non-concordance models. With the increasing sensitivity of surveys and joint analyses of multiple probes, the predictions of the models will be stress-tested with unprecedented force. There is a bright future for cosmology. The vast amount of data that are about (or just have) become available to the community allow for a much finer model selection and thus ruling out of unfitting theories. GR is very much in the spotlights nowadays, a hundred years after its scientific dawn. Even though doubts exists about its validity, GR is a very-well posed theory to remain at the forefront

for many years to come.

2.3 Gravitational lensing effect

We turn to the gravitational lensing effect, the different regimes, their applications and implications for astrophysics and cosmology. The goal of this section is to provide a background that motivates the aims of this thesis. The interested reader can delve further into this subject by studying the following reviews and monographs: [Schneider et al. \(1992\)](#); [Kaiser \(1998\)](#); [Bartelmann and Schneider \(2001\)](#); [Wittman \(2002\)](#); [Munshi et al. \(2008\)](#); [Schneider et al. \(2006\)](#); [Hoekstra and Jain \(2008\)](#); [Huterer \(2010\)](#); [Massey et al. \(2010\)](#); [Kilbinger \(2015\)](#); [Bartelmann and Maturi \(2016\)](#); [Mandelbaum \(2017\)](#).

2.3.1 Light travelling in a gravitational field

The idea that the trajectory of light could be bent by the presence of massive objects emerged in the early days of the Newtonian theory of gravitation. This well-known theory predicts that the gravitational acceleration acting on a test particle of mass m is independent of m . This counter-intuitive notion was tested numerous times, but notably during in an unusual experiment. The commander of the Apollo 15 mission to the surface of the Moon, David Scott, simultaneously dropped a 1.32-kg hammer and a 0.03-kg feather from the same height. Both objected fell to the ground in the same time, in accordance to Newtonian theory ([Allen 1972](#)). If light can be treated as a particle, it means that photons would be sensitive to a gravitational potential and its trajectory influenced by this potential. A photon travelling in the gravitational field of a massive spherical object of mass M , say the Sun, is deflected from its original trajectory. Integrating the acceleration over the trajectory of the particle yields a deflection angle of

$$\hat{\alpha}_N = \frac{2GM}{c^2} \frac{\xi}{|\xi|^2}, \quad (2.27)$$

where ξ is the impact parameter. This deflection angle was discussed early on, even before the concept of the photon by [Michell \(1784\)](#) and [von Soldner \(1804\)](#). In addition, at the same period, the great scientist Pierre-Simon de Laplace noticed that if the mass of a spherical object is large enough, the escape velocity could be higher than the speed of light, which arises when the radius of the object R is

$$R \leq r_s = \frac{2GM}{c^2}, \quad (2.28)$$

what will be referred to the Schwarzschild metric in modern days ([de Laplace 1798](#)). Introducing the work by Albert Einstein and considering the general relativity formalism and

computing the same angle yields a deflection angle twice as large

$$\hat{\alpha} = \frac{4GM}{c^2} \frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|^2}, \quad (2.29)$$

provided that the ray impact parameter ξ is much larger than the Schwarzschild radius. This can be computed by deriving the equations of movements, considering that photons follow null geodesics in the Schwarzschild metric,

$$ds^2 = \left(1 - \frac{r_s}{r}\right) c^2 dt^2 - \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (2.30)$$

with r the radial coordinate, θ and φ two angular coordinates (the three spatial dimensions) (Schwarzschild 1916) and integrating along its trajectory. The assumption of a much larger impact parameter than the Schwarzschild radius implies that the Newtonian gravitational field is weak, i.e. $\phi_N/c^2 \ll 1$. In that case, the field equations can be linearised.

The deflection by an ensemble of point masses is, in this simplified linearised picture, the vectorial sum of the individual deflection angles. All the individual deflection angles are small, thus the trajectory of the photons can be associated with straight lines close to the deflecting masses. The light propagating through a 3D mass distribution will cross the geometrically-thin lens plane at a location (ξ_1, ξ_2, r_3) . The coordinates were chosen such that the ξ_1, ξ_2 coordinates are perpendicular to the line-of-sight. The line-of-sight runs along r_3 . The light, which propagates from the source object towards the observer, crosses the thin lens plane located at r'_3 and at a position $\boldsymbol{\xi} = (\xi_1, \xi_2)$. A mass element dm in the lens plane is spatially described by $\mathbf{r}' = (\xi'_1, \xi'_2, r'_3)$, thus the impact parameters between the incoming light and the mass is $\boldsymbol{\xi} - \boldsymbol{\xi}'$. The total deflection angle can now be expressed as

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \sum dm(\xi'_1, \xi'_2, r'_3) \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} = \frac{4G}{c^2} \int d^2\xi' \int dr'_3 \rho(\xi'_1, \xi'_2, r'_3) \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}, \quad (2.31)$$

where we used the fact that the mass elements dm can be expressed as $dm = \rho(\mathbf{r})dV$. The notation can be simplified by defining the surface mass density

$$\Sigma(\boldsymbol{\xi}) = \int dr_3 \rho(\xi_1, \xi_2, r_3). \quad (2.32)$$

Equation (2.31), the deflection angle produced by an arbitrary density distribution, now reads

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int d^2\xi' \Sigma(\boldsymbol{\xi}') \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}. \quad (2.33)$$

This angle is a two-dimensional vector, since it depends on position in the lens plane $\boldsymbol{\xi}$, which is a two-dimensional vector too. An assumption used to derive the above equation is that the deflection of the light ray is small compared to a scale on which the mass distribution changes. The deflection angles are typically very small. This assumption is referred to in quantum mechanics as the Born approximation. This assumption is no longer valid in cases where the

light emitted by a distant source propagates through the large scale structures. This case is typical of weak gravitational lensing and will be treated in Sect. 2.4.2.

2.3.2 The eclipse experiments

In 1911, when Albert Einstein first worked on the problem of the deflection angle in the framework of special relativity, he ended up with the same equation as the Newtonian deflection angle, i.e.

$$\alpha_N = 0.875'' \left(\frac{M}{M_\odot} \right) \left(\frac{\xi}{R_\odot} \right)^{-1}. \quad (2.34)$$

It is worth noting that this value of the deflection angle is valid for an object whose impact parameter is the radius of the limb of the Sun. In 1915, he also considered the curvature of space around the Sun and wrote the general relativity deflection angle (Eq. 2.29)

$$\alpha = 1.75'' \left(\frac{M}{M_\odot} \right) \left(\frac{\xi}{R_\odot} \right)^{-1}. \quad (2.35)$$

Einstein tried to get observational confirmation as early as 1912, and quickly reached the conclusion that the only option was to measure the positional change of stars on photographic plate taken during a total solar eclipse (Ellis 2010). After a number of failed experiments⁶, Arthur Eddington and Frank Dyson proposed an expedition to observe the total solar eclipse on 29 May 1919 from two locations in Brazil and off the coast of western Africa. The results presented at the Royal Astronomical Society showed good agreement with the expected angle (eq. 2.35) (Eddington 1919; Dyson et al. 1920). The statistical significance of the 1919 eclipse was low, and a second experiment was devised to observe another eclipse from Australia in 1922. The measurements corroborated the 1919 results with increased precision (Campbell and Trumpler 1923). Further experiments failed to measure the displacement of stars with better precision.

2.3.3 The dawn of gravitational lensing

In the early days of gravitational lensing, the notion of galaxies as an extra-galactic collection of vast amounts of materials was only brewing (for a historical approach to the nature of galaxies, see e.g. Fernie 1970). Thus, the efforts of the time were concentrated on gravitational lensing by objects located in our Milky Way such as massive stars. In 1924, Orest Chwolson pondered the particular case in which the source, the lens and the observer were aligned. He noticed that, if the mass of the star that was playing the role of the lens was large enough, the source would appear as a ring (Chwolson 1924). Later, Albert Einstein published a note that described the gravitational lensing of one star on another, but concluded that there was no hope of actually observing it as the resolving power of instruments was too low for the

⁶Multiple expeditions were cancelled or abridged due to the 1st World War or instrumental issues.

tiny scale (Einstein 1936). It was learned much later, thanks to the discovery of unpublished notes, that Einstein had already studied the phenomenon of ring images in 1912 (Renn et al. 1997). In 1937, Fritz Zwicky published two influential papers in which he postulated that galaxies (known as “extra-galactic nebulae” at the time) could easily act as gravitational lenses. Nowadays, distant extra-galactic sources are not a scarce resource, but they were at the time. No lenses or lensed objects were catalogued. However, Zwicky proposed that gravitational lensing would be useful to (i) weigh the lensing galaxies and (ii) use the phenomenon as a natural telescope to see galaxies brighter than they would be in the absence of lensing (Zwicky 1937a). Moreover, he concluded that the probability of producing multiple images by a lens was roughly 0.25%, which is currently known as the lensing cross section and predicted that finding a multiple-lensed image is practically a certainty (Zwicky 1937b). However, after these optimistic predictions, the field, which was still in its infancy, came to a long standstill. It was only in the 1960s that there was a renewed interest in the field, with two studies by Klimov (1963) and Liebes (1964) on galaxy-galaxy lensing and on a prediction of lensing statistics respectively. The latter study will actually give rise to surveys to detect Milky Way stars acting as lenses for stars in the Magellanic Clouds. Those surveys will be implemented decades later, in the 1990s. Still in 1964, Refsdal (1964) realised that, if there are multiple images, the light follows different paths of different length. The light travel times are thus not the same. Measuring this time delay would allow to measure (i) the radial mass profile of the lens galaxies and (ii) the distances between the different objects, which is the first step to measure the Hubble constant H_0 . Quasars are peculiar objects: discovered by Hazard et al. (1963), they are some of the brightest astronomical objects, visible at the cosmological scale and yet point-source-like. Their power source is the accretion of mass onto a super-massive black hole. The quasar QSO 0957+561 was the first lensed object ever to be observed (Walsh et al. 1979). They are very distant, hence they are the ideal sources craved by Zwicky over 40 years earlier and they are photometrically variable, hence ideal for Refsdal. In the decades that followed, more lensed quasars were discovered, with two or four visible images of the sources, including the so-called Einstein cross (QSO 2237+0305, Huchra et al. 1985). Those multiple imaged quasars were still not what Einstein and Chwolson predicted, but in the mid 1980s, two teams discovered an arc in the galaxy cluster Abell 370 (Lynds and Petrosian 1986; Soucail et al. 1987). The nature of this bluish arc was not grasped at once. Paczynski (1987) proposed the, correct, lensing interpretation. Finally, the first full Einstein ring was discovered by Hewitt et al. (1988) in the radio domain.

2.3.4 The lens equation

A typical lensed system is shown on Fig. 2.5, with the source object located at a distance D_s and the lens plane at a distance D_l , both from the observer. The distance from the lens to the source is D_{ls} . If the lens plane is thin enough (i.e. its thickness is much smaller than D_l and D_{ls}), the light propagation can be considered straight before and after the deflection. The deflection is described by the two-dimensional vector $\hat{\alpha}$, which depends on the impact vector and the mass distribution. In the source plane, the object is located at a true position η which,

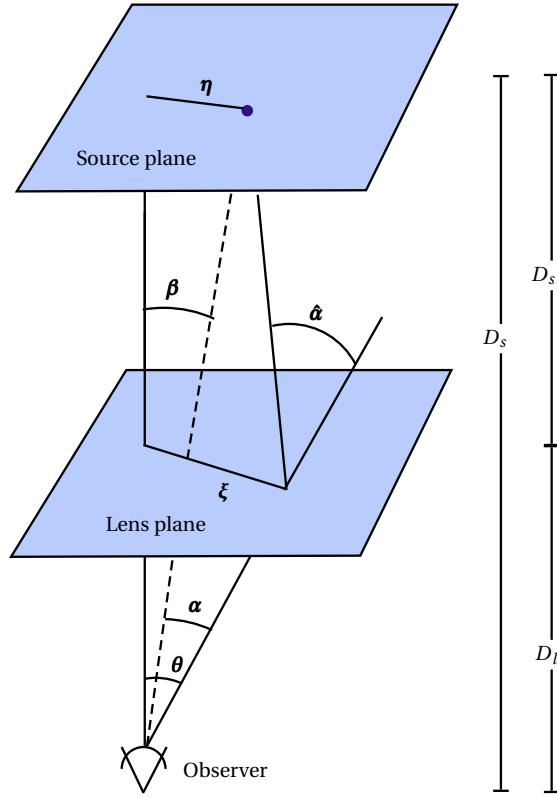


Figure 2.5 – Representation a lens system. The source is the blue dot in the source plane at a distance D_s , while the lensing masses are distributed in the lens plane, at a distance D_l .

if observed unlensed, would appear with an angular position of β . As the system is lensed, the observed source position is θ . From Fig. 2.5 and assuming small angles, the true position η can be related to the impact vector and the deflection angle

$$\eta = \frac{D_s}{D_l} \xi - D_{ls} \hat{\alpha}(\xi). \quad (2.36)$$

The true position and the impact vectors can be rewritten as $\eta = D_s \beta$ and $\xi = D_l \theta$ respectively. This leads to the lens equation

$$\beta = \theta - \frac{D_{ls}}{D_s} \hat{\alpha}(D_l \theta) = \theta - \alpha(\theta), \quad (2.37)$$

where $\hat{\alpha}$ is called the scaled or reduced deflection angle. This lens equation is a mapping from the lens plane to the source plane. If there are more than one solutions to the lens equation, then multiple images of the same source object can be seen. We define the convergence $\kappa(\theta)$ as

$$\kappa(\theta) \equiv \frac{\Sigma(D_l \theta)}{\Sigma_{cr}} \quad \text{with} \quad \Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_s}{D_l D_{ls}}, \quad (2.38)$$

Chapter 2. Cosmology and gravitational lensing

in which we used the critical surface mass density Σ_{cr} . The scaled deflection angle can be expressed as

$$\alpha(\theta) = \frac{1}{\pi} \int d^2\theta' \kappa(\theta') \frac{\theta - \theta'}{|\theta - \theta'|^2}. \quad (2.39)$$

A mass distribution which $\kappa \geq 1 \Leftrightarrow \Sigma \geq \Sigma_{\text{cr}}$ can produce multiple images. The convergence is a criterion to separate between weak and strong regimes. Using the identity $\nabla \ln |\mathbf{x}| = \mathbf{x}/|\mathbf{x}|^2$, the above equation is written as

$$\alpha(\theta) = \frac{1}{\pi} \int d^2\theta' \kappa(\theta') \nabla \ln |\theta - \theta'| = \nabla \underbrace{\left(\frac{1}{\pi} \int d^2\theta' \kappa(\theta') \ln |\theta - \theta'| \right)}_{\equiv \psi(\theta)}, \quad (2.40)$$

with ψ the deflection potential, in analogy with classical mechanics. Taking the Laplacian of the deflection potential (eq. 2.40) and taking advantage of another identity ($\nabla^2 \ln |\theta| = 2\pi\delta(\theta)$) yields

$$\nabla^2 \psi(\theta) = \frac{1}{\pi} \int d^2\theta' \kappa(\theta') \nabla^2 \ln |\theta - \theta'| = 2\kappa(\theta). \quad (2.41)$$

The convergence is analogous to the 3D mass density. We derived two important relations that will be fundamental in the description of weak gravitational lensing: (i) the deflection angle in function as the gradient of the deflection potential (eq. 2.40) and (ii) the Poisson equation in 2D (eq. 2.41).

In the unlikely case where the source and the observer are aligned the lens equation becomes

$$\beta = \theta - \alpha(\theta) = 0 \implies \theta = \alpha(\theta). \quad (2.42)$$

and using the relationship between the deflection angle and reduced deflection angle in eq. (2.37) leads to

$$\theta = \frac{D_{ls}}{D_s} \frac{4GM}{c^2 D_l \theta} \implies \theta_E \equiv \sqrt{\frac{4GM}{c^2} \frac{D_{ls}}{D_s D_l}}, \quad (2.43)$$

which is called the Einstein radius. Linking the Einstein radius (eq. 2.43) to the critical surface mass and rearranging to emphasise the mass contained within the Einstein radius yields

$$M = \pi \Sigma \xi^2 \implies \kappa = \frac{\Sigma}{\Sigma_{\text{cr}}} = \left(\frac{\xi_E}{\xi} \right)^2, \quad (2.44)$$

i.e. the convergence is larger than 1 if the impact parameter is smaller than the Einstein radius and smaller than one if the impact parameter is larger than θ_E . The effect of the convergence on a source object is shown in Fig. 2.6.

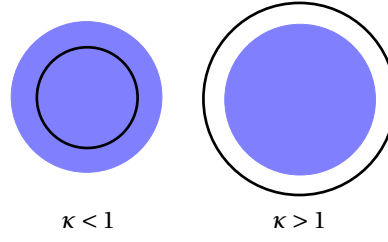


Figure 2.6 – The source (blue disk) is mapped to the lensed image (black circle) with convergence only.

2.3.5 Magnifying and distorting the source image

The observed image of a source is distorted by the lens as the light rays are affected differently depending on their exact impact vector ξ . The exact shape of an extended source object can be derived by evaluating the lens equation for every point of the source object. Gravitational lensing conserves surface brightness⁷ and is achromatic: the wavelength of the photon does not influence its deflection angle. To be able to write down a distortion function between the source and observed images, we also assume that the angular size of the source is much smaller than the scale on which the lens properties change. These two statements allow for a linearisation of the mapping source image to observed image. The source has a surface $d\eta_1 d\eta_2$ as measured in the source plane. If there were no lens, the source would be observed at a solid angle

$$d\Omega_s = \frac{d\eta_1 d\eta_2}{D_s^2}. \quad (2.45)$$

However, due to the presence of a lens, it is observed as

$$d\Omega_l = \frac{d\xi_1 d\xi_2}{D_l^2}. \quad (2.46)$$

The flux is expressed as $F = I d\Omega$ with I the intensity. The ratio of fluxes, called magnification is written (remembering that the angles $\beta = \eta/D_s$ and $\theta = \xi/D_l$)

$$\mu = \frac{F}{F_0} = \frac{\Omega_l}{\Omega_s} = \frac{d\xi_1 d\xi_2}{d\eta_1 d\eta_2} \left(\frac{D_s}{D_l} \right)^2 = \frac{d\beta_1 d\beta_2}{d\theta_1 d\theta_2} = \det \frac{\partial \beta(\theta)}{\partial \theta} \frac{d\theta_1 d\theta_2}{d\theta_1 d\theta_2} = \det \frac{\partial \beta(\theta)}{\partial \theta}. \quad (2.47)$$

where we used the assumption of the small angular size described in the above. The inverse distortion of an image is then described by the Jacobian matrix A , which can be rearranged using the lens equation

$$A = \frac{\partial \beta}{\partial \theta} = \frac{\partial \theta}{\partial \theta} - \frac{\partial \alpha(\theta)}{\partial \theta} \Rightarrow A_{ij} = \delta_{ij} - \frac{\partial^2 \psi(\theta)}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (2.48)$$

⁷This is because photons undergoing gravitational lensing deflection are neither absorbed nor emitted by the lens and is a consequence of the Liouville theorem. See [Schneider et al. \(1992\)](#) for extended discussion of the surface brightness conservation.

where we made use of the concept of shear $\gamma \equiv \gamma_1 + i\gamma_2 = |\gamma|e^{2i\varphi}$ with φ the position angle, or

$$\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22}), \quad \gamma_2 = \psi_{,12}, \quad \kappa = \frac{1}{2}(\psi_{,11} + \psi_{,22}), \quad (2.49)$$

where we used the coma convention to mean $\cdot_{,i} = \partial \cdot / \partial \theta_i$. The Jacobian matrix A has two eigenvectors $1 - \kappa \pm \gamma_1$.

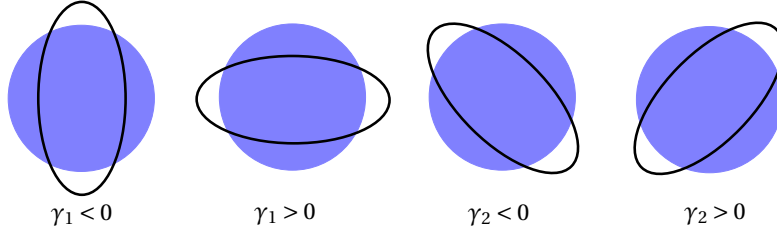


Figure 2.7 – The source (blue disk) is mapped to the lensed image (black circle) affected by shear only.

The observed images will be distorted both in size and in shape as shown in Fig. 2.7. The stretching of the shape is influenced by the shear γ . As the deflection angle depends on the impact vector, the shape is distorted because the light bundles follow slightly different paths. The change of size is caused by a focussing of the light by the convergence κ and the shear γ for the isotropic and anisotropic parts respectively. We will often refer to the reduced shear g in the following,

$$g \equiv \frac{\gamma}{1 - \kappa} = \frac{|\gamma|}{1 - \kappa} e^{2i\varphi}, \quad (2.50)$$

as the reduced shear is the main observable in weak gravitational lensing.

2.4 Weak gravitational lensing and its applications

2.4.1 The weak regime

The amount by which gravitational lensing distorts the source images depends on the mass distribution along the line-of-sight. These distortions can be understood as a change of coordinates between the unlensed coordinates β and the lensed coordinates θ of an image, i.e. the distortions is described by the (inverse of the) Jacobian matrix A as written in eq. (2.48). In the weak gravitational lensing or weak lensing (WL) regime, the transformations are well approximated by a linear transformation and the Jacobian matrix A is close to unity, thus the distortions of the sources are weak and the magnification small. This corresponds to a case for which $\Sigma \ll \Sigma_{\text{cr}} \Leftrightarrow \kappa(\theta) \ll 1$. Due to the weak tidal field there are no multiple lensed images of the source. The source image is only slightly distorted and slightly magnified. The typical variation of the source image shape is of the order of 1% in ellipticity.

The original, unlensed, shape is not accessible; a far-away object can only be observed through

the prism of the natural telescope of gravitational lensing. For that reason, weak lensing analyses require statistical approaches to extract the signal of lensing, i.e. the reduced shear. The magnification is not usually measured as it is a particularly difficult measurement. Remember that we do not know the size of galaxies in the source plane. In recent years, with the development of better acquisition and analysis tools, a few studies have begun to measure convergence in weak lensing simulations and surveys, resulting in proof-of-concept detections (Casaponsa et al. 2013; Alsing et al. 2015). The type of objects used to detect weak lensing signal is distant, high redshift, galaxies. The light incoming from distant galaxies travels through the large scale structures of the universe, and the image of these distant galaxies are lensed by the massive objects on the way. The deeper the image, the fainter the source galaxies and also the smaller. Smaller galaxies are affected more by the convolution of the system (and if from the ground, atmospheric) PSF. Its correction is one of the critical step of the analysis. If neglected or insufficiently corrected, significant biases arises in the measure of the reduced shear thus affecting the measurement of the cosmological parameters (e.g. Kaiser et al. 1995; Erben et al. 2001; Hoekstra 2004; Bartelmann et al. 2012). The effects of the PSF on the shapes are typically stronger than the effect of lensing (Wittman 2002). This convolution of the image by the PSF is part of the so-call forward process. We will discuss this forward process, i.e. the different stages that separate the emission of a photon in the source object to its registration in the memory bank of a telescope station, in section 3.1.

Weak lensing is a tool that directly measures the mass and its distribution in the Universe. Lensing at all regimes do not require astrophysical assumptions to probe the mass of a structure. Strong lensing is impressive by its sheer effect on the observed image of a source, but can only be produced at very specific locations, namely at the centre of very massive clusters. Weak lensing is certainly more difficult to detect and properly measure, but distorts almost all distant objects in the universe. It is only the advances in image acquisition, i.e. the advent of charged coupled devices (CCDs), which allowed to better register incoming photons and to better measure of the light distribution. CCDs, in turns, enabled the detection of (reduced) shear. The first detection by Tyson et al. (1990) was made by observing clusters of galaxies and measuring the alignment of the faint blue galaxies in the image. They showed that the galaxies tend to align perpendicularly to the vector linking their position to the centre of the cluster. The detection of weak lensing by large-scale structures, also known as cosmic shear, was made at the turn of the millennia by five different teams (Bacon et al. 2000; Kaiser et al. 2000; Wittman et al. 2000; Maoli et al. 2001; Van Waerbeke et al. 2000, 2001). Combining the ellipticity measurement with the redshift of sources generates additional information such as constraints on dark energy or the 3D reconstruction of the mass density (which is dominated by dark matter).

The weak lensing tool can provide answers (or at least hints) to a range of subjects dealing with mass (dark and luminous) in the universe and its distribution. A few are listed here: Where are located the overdensities and are their position correlated with the location of clusters? Simply put, how does the luminous mass correlate with the total mass, measured with weak lensing? What is dark matter? Is the dark energy simply a cosmological constant, or does it

evolve with time? How are structured the large-scale structures (their voids, filaments and walls), where are they and can we trace their evolution? What cosmology do they best fit? Do modified theories model the growth of large-scale structure better than the current most popular cosmological models?

2.4.2 The deflection angle for weak lensing

In the derivation of the deflection angle, it was assumed that the lens is small (in the line-of-sight axis) compared to the distances involved. Weakly lensed objects are mostly lensed by the large-scale structures, which violate the assumption of smallness of the lens. The derivation of the results presented in the following is fairly involved and detailed in [Bartelmann and Schneider \(2001\)](#). Here, we summarise the main ideas and results. In lensing, the distances are usually cosmological, and in a universe in which space-time is expanding and which may not be flat, there are different ways of defining distances. We use the notion of angular-diameter distance (as defined in Sect. 2.1.5).

According to Einstein's general relativity, the light propagates along the null geodesics. We will use this principle again. We assume that the universe in which the light propagates is slightly inhomogeneous. The Friedmann–Lemaître–Robertson–Walker metrics can be rewritten to include these slight perturbations. Computing the null geodesics and, in analogy to the deflection angle for thin lenses, a Jacobian matrix can be defined

$$A(\boldsymbol{\theta}, \omega) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{1}{f_k} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}. \quad (2.51)$$

The inverse of the Jacobian matrix describes the locally linearised mapping of the source light distribution to image coordinates. From this, a lensing deflection potential can be established

$$\psi(\boldsymbol{\theta}, \omega) = \frac{2}{c^2} \int_0^\omega d\omega' \frac{f_k(\omega - \omega')}{f_k(\omega) f_k(\omega')} \Phi(f_k(\omega') \boldsymbol{\theta}, \omega'). \quad (2.52)$$

where Φ is the perturbative peculiar Newtonian potential. The reader will have noticed that the matrix $A_{ij} = \delta_{ij} - \psi_{,ij}$ has the same form as the Jacobian matrix in the thin lens equation. The Born approximation has been used to derive eq. (2.52). In this approximation, lensing by an extended 3D distribution of matter, can be treated as an equivalent thin lens plane.

The efficiency of lensing g_i by gravitational potential fluctuations Φ is

$$g_i(\omega) = \int_\omega^{\omega_{\text{lim}}} d\omega' n(\omega') \frac{f_k(\omega' - \omega)}{f_k(\omega)}, \quad (2.53)$$

where we used the source galaxy probability distribution in comoving distance $n(\omega)d\omega$. The galaxy probability distribution is normalised such that $\int_0^\infty dz n(z) = 1$. Rewriting the above

equation in terms of redshift and the common distances in lensing,

$$g_i(z_l) = \int_{z_l} dz_s n(z_s) \frac{D_l D_{sl}}{D_s}, \quad (2.54)$$

i.e. the lensing efficiency is simple the ratio of distances $D_l D_{sl}/D_s$ weighted by the galaxy probability distribution. Let's take a sample of galaxies for which the redshift are known to arbitrary precision and divide it into two bins. The galaxies with $z < z_{\text{median}}$ are classified into bin 1 and the others into bin 2. Eq. (2.54) implies that even though the two galaxy distributions do not overlap, the lensing efficiency strongly correlate. This translates the fact that higher and lower redshift source galaxies are both lensed by low-redshift structures. Eq. (2.54) also highlights that the lensing effect is higher when the distances to the sources are larger.

2.4.3 Reduced shear estimator from galaxy shapes

To estimate the shear from the data, galaxy shapes are measured. The ellipticity ε of a galaxy is the observable in weak lensing studies. There are a few definitions of the word "ellipticity". In the following, unless stated otherwise, we will use

$$\varepsilon = \frac{a-b}{a+b} \exp 2i\varphi, \quad (2.55)$$

with the minor-to-major axis ratio b/a and the position angle φ . We define also the undisturbed, intrinsic, ellipticity of a source as ε^s . The observed ellipticity ε relates the intrinsic ellipticity to the reduced shear g by

$$\varepsilon = \frac{\varepsilon^s + g}{1 + g^* \varepsilon^s}. \quad (2.56)$$

We have used the notation $*$ to denote complex conjugation. This equation is valid in the case in which $|g| < 1$. In the weak-lensing regime, the above equation reduces to (Seitz and Schneider 1997)

$$\varepsilon \approx \varepsilon^s + g. \quad (2.57)$$

If there is no preferred direction to which the galaxies align and a sufficiently large number of observable over which to average, the mean value of the intrinsic ellipticity reduces to zero $\langle \varepsilon^s \rangle = 0$, thus the expectation value of the observed ellipticities is an unbiased estimation of the reduced shear,

$$\langle \varepsilon \rangle \approx g. \quad (2.58)$$

The assumption that there is no preferred direction is however not correct. Galaxies tend to align to each other, thus violating the assumption $\langle \varepsilon^s \rangle = 0$. This topic of intrinsic alignment is discussed further in Sect. 2.4.7. There are other estimators for the shear, in particular in terms

of a combination of second order moments of the image (see, e.g. [Schneider and Seitz 1995](#)). We will describe how to relate second order moments to the reduced shear in Sect. 3.2.1.

2.4.4 Mass map reconstruction

The convergence κ is half of the Laplacian of the potential ψ (See eq. 2.49). The convergence is a direct probe of the (projected) total matter density of the lens. Mass reconstruction from lensing measurements is interesting because no astrophysical assumptions (like relaxation or mass-to-light ratio) should be made. The first dark matter mass maps were established by early measurements of weak lensing in the neighbourhood of galaxy clusters ([Fahlman et al. 1994](#); [Bonnet and Mellier 1995](#)). Comparing mass maps from lensing and those drafted from luminous matter distribution show can differentiated mass distribution for dark and luminous matter. I.e., it's a probe of the nature of dark matter. We will describe this more in Sect. 2.4.8.

The (projected) mass map is the convergence κ map reconstructed from the shear measurement. The shear field γ_i and the convergence κ all derive from the gravitational potential ψ as written in eq. (2.49). Assuming weak lensing regime, [Kaiser and Squires \(1993\)](#) showed that the smooth convergence map can be obtained from a smooth shear map in the Fourier domain. The relationship between convergence and shear field is linear. The observable in weak lensing studies is the reduced shear $g = \gamma/(1 - \kappa)$. The convergence map is a non-linear function of the reduced shear. The caveat of this method is that the shear map must be smoothed to prevent the variance of the mass map from blowing up. The smoothing window is introduced to mitigate the effect of the non-vanishing white noise due to the random intrinsic ellipticities of the source galaxies. Masked regions must be accounted for, the resolution is reduced and correlated noise arise. The convergence map is determined up to an additive constant. This entails that a constant convergence does not induce a shear. A constant convergence is not observable and introduces the mass-sheet degeneracy. Additional information (e.g. the magnification) is required to lift the degeneracy.

If the condition of weak lensing is not met (i.e. $\kappa \sim \gamma_i \sim 1$), the above method breaks down. A likelihood-minimisation method to find the best gravitational potential fitting the (reduced) shear can then be applied (e.g., [Bartelmann et al. 1996](#)). In addition to the choice of smoothing window and the intrinsic ellipticities, an important source of noise is the number density of galaxies (for an early review of the sources of noise see e.g., [Waerbeke 2000](#)). To avoid dealing with masking, in-paintings methods have been proposed ([Pires et al. 2009a](#)). Other techniques to reconstruct the convergence map also exist. They are using non-linear lensing signal ([Pires and Amara 2010](#)), galaxy clustering ([Simon, Patrick 2013](#); [Szepietowski et al. 2014](#)) or redshift information and sparsity (e.g., [Leonard et al. 2014](#)).

There are definite advantages of mass maps compared to shear-based analysis in terms of complexity. The convergence is a scalar quantity whereas the shear is a complex quantity with a spin of 2. κ is a direct probe of the mass distribution which eases a multiple-probe analysis. The reconstruction process however makes it difficult to tract errors and systematics.

2.4.5 Decomposition in E- and B-modes

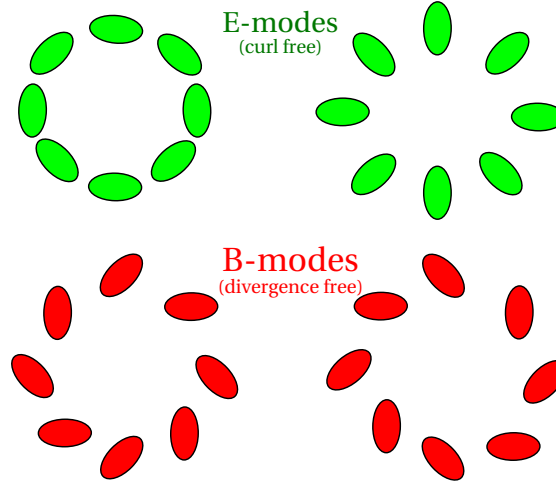


Figure 2.8 – E- and B-mode decomposition. *Top.* Lensing causes galaxies shape to be tangentially aligned to an overdensity (left) and radially for an underdensity (right). *Bottom.* Illustration of B-modes. Adapted from [Van Waerbeke and Mellier \(2003\)](#).

In the previous paragraph, we outlined the computation of the lensing effect by an extended mass distribution which produced a equivalent thin lens plane. For that equivalent plane, we derived a Jacobian matrix which, like in the thin lens approximation, is symmetric. The two shear components are not independent from each other since they both derive from the same scalar field. The shear field can be described by a symmetric and traceless 2D tensor. Such tensor field can be decomposed into a curl-free E-mode and a divergence-free B-mode (see Fig. 2.8). In the case of cosmological shear fields, the distortions will be tangential for an overdensity and radial for an underdensity, and curl-free. In an E-mode only regime, the B-modes will be negligible ([Kaiser 1992](#); [Stebbins 1996](#); [Crittenden et al. 2002](#)). There are peculiar cases from which B-modes can arise linked to the violation of the Born approximation or a clustering of sources or intrinsic alignments, but the effects have been shown to be of meagre importance (e.g. [Jain et al. 2000](#); [Schneider et al. 2002b](#); [Hirata and Seljak 2004](#); [Vale et al. 2004](#)). This property of vanishing B-modes is useful when scrutinising the measurements. If non-negligible B-modes are present (and the above possibilities for B-modes explored), they can be traced to systematic biases in the shape measurement or imperfect PSF correction.

2.4.6 Lensing by large-scale structures

Lensing by the large-scale structures causes distortions of images of distant, background, galaxies. The distortions of the image depend on the inhomogeneity of the matter density along the light path. These inhomogeneities in the mass distribution create a tidal gravitational field. Light rays propagating in this field are affected differentially, thus distorting the images of the background sources differentially too. As the cosmic web is made up of building blocks (clusters, filaments and voids) which differ in mass and extent, the effect on the light bundles

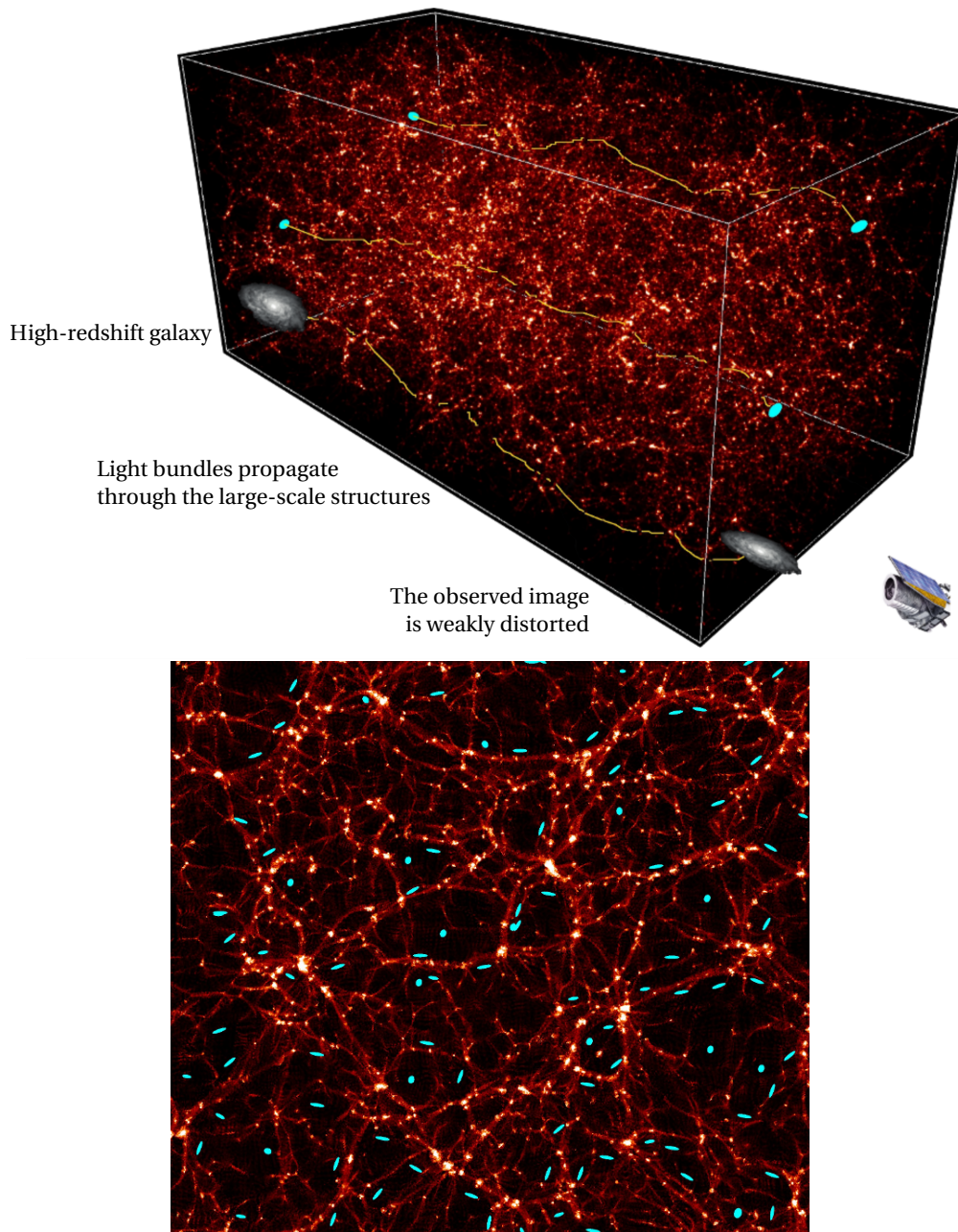


Figure 2.9 – *Top*. Illustration of lensing by large-scale structures (in red). The light bundles, in yellow, are emitted by distant galaxies (*left*), propagate through the large-scale structures and are deflected by the different components of the cosmic web before being registered on the detector of an instrument (*right*). – *Bottom*. Simulated image of distant galaxies affected by lensing by large-scale structures (red) on distant galaxies (blue). The images of the source tend to align to the large-scale structure, even though the intrinsic shape distribution is not correlated to those structures. Adapted from the Canada Hawaii France Telescope (CHFT) lensing webpage – <http://www.cfht.hawaii.edu/News/Lensing/#1C>, simulation by Colombi, S.

is different from one line-of-sight to another. Cosmic shear induces an alignment of the shape of background, high-redshift, galaxies to the tidal gravitational field. Fig. 2.9 illustrates lensing by large-scale structures. The amplitude of the distortions depends on the magnitude of the matter inhomogeneity. In the following, we will refer to “cosmic shear” rather than the mouthful lensing by large-scale structures.

The typical distortion of the shape of background galaxies is of the order of a percent, barely noticeable by a trained human eye. Of course, we do not have access to the intrinsic shape of the source galaxy. It is only through a statistical approach that one can detect a cosmic shear signal. To compute the mean shear, a few hundreds of galaxies are needed to obtain a signal-to-noise of one (Kilbinger 2015). This translate into a region of a few sq arcmin⁸ (Munshi et al. 2008). As comic shear depends on the matter inhomogeneities, it allows to fathom the dark and luminous matter distribution in the Universe and trace the evolution of the large-scale structures (Hu et al. 1998; Hu 1999). The evolution of these structures depends on dark energy, so cosmic shear studies can indirectly constrain dark energy (Benabed and Van Waerbeke 2004). Cosmic shear studies can, by extension, shine some lights on the nature of dark energy, or perhaps point towards modified gravity theories (Bartelmann and Schneider 2001; Huterer 2002). In cosmic shear analysis, a correct inference of the redshift distribution is crucial to produce unbiased cosmological parameters estimates. Two cosmological parameters are studied with this technique: the matter density Ω_m and the amplitude of mass fluctuations in the Universe σ_8 . However weak lensing measurement alone can only constrain a degenerate combination of the form $\sigma_8 \Omega_m^\alpha$, where α is an exponent ranging from 0.2 to ~ 1 . Weak lensing is also an useful probe as the degeneracies in the determination of the cosmological parameters is different than other probes (e.g. CMB). A combination of different probes yields tighter constraints on the parameters.

As eluded to in the above paragraph, knowing the redshift distribution of the sources is important to get an unbiased probe. Determination of the spectroscopic redshift (spec-z) for all sources would require an extraordinary amount of resources (especially in the light of the coming large surveys). For this reason, the distances to the most of the sources are determined by photometric redshift (photo-z) techniques. Photo-z methods are based on a photometric measurements in a limited number of bands. Photo-zs are less accurate than spec-zs, but great strides have been made in the recent past (e.g. Koo 1985; Arnouts et al. 1999; Bolzonella et al. 2000; Collister and Lahav 2004; Hildebrandt et al. 2010; Abdalla et al. 2011; Sánchez et al. 2014; Cavuoti et al. 2015; Bonnett 2015; Sadeh et al. 2016; Beck et al. 2017; Cavuoti et al. 2017). Using one of the redshift-determination methods, each galaxy can be assigned a redshift. The added information can be exploited to divide the sample of galaxies into shells (we will use both terms bin and shell). This application is called tomography and its added value is the cross-correlation of the signal between each tomographic bins. The lensing efficiency of distant bins is higher than nearby ones. The gain in information from tomography is dependent on the application. Taking advantage of knowledge of the distance

⁸Assuming a survey with a galaxy number density of thirty or forty galaxies per sq arcmin.

to each individual sources rather than the distribution of galaxies is called 3D lensing. A full 3D treatment consists in doing tomography with infinitely thin shells. 3D lensing thus probes locally the gravitational tidal field, with the goal of drafting a 3D map of the mass density (an application envisioned by [Taylor 2001](#); [Jain and Taylor 2003](#)). Tomography studies are especially in the study of dark energy properties ([Bernstein and Jain 2004](#)). There are usually joined to other probes to lift degeneracies and measure ω_0 to reasonable accuracy. [Massey et al. \(2007b\)](#) presented a three-dimensional cosmic shear analysis of the Hubble Space Telescope COSMOS survey, the largest ever optical imaging program performed in space, which allowed to constrain a combination of cosmological parameters Ω_m and σ_8 and measure the growth of structures ([Massey et al. 2007c](#)).

Gravitational lensing is achromatic, hence its effects is not limited to optical wavelengths. Weak lensing and cosmic shear signals can also be detected at other wavelengths. Radio lensing dates back a to the dawn of lensing, with the first full Einstein ring discovered ([Soucail et al. 1987](#)), putting early constraints on cosmological models ([Falco et al. 1998](#)), e.g., the study of galaxy formation and evolution lensed by galaxy clusters ([Smail et al. 1997](#)). The Square Kilometre Array telescope (SKA)⁹ will provide an unprecedented amount of high-quality radio data. Radio data has a number of advantages over optical measurements. For example, radio surveys are less affected by instrumental biases and can resolve many more sources (see [Brown et al. 2015](#)). Forecasts show that the SKA, in its full configuration, will be competitive in terms of constraints with optical counterparts ([Harrison et al. 2016](#)). The cross-correlation of radio- and optical-based maps mitigates the systematic biases and thus puts tighter constraints on cosmology ([Demetroullas and Brown 2016](#)). The CMB photons are also lensed (Sect. 2.4.11).

Numerical simulations are a key part of the cosmic shear modelling and analysis ([Bartelmann and Schneider 1992](#); [Jain and Seljak 1997](#); [Jain et al. 2000](#); [Hamana and Mellier 2001](#)). They are necessary as lensing has a non-linear relationship with the evolution of large-scale structures, thus analytical predictions are extremely difficult. The distribution of the mass acting as lenses is also non-linear, which implies a non-Gaussian convergence. Incorrect inferences about convergence can bias the error estimations of the cosmological parameters ([Semboloni et al. 2011](#)). As the surveys become larger and the cosmological prediction better, baryons cannot be neglected any more and must be included. Importantly, simulations provide systematic test opportunities and a mean to calibrate remaining biases (e.g., [Vale and White 2003](#); [Sato et al. 2009](#); [Killedar et al. 2012](#); [Harnois-Déraps and van Waerbeke 2015](#); [Mandelbaum et al. 2017](#)) and are crucial for cosmological model selection (e.g., [Pires et al. 2009b](#); [van Daalen et al. 2011](#); [Melia et al. 2015](#); [Shirasaki et al. 2015](#)).

Cosmic shear analysis provides a measurement of the combination of parameters $\sigma_8\Omega_m^\alpha$. The first publication of the above parameter combination on COSmological Evolution Survey (COSMOS) survey (A 1.64 sq deg space-based survey [Scoville et al. 2007](#)) was made by [Massey et al. \(2007c\)](#). Earlier measurements on small fields lead to a measurement in systematic tensions with CMB-based observations. The Wilkinson Microwave Anisotropy Probe (WMAP3)

⁹<http://skatelescope.org/>

experiment determined a value (Spergel et al. 2007) lower than the weak lensing studies (for a compilation, see Hettercheidt et al. 2007). It turns out that the weak lensing studies were biased because there was no knowledge of the redshift distribution of the source galaxies. Breaking the degeneracy between the parameters σ_8 and Ω_m as measured by weak lensing require to use additional information. 3D lensing makes use of the redshift of sources which breaks the degeneracy. High order statistics can also break the statistics by studying the non-Gaussianities in the data (e.g., Pires et al. 2010).

The main observable of cosmic shear is (weakly) lensed galaxy shapes. Instrumental effects, like the convolution by the PSF, perturb the shape typically more than the lensing signal. Astrophysical sources of noise must also be taken into account. We will review the different effects and their correction (or at least mitigation) in Sect. 3.2. External observables, like redshift, directly and strongly affect weak-lensing parameter estimations.

2.4.7 Contamination of the lensing signal by intrinsic alignment

One of the important assumptions made in the above is that, in the absence of lensing, the shapes of galaxies are uncorrelated to each other. Galaxies shapes correlate with the density field in which they reside (Hirata and Seljak 2004). In particular, there is a slight tendency for galaxy ellipticities to align to one another. This intrinsic alignment (IA) produce a relatively small, but measurable, effect that mimics the lensing correlation and can contaminate the signal. If the ellipticities of two galaxies are correlated, the following would be returned

$$\underbrace{\langle \varepsilon_i \varepsilon_j \rangle}_{\text{observed}} = \underbrace{\langle \gamma_i \gamma_j \rangle}_{\text{GG}} + \underbrace{\langle \varepsilon_i^s \varepsilon_j^s \rangle}_{\text{II}} + \underbrace{\langle \gamma_i \varepsilon_j^s + \varepsilon_j^s \gamma_i \rangle}_{\text{GI}}, \quad (2.59)$$

where GG is the pure lensing signal, II the pure intrinsic alignment signal and GI their correlation (Joachimimi et al. 2015). The cosmological informations are extracted from the GG term only. Galaxies align to each other causing a non-zero II term. In addition, from their alignment to the large-scale structure, no-zero GI terms arise too. In a tomographic analysis, the lensing of galaxies is correlated across redshift bins as the light paths are shared. Intrinsic alignment, however, is a local process and is therefore not correlated across redshift bins (this holds for II, but not fully for GI). If IAs are neglected, cosmological parameters will be measured with a large bias (Kirk et al. 2012). To account for them and thus mitigate the biases, intrinsic alignments must be modelled, which is currently a challenge (Kiessling et al. 2015). Redshift must also be measured with good precision and accuracy (Hoekstra and Jain 2008). The large and deep surveys conducted in the recent years (and those foreseen in the future) improve the observations of IA, especially thanks to strides in the shape measurement and PSF handling (Kirk et al. 2015). For now, contamination of the cosmological signal by IA is still present and further efforts are necessary to solve this unresolved issue (Troxel and Ishak 2015).

2.4.8 Cluster lensing

Another interesting class of lensing object is galaxy clusters. Clusters are also interesting from a cosmological point-of-view as being the most massive virialised structures in the Universe. As their crossing time is very large (of the order of the Hubble time), they can be mined to extract their formation history. This type of lensing measures the localisation and mass of the dark and luminous matter. The work on lensing is preceded by the finding of these clusters, probed by comparing the surface-brightness distribution of X-rays. X-rays surface brightness is a tracer of matter density (Clowe et al. 2000; Hoekstra et al. 2000) from which clusters can be detected and their luminous mass measured. Luminosities are compared to the mass measured by weak lensing giving the mass-to-light ratio. By studying merging cluster, one can first directly detect dark matter and second measure the cross-section of dark matter (Kahlhoefer et al. 2014). The principle of measurement is to compare the position of the gas, the stars and the dark matter halo to deduce a cross-section. Galaxy-galaxy lensing measurements provide a proxy for the position of the halo. The first detection of dark matter was made in the now famous Bullet cluster by Clowe et al. (2004, 2006). Treating mergers as giant particle colliders, Harvey et al. (2015) have measured 72 merging clusters to (i) detect dark matter at high confidence and (ii) constrain the self-interaction cross-section of dark matter.

2.4.9 Galaxy-galaxy lensing

Foreground galaxies or galaxy clusters can also act as lenses of the distant galaxies. This is called galaxy-galaxy lensing. As for the lensing by large-scale structures, the information on the original shape is inaccessible, and galaxy-galaxy lensing can only be measured statistically (Tyson et al. 1984). In practice, faint galaxies will tend to slightly align tangentially around brighter galaxies. In short, galaxy-galaxy lensing alters the distribution of the position angles of the background galaxies. The magnitude of the effect depends on the distances to both the source and lens galaxies, the mass of the lens and the angular separation of the two objects. The first detection of galaxy-galaxy lensing dates back to 1996 when two independent teams measured a few thousand galaxy shapes and detected the tangential alignment (Brainerd et al. 1996; Dell’Antonio and Tyson 1996). The lensing signal being too weak to detect from single galaxies, similar (in terms of e.g. colour, luminosity or redshift) lens galaxies are selected and stacked to improve the signal-to-noise. Galaxy-galaxy lensing of galaxy cluster members was first detected a few years later (Bernhard and Schneider 1999; Natarajan et al. 2002). Galaxy-galaxy lensing measurements provide constraints on the size and mass of the lensing galaxies, including its dark matter content. This technique measures the surface mass density contrast $\Delta\Sigma = \Sigma_{\text{cr}}\gamma_t$ which is the critical surface density multiplied by the tangential shear. The critical surface density can be computed, while the tangential shear is an observable. The surface density can be inverted to yield the density profile (Wilson et al. 2001; Sheldon et al. 2004). A large amount of information can be harvested: profiles of dark matter halos (e.g. Hoekstra et al. 2004; Coupon et al. 2013; Viola et al. 2015; Chalela et al. 2017), mass-richness relation

(Simet et al. 2017), the mass distribution (Fischer et al. 2000; Hoekstra et al. 2015; Kettula et al. 2015; Mantz et al. 2016), the mass-to-light ratio (Heymans et al. 2006; Cacciato et al. 2009; Li et al. 2009; Velander et al. 2014) and cosmological information when in a joint analysis with galaxy clustering (Leauthaud et al. 2010; More et al. 2015). The advantages of lensing to study the galaxy mass are that (i) the technique is capable of probing the halos of galaxies at very large radii and that (ii) no assumption is necessary about the dynamical state or geometry. This technique requires, in addition to a good shape measurement method, a large number of lens-galaxy pairs (a few millions is a good start) to derive statistically meaningful results and a good knowledge of the redshift of the foreground and background objects. This implies that surveys designed to measure galaxy-galaxy must go deep (i.e. expose for a long time, there are ~ 40 galaxies per square arcminute), go wide (observe a large portion of the sky), – or both! Large surveys like SDSS, Canada-France-Hawaii Telescope Legacy Survey (CFHTLS), Kilo Degree Survey (KiDS) or Dark Energy Survey (DES)¹⁰ provide a sufficient quantity of pairs and a good enough redshift determination to extract robust information from galaxy-galaxy lensing signal (Mandelbaum et al. 2006, 2013; Kuijken et al. 2015; Parker et al. 2007; Clampitt et al. 2017).

2.4.10 Lensing by voids and filaments

The large-scale structures, also called the cosmic web, comprises clusters of galaxies linked to one another by filaments and surrounded by giant empty regions subtly named voids. Interest in the lensing of background objects by voids and filaments has been picking up these past few years. Most of the mass in the Universe is located in the clusters, but most of the space is actually empty¹¹. Voids evolved from the primordial underdensities and expanded into vast empty regions, bordered by filaments. Their study allows to explore the initial cosmological conditions and its evolution as well as to probe the initial density distribution and even to fathom dark energy as it is currently dominant in these regions (Van De Weygaert and Platen 2011; Sutter et al. 2012, 2014). The idea of studying voids with lensing dates back to 1999, when Amendola et al. (1999) proposed that large spherical voids (with radii $\sim 100h^{-1}$ Mpc) only could have a sufficiently large lensing effect to be observed. Since then, new ideas blossomed to detect the lensing signal of smaller objects (e.g. Krause et al. 2013). The first measurements of gravitational lensing by voids was published not long ago (Melchior et al. 2014; Clampitt and Jain 2015; Chantavat et al. 2017), but the studies are for now focused on robustly detecting the signal and not yet constraining cosmological parameters.

Filaments that link the galaxy clusters also produce a weak lensing effect. The measurement of the mass of the filaments with lensing can indicate whether they are dark-matter dominated and thus deduce cosmological constraints (Higuchi et al. 2013). If this subfield, like lensing by voids, is still in its infancy, detections of individual filaments by weak lensing analysis have been claimed (Dietrich et al. 2012; Higuchi et al. 2015). Stacked-signal investigations were able

¹⁰We will expand on these surveys in Sect. 3.5.2, 3.5.3 and 3.6.

¹¹Or filled with “voids” if that makes any sense.

to detect filament lensing with confidence ([Clampitt et al. 2016](#)) and even deduce an estimate for the mass of the filaments ([Epps and Hudson 2017](#)).

2.4.11 CMB lensing

CMB photons are travelling, since their release, on the null geodesics through the emerging, growing and established large-scale structures and experience weak lensing. The photons are not flying along unperturbed null geodesics, but rather slightly perturbed ones. The observed image of CMB is in fact a slight sheared and magnified image of itself. Gravitational lensing does not significantly alter the CMB power spectrum on large scales, but becomes important at small scales. Weak lensing smooths out the sharp features in the CMB correlation functions ([Seljak 1996](#)) and correlates Fourier modes that should not be correlated. The temperature fluctuations are independent, thus their Fourier modes should be uncorrelated. The smoothing effect of lensing creates a slight correlation of the temperature Fourier modes. The statistical effect of lensing can however be removed from the CMB data, there is no need to measure galaxy shapes in the CMB lensing framework. In other words, the temperature and polarisation anisotropies in the CMB are Gaussian random fields, which is not the case for the effect of weak lensing. The distinct mark of lensing in the CMB data can be exploited to extract additional cosmological information ([Okamoto and Hu 2003](#)). Using the CMB lensing alone, the *Planck* collaboration could constrain the cosmological parameter combination $\sigma_8 \Omega_m^{0.25}$ to a few percent. In addition, they could detect lensing B-modes at high confidence and the lensing-integrated Sachs-Wolfe bispectrum. ([Planck Collaboration 2016b](#)). For a more in-depth review of the effect of lensing on the CMB, we refer the interested reader to [Lewis and Challinor \(2006\)](#).

2.4.12 Flexion: powerful but infeasible?

Flexion is the effect of the next-to-linear term in the derivation of the observed image distortions. This order effectively skews the images and shapes them in an arc-like fashion ([Goldberg and Natarajan 2002](#); [Bacon et al. 2006](#); [Schneider and Er 2008](#)). It measures local changes in the shear field. Flexion measurements (usually in combination to shear analysis) lead to precise determinations of the galaxy mass and dark matter halos, but are extremely challenging ([Cain et al. 2011](#); [Rowe et al. 2013](#)). It can also be used to reconstruct mass maps and break the mass-sheet degeneracy ([Rexroth et al. 2016](#)). It is poised as effective at measuring the substructures ([Irwin and Shmakova 2006](#); [Pires and Amara 2010](#); [Rexroth 2015](#)). If reliable flexion measurements can be obtained for future surveys, they will put tight constraints on the shape of dark matter halos ([Hawken and Bridle 2009](#)). The lensing community as a whole currently agrees that the path to significant detections of flexion and analysis is at best long and strenuous.

2.5 Other gravitational lensing regimes

While this thesis is mainly concerned by weak lensing and its applications, there are other important applications of gravitational lensing ranging from exoplanets detection (e.g. [Gaudi 2012](#)) to the measurement of cosmological parameters (e.g. [Bonvin et al. 2017](#)). In the following, we summarise the main applications and their reach into the scientific community.

2.5.1 Strong lensing

The strong lensing regime produces impressive effects. One can see several images of the same point-like sources or severe distortions of an extended source. The number of known strongly lensed systems is low: of the order of a few hundreds, mostly because their discovery is either serendipitous or time-consuming (e.g. [Browne et al. 2003](#); [Bolton et al. 2008](#); [Dark Energy Collaboration et al. 2016](#); [Napolitano et al. 2016](#)). Until recently, strong lenses were found by a systematic human search through survey data. The domination of human-lead searches is over (even if the presence of a human in the loop is still foreseeable and might be wanted for confirmation); we will discuss this further in chapter 6. To enter the strong lensing regime, the surface mass density of the lens must be greater than the critical surface mass density $\Sigma > \Sigma_{\text{cr}}$. A point-source background object, for example a quasar, will typically produce multiple images. An uneven number of lensed images are produced by the lens plane ([Burke 1981](#)), even if one (or more) can be merged with the image of the lens. If the source is an extended object, such as a galaxy, the lensed images will be distorted into an arc-like feature or even a ring. How the image of a source is distorted depends on the alignment of source-lens-observer and the gravitational potential of the lens. A model of the lens is required for many applications of strong lensing. Lens modelling remains a challenging endeavour, in particular because there are degeneracies (for a practical case e.g. [Treu and Koopmans 2002](#)). If additional information can be harvested, like the intrinsic size or luminosity (by e.g. observing a lensed SNIa) or the true mass of the lens (measured via, e.g. stellar kinematics), the degeneracies can be lifted.

Observing strongly lensed systems enable to probe cosmology in a number of ways. The first of these applications is cosmography, i.e. the determination of the geometry of the Universe at cosmological scales ([Blandford and Narayan 1992](#)). The determination of the Hubble constant H_0 is the objective of time-delay studies. Alike conventional optics, the multiple images of the sources form at the extrema of the arrival time surfaces. The light travel time, which depends on the speed of light of a material in conventional optics, depends here on the gravitational time delay effect, called the Shapiro time delay. The total time delay is the addition of the Shapiro and the geometric delays. The latter is simply due to a difference in the length of the light path. The time delay is inversely proportional to H_0 , as first proposed by [Refsdal \(1964\)](#). Multiply-lensed quasars represent ideal targets for time-delay measurements as they are intrinsically variable, easing the determination of the difference in arrival time. Recent results by the H_0 Lenses in COSMOGRAIL's Wellspring (H0LICOW) collaboration based on

data from the COSmological MONitoring of GRAVItational Lenses (COSMOGRAIL) survey constrained the Hubble constant to 3.8% using time delay measurements of three lenses only (Suyu et al. 2017; Wong et al. 2017; Bonvin et al. 2017).

Strong lensing allows to measure the dynamical mass of the lens within the Einstein radius, and thus to conduct light-to-mass studies, however this application is subject to mass-sheet degeneracies. This application is useful to reconstruct the mass distribution in clusters (e.g., Kneib et al. 1996). The same approach leads to the inspection of dark matter halos substructures where again there are discrepancies between the Λ CDM model and observations (they were recognised almost two decades ago, Moore et al. 1999). These investigations (mostly done in the radio band) test the cosmological models (i.e. smooth mass models or grainy or presence of dark matter substructures). Those tests are not yet definitive (Metcalf and Amara 2012; Vegetti et al. 2012; Vegetti and Vogelsberger 2014; Xu et al. 2015), but the Λ CDM model shows difficulties at small scales (Weinberg et al. 2015; Del Popolo and Le Delliou 2017). This is called the “missing satellite problem” (or “excess sub-halos problem”). Probing the dark matter profile in strong lenses also answers the question of the universality of the mass profile. A profile of the form proposed by Navarro et al. (1997, i.e. the NFW profile) is almost always predicted by cosmological simulations, however there are a number of discrepancies in the data (Treu 2010).

The natural telescope properties of strong lensing allow to study the faintest sources that would not be possible to observe otherwise (e.g., Stark et al. 2008; Livermore et al. 2015). This property may give rise to a selection bias in the luminosity of the sources called magnification bias (e.g., Negrello et al. 2010). An impressive and serendipitous result is observation of a supernovæ in a strongly lensed galaxy, which can be seen going off multiple times. It provides an opportunity to blindly test the methodology of mass model reconstruction (Kelly et al. 2015; Rodney et al. 2015; Treu et al. 2016).

Strong lensing is a useful tool of cosmology and astrophysics which can help understanding three major domains: (i) cosmography and the content of the Universe (notably Ω_m and Ω_Λ), (ii) mass distribution at small (sub-kpc to kpc) scales and especially the distribution of dark matter, and (iii) the formation and evolution of extremely distant objects. While there is no strict and sudden differentiation between weak and strong lensing, the data analysis tools and challenges are not the same. Observables in strong lensing are photometric time series or particularly distorted images, and some of the challenges are to find and to model the object involved. The large amount of data, however, is also becoming an issue for strong lensing, and currently the focus is on finding those strongly lensed systems.

2.5.2 Microlensing

Compact objects such as black holes, stars or planets can also act as lenses. When Einstein and Chwolson worked on gravitational lenses, they thought of compact objects (namely stars) as lenses. This regime of gravitational lensing gives rise to multiple images too, but their deflection angles (milli- or micro-arcsecond) are too small to be resolved. Hence the “microlensing” term ([Paczynski 1986a](#)). The observable effect is a magnification of an object. We often refer to a microlensing event, because the magnification lasts only a relatively short amount of time. The microlensing event is the fluctuation of the magnitude of the source object. The alignment of the observer, the lens and the source is dictated by their relative positions and motions. The magnification of the source object is of course stronger as the impact parameter is smaller. The main observable in microlensing is photometry, which gives a light curve. The difficulties from the point-of-view of data analysis consists in modelling light curves and extracting the physical information.

The study by [Liebes \(1964\)](#) on lensing statistics gave rise in the early 1990s to a search for microlensing events towards the Small and Large Magellanic Clouds. The idea of this study was to monitor a few million stars in the Magellanic Clouds to detect “dark halos objects”. ([Paczynski 1986b](#)). These MAssive Compact Halo Objects (MACHOs) are impossible to observe on their own, because they are too dark. Studying microlensing light curves could, indirectly, reveal the presence of MACHOs and allow for a statistical treatment to assess their importance. Three independent teams (MACHO, EROS, and OGLE) rushed to observe millions of stars and analyse the microlensing events in the dataset. The results of these studies rule out the possibility that dark matter significantly consisting of low-mass compact object ([Alcock et al. 1998](#); [Alcock et al. 2000](#); [Calchi Novati et al. 2013](#)). Compact objects must only not be stars or black holes, but also planets and moons. [Mao and Paczynski \(1991\)](#) were the first to propose to use microlensing to detect a planet orbiting a star. A standard microlensing light curve produced by a star is a single peak of luminosity. When there is a planet orbiting the lens star, a secondary peak is created. This method allows to measure the mass ratio planet to star and the angular separation at the time of the event. Microlensing can detect smaller planets, small planets at large radii and planets at very large distances from Earth. Using microlensing, a probable planet was discovered that likely reside in the Andromeda Galaxy ([An et al. 2004](#); [Ingrosso et al. 2009](#)). Objects much less massive than their host stars (by as much as a multiplicative factor of 10^{-5}) can be detected ([Bond et al. 2017](#)). The main drawback of lensing is that the event happens only once; it cannot be a follow-up and characterisation tool.

In those applications of microlensing, the distances are (mostly) galactic. However, microlensing also arise at cosmological scales. Individual stars in a lens galaxy can produce microlensing events of the source object. These individual lensing events introduce another signal, in addition to signal of the source. When measuring time delays, this additional signal is noise, as the microlensing signal is local and affect each of the image differently ([Tewes et al. 2013](#)). Microlensing can be used to extract information on size and shape of the background source and motions and mass in the lensing system (e.g. [Eigenbrod et al. 2008](#)). Those lensing masses

can also be small to intermediate-mass black holes. Microlensing can serve to constrain the mass and abundance of such intermediate-scale black holes (e.g., [Mediavilla et al. 2017](#)) The distribution of dark matter (and its graininess) on small scales can also be studied with microlensing. A paper by [Schechter and Wambsganss \(2002\)](#) studies the flux ratio of multiply imaged objects in a different light. They propose to use microlensing to study how smoothly dark matter is distributed (clumpier mass distribution increases the luminosity fluctuation produced by microlensing). This has yet to prove conclusive however ([Motta et al. 2017](#)).

3 Surveys and data analysis for weak gravitational lensing

Overview

Data processing is becoming an increasingly important task in astrophysics and cosmology. This is not only due to the increasingly complex data (e.g. gravitational wave analysis, [Messick et al. 2017](#)), but also to the minuscule signals to be extracted with great robustness (e.g. cosmic shear) and the vast amount of data to treat (the best example in that area would be the SKA, [Hall 2005](#)). Algorithms tackling data coming from the *Euclid* survey will have to deal with an avalanche of faint signals to be measured with exquisite performance.

In this chapter, we present the data aspect, always with weak lensing in mind. We start by presenting data acquisition and its inverse process: the observable to signal route. Once the data is measured with the required performance, it must be interpreted. We will overview the different weak lensing statistics derived from the measured signals that use catalogues of galaxy shapes. Then, we will turn to highlight the most interesting weak-lensing surveys and their data challenges with a particular focus on the *Euclid* mission.

3.1 The forward process

In this qualitative section, we follow photons from emission to their capture by a detector mounted on a telescope. There are many astrophysical and instrumental processes that can take place between the emission and its registration in the memory bank. We will concentrate on a few key and universal concepts, found in every weakly lensed objects. A now famous and widely used sketch to summarise this “forward process” is shown in Fig. [3.1](#).

Photons emitted by a distance galaxy form a light bundle; we decide to follow one that will be later captured by a (human) telescope. This light bundle first forms an image of the galaxy. A shape can be defined: this is the intrinsic shape. The light bundle continues its flight and is perturbed by the large mass distributions encountered along the way. Due to the tidal gravitational fields in the vicinity of those large-scale structures or galaxy clusters, the light bundle is affected differentially (Sect. [2.4](#)), thus shearing the intrinsic shape. At this point the

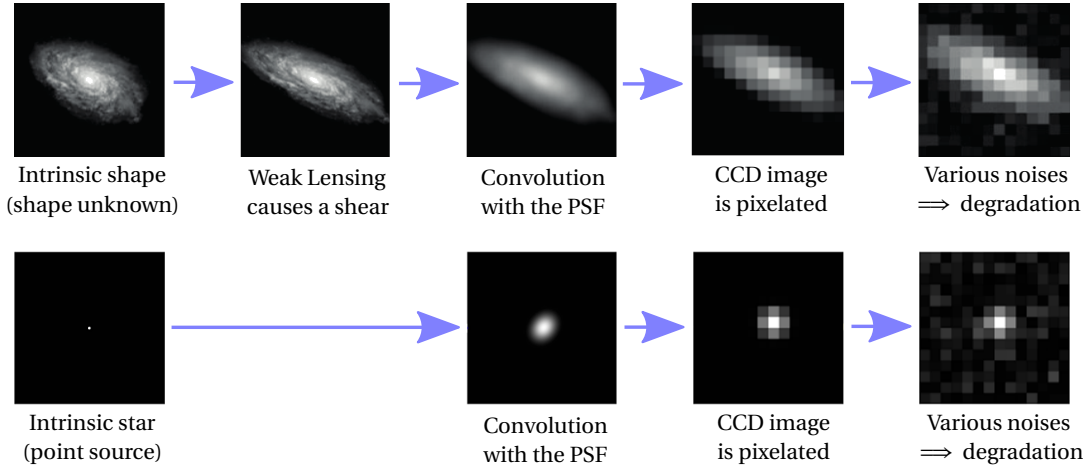


Figure 3.1 – *Top*. Forward process in weak lensing studies. The unlensed image (which is inaccessible to the observer) of a source galaxy is first sheared (here exaggerated by a factor of ~ 10), then enters the atmosphere and the optics of the telescope which smear out the image. This is the convolution by the PSF. Finally, as the photons hit the detector, the image is pixelated and the quality of the image is degraded by various astrophysical or instrumental sources of noise. *Bottom*. A star follows the forward process. The angular diameter of a star is so small that can be considered a point source. The light travels through space, *a priori*, undisturbed by gravitational lens systems. After going through the atmosphere and the optics, the light is registered on CCDs. Adapted from the GREAT08 handbook (Bridle et al. 2009).

intrinsic shape information is lost. The light bundle now approaches Earth. It may collide with the optics of a space-borne telescope or enter the atmosphere before being captured by a ground-based telescope. If it enters the atmosphere, the resolution of its image is limited by dynamic turbulences in the atmosphere. Technically, the bundle undergoes a convolution. When the bundle finally reaches the telescope, it is deformed by the diffraction pattern of the telescope. In other words, it is convolved with the PSF. The detectors, which are CCDs, consist of an array of small photo-sensitive surfaces. The image of the galaxy suffers from pixelisation at this stage. The astrophysical processes and imperfections of the CCDs generate noise. Examples of the former are zodiacal light, which creates a diffuse background, stray light from stars close to the line-of-sight and cosmic rays may ruin localised portions of the image. The CCDs themselves are a source of noise: the read-out of the array creates Gaussian noise, read-out of bright objects create leakage by charge transfer inefficiency (CTI) or the temperature of the detector which induces dark current. Other subtle effects include bright stars seen to be slightly larger than faint star, which should not be the case. This is known as the brighter-fatter effect. On top of those noises, the source photon counts is riddled with Poisson noise.

As eluded to in Sect. 2.4, PSF correction is an essential step of weak lensing analysis. PSFs are neither stable spatially nor in time. This means that the PSF changes as a function of location

on the detector. Stars have extremely small angular diameter¹. As such, they appear as point sources to a distant observer. This property implies that the image of a star after convolution by the PSF is the PSF itself.

3.2 The inverse process

Images are stored in a memory bank. Pre-processing tasks (flat-fielding, cosmic hits removal, CTI corrections, ...) have been performed such that the image is of good quality, even if still noisy. The goal of the inverse process now consists in going back upstream from the image to measure the cosmologically meaningful signal: shear. In this section, we first overview the different sophisticated shear estimators. Then, we will review the major adversarial effects that hamper the galaxy shape measurement.

3.2.1 Estimators of shapes

A major challenge in the upcoming surveys is the measuring the shape of galaxies. There is a wealth of different methods proposed in the literature. They range from direct measurement in pixel space to profile-fitting methods which include a number of sophisticated tools. In this section, we briefly discuss the main methods and trends in the field of shape measurement.

The goal of galaxy shape measurement methods is to extract the complex ellipticity ε from the light profile $I(\boldsymbol{\theta})$. There are two main families of methods. The first is based on direct estimation from the light profile, e.g. by measuring the moments of the image or decomposing the light profile onto a set of basis functions. The second broad family of methods is forward-fitting of a galaxy convolved with a model PSF on the data. One of the parameter of this fit is the ellipticity. Contrary to direct estimation methods, the treatment of the PSF in forward-fitting model is straightforward. The former however are less demanding in terms of assumptions and computation cost.

Direct shape measurements

Moment-based methods are certainly the most well-known and, also the oldest, as proposed by Kaiser et al. (1995). The so-called KSB and KSB+ method class (later further developed by Luppino and Kaiser 1997; Hoekstra et al. 1998; Kaiser 2000; Bernstein and Jarvis 2002) rely on second-order moments. The quadratic moment q_{ij} of a light profile is computed *via*

$$q_{ij} = \frac{\int w(\boldsymbol{\theta}) I(\boldsymbol{\theta}) (\theta_i - \bar{\theta}_i) (\theta_j - \bar{\theta}_j) d^2\theta}{\int w(\boldsymbol{\theta}) I(\boldsymbol{\theta}) d^2\theta}, \quad i, j \in \{1, 2\}, \quad (3.1)$$

where $w(\boldsymbol{\theta})$ is an optional weighting function designed to reduce the noise in the estimation of

¹A star of the radius of the Sun, $R = R_\odot$, at a distance of $d = 8$ kpc from a telescope has an angular diameter of $\theta = 2R/d \sim 1.1 \cdot 10^{-6}$ arcsec – or roughly a factor 10^5 smaller than the *Euclid* pixel size.

q , and $\bar{\theta}$ is the centroid of the image. A parametrisation of the ellipticity, denoted χ is given by

$$[\chi_1, \chi_2] = \left[\frac{q_{11} - q_{22}}{q_{11} + q_{22}}, \frac{2q_{12}}{q_{11} + q_{22}} \right], \quad (3.2)$$

which transforms into the complex ellipticity ϵ via

$$\epsilon = \frac{\chi}{1 + (1 - |\chi|^2)^{1/2}}. \quad (3.3)$$

There is a subtle difference in meaning between χ and ϵ (see Sect. 2.4.3), indeed

$$|\chi| = \frac{a^2 - b^2}{a^2 + b^2}, \quad |\epsilon| = \frac{a - b}{a + b}. \quad (3.4)$$

The reduced shear is derived by

$$\mathbf{g} = \frac{1}{2} \langle \chi \rangle, \quad \mathbf{g} = \langle \epsilon \rangle. \quad (3.5)$$

Bearing in mind that we averaged over $N \gg 1$ galaxies and the dispersion of ellipticities is σ_e , the lensing signal noise ratio S/N_{wl} is

$$S/N_{wl} = \frac{|g| \sqrt{N}}{\sigma_e}. \quad (3.6)$$

Some authors have added higher order terms to the KSB method class to remove biases (Okura and Futamase 2009; Melchior et al. 2011; Viola et al. 2011; Okura and Futamase 2012).

Another approach to direct shape measurement is the decomposition of the galaxy image and the PSF in a set of basis functions, called shapelets (orthogonal Gauss-Laguerre polynomials, Refregier 2003). As the PSF is modelled in the same set, the determination of the shape is a simple inverse matrix operation. A few different families of basis functions have been tested to efficiently represent the galaxies and PSFs with as few components as possible (e.g. Bernstein and Jarvis 2002; Refregier 2003; Refregier and Bacon 2003; Massey and Refregier 2005; Kuijken 2006; Nakajima and Bernstein 2007; Ngan et al. 2009; Andrae et al. 2011). There are a number of biases arising from these methods, which are due to the decomposition (e.g., Melchior et al. 2010). The decomposition of a galaxy image in the presence of noise and a PSF correlates the coefficients and lead to a biasing truncation of the coefficients. Another approach is to measure the shear in the Fourier transform space of the image (Fourier-Domain Null tests, FDNT, Bernstein 2010; Bernstein et al. 2016).

A class of methods referred to as (re)gaussianisation first convolve the galaxy profile with another kernel. The aim is to make the PSF more isotropic and tending towards a Gaussian profile, thus more closely matching an assumption of the KSB family. This assumption states that the PSF should be small and that the PSF anisotropic component is convolved with a dominant isotropic component. Regaussianisation reduces the biases in weighted second-

order moment approach (Hirata and Seljak 2003; Mandelbaum et al. 2005, 2012; Herbonnet et al. 2017). This approach is currently regarded as the most successful of the direct shape measurement family.

Forward-fitting methods

In this class of methods, one assumes a profile to which the data is fitted. Proposed by Kuijken (1999) to correct the KSB PSF anisotropic biases, it has become a successful approach to measuring shapes. Properties of the methods can be decided from the underlying mathematical formalism. A widely recognised method is the Bayesian forward-fitting *lensfit* (Miller et al. 2007; Kitching et al. 2008; Miller et al. 2013). Some method hinge on the side of simplicity to guarantee speed like *gfit* using maximum likelihood (Gentile et al. 2012) or *Im3shape* (Zuntz et al. 2013). Both methods use simple profiles (a sum of two Sérsic functions, see eq. 5.32) with the convolution with a PSF to fit the data, but involved minimisation techniques. Having only a few degrees of freedom to fit a real galaxy profile (or even a simulated one) can lead to biases that must be corrected (See Sect. 3.2.2).

Learned methods

Recently, Bernstein and Armstrong (2014) and Sheldon (2014) proposed a Bayesian method that estimates the posterior probability of the shear given the data, without having to define ellipticities (or shear) for individual galaxies. The drawback of this method is that the prior distribution must be inferred from a realistic distribution, like those obtained *via* deep images. Calibrating methods is another option. First, one measures the systematic bias of any method on realistic mock images. Then, correcting factors are applied such that the multiple and additive biases are consistent with zero. This can be iteratively by modifying the simulations or/and the method parameters (Bergé et al. 2013; Bruderer et al. 2016; Fenech Conti et al. 2017) or by perturbing the original data by a known quantity (or “metacalibration,” Huff and Mandelbaum 2017; Sheldon and Huff 2017). The DeepZot method first fits a model using maximum likelihood and then corrects for the remaining bias using neural networks (see the appendices of Kitching et al. 2012). The method learns the calibration from a set of simulations. This method was very successful in GREAT10, one of the shape measurement competition that the community organised (see Sect. 3.5.1). A mapping can be learned between the measured (and thus biased) properties of a galaxy light profile and the unbiased and deconvolved shape parameters. A look-up query in this N -dimensional table is very fast. The method, dubbed MegaLUT (Tewes et al. 2012) is the basis of the work on shape measurement in this thesis. The original idea behind MegaLUT was to classify galaxies and their PSFs depending on the measured parameters and query a look-up table to search for the ellipticity corrections to apply to the measurements. We expanded the method to include machine-learning to create an efficient and accurate mapping between measured galaxy properties and reduced shear. This method is detailed in Sect. 5.4. The strengths of this family of method are its versatility and speed. The major disadvantage is that the training set must be well controlled.

3.2.2 Adversarial effects

High-redshift galaxies, which yield the most cosmological information, are best observed in the bands i , r or z . They are usually faint, implying that their signal-to-noise is low. An image of a galaxy is typically small: of the order of the arcsec, translating into a few pixel across in the physical image. As weak lensing studies are based on the analysis of an ensemble of galaxies, each shape of galaxy must not be measured with precision, but with accuracy. Cutting-edge methods can reasonably measure galaxies down to $S/N \sim 10$. The measurement of the ensemble of shapes must be unbiased. Any remaining bias impacts the determination of the inferred cosmological constraints. For a recent review of the systematics in weak lensing analysis, see [Mandelbaum \(2017\)](#).

The biases on the shape are usually characterised by two factors: a multiplicative factor, μ , and an additive one, c . These parameters relate the observed and the true shear (e.g., [Huterer et al. 2006](#); [Heymans et al. 2006](#); [Bridle et al. 2009](#); [Mandelbaum et al. 2014](#)),

$$g_i^{\text{obs}} - g_i^{\text{true}} = \mu_i g_i^{\text{true}} + c_i, \quad i = 1, 2. \quad (3.7)$$

In general, the bias factors μ and c are dependant on the galaxy properties (and on outsider parameters like redshift biases). Ideally, $\mu_i = c_i = 0$, and $|\partial\mu/\partial p| = |\partial c/\partial p| = 0$, where p is any parameter. That is the method is unbiased and not sensitive to any parameter.

In the following, we will not discuss effects that cannot be corrected by shape measurement methods, like redshift distribution (see Sect. 2.4.6) or intrinsic alignment (see Sect. 2.4.7).

PSF correction

The largest source of distortion in weak lensing surveys is the convolution by the PSF. It can create galaxy shape correlations that are orders of magnitude larger than, e.g., the cosmic shear correlations. The additive bias c is directly linked to the PSF ellipticity and both factors depend on the size of the PSF (see [Massey et al. 2013](#), for a derivation of this result). A local and noisy representation of the PSF can be recovered from the observations of stars, but it is still required to interpolate the PSF at the galaxy coordinates and, if the passband is broad enough, with the right spectrum.

The effect of the isotropic part of the PSF is to induce biases in the measured shape of low S/N galaxies by making the galaxy images rounder (see illustration in Fig. 3.2). The anisotropic component of the PSF shears the image and thus introduces a preferred direction in the galaxy orientations. The anisotropic effects of the PSF are due to guiding errors (often referred to as jitter), optics misalignments and deformations of the mirrors. Misaligned optics produce effects that vary across the field. This is the most problematic of the above sources ([Jarvis et al. 2008](#)). On the ground, the seeing, the turbulence of the atmosphere, is an important contributor to the isotropic part.

This is a well known effect, and the community made a lot of efforts to find ingenious and

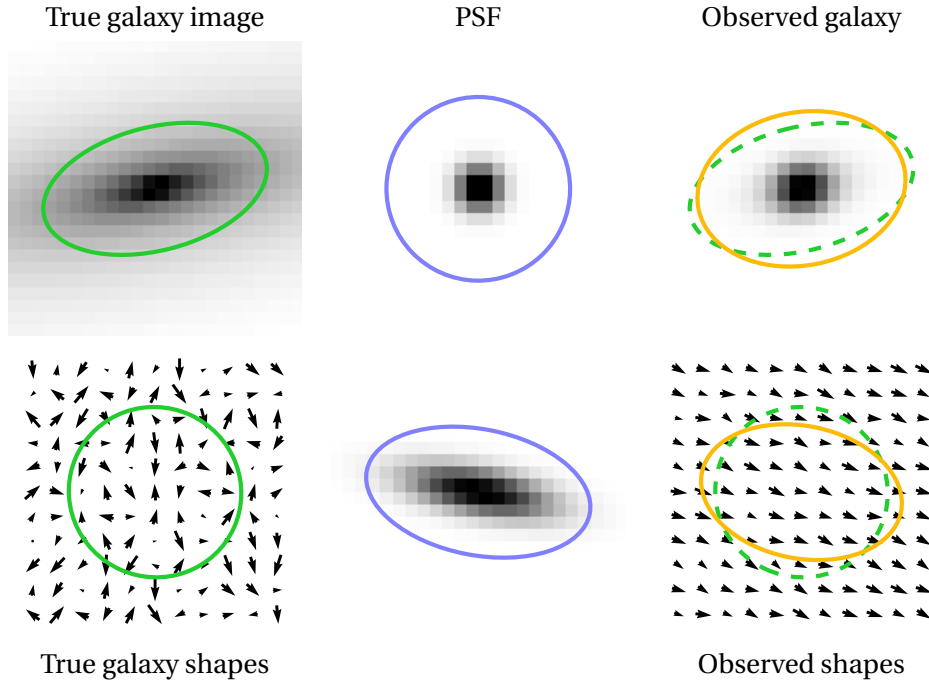


Figure 3.2 – Illustration of the effect of the PSF on galaxy shapes. Green ellipses highlight the true galaxy profile. Blue ellipses show the PSF ellipticity while yellow ellipses show the observed profiles of the galaxies convolved with the PSF. *Top.* Effect of an isotropic Gaussian PSF on a Sérsic-profile galaxy that has random ellipticities. The rounding effect, while being exaggerated, is clearly visible. Note also that this galaxy is small in comparison to the PSF and thus much more affected by it than a more extended galaxy. *Bottom.* Effect of an anisotropic Gaussian PSF on a population of Sérsic-profile galaxies. The anisotropy of the PSF in this panel is extreme. The green and yellow ellipses materialise the mean profile of the galaxies in the field. The true mean galaxy shape is round, but, due to the anisotropic PSF, a preferred direction is detected and is aligned with the PSF. The quiver plots present the ϵ_1 , ϵ_2 complex ellipticity components.

reliable methods to account for the PSF. The results of this is a wealth of methods correcting the PSF effects at different levels depending on the type of methods (e.g. [Kaiser et al. 1995, 2000](#); [Bernstein and Jarvis 2002](#); [Refregier and Bacon 2003](#); [Hirata and Seljak 2004](#); [Massey and Refregier 2005](#); [Kuijken 2006](#); [Miller et al. 2007](#); [Rowe 2010](#); [Soulez et al. 2016](#); [Lu et al. 2017](#)). Moment-based methods usually correct the ellipticity by adding PSF-related terms to the shear estimators. Those correction terms can simply be the ellipticity and size, but could be based on another parametrisation of the profile. The inclusion of the PSF in model-fitting methods is straightforward. The model of the galaxy is convolved with the PSF before its comparison to the image. This requires however to know the profile of the PSF, and not a parametric representation. Calibrations based on the measurement of simulated, and thus perfectly known, galaxies can also be used, with the same requirement to know the PSF profile.

The next generation of weak lensing survey will require exquisite PSF corrections (see e.g.,

[Paulin-Henriksson et al. 2008](#)). In this thesis, we have continued the effort to improve this correction. We discuss the work done in Sect. 5.3.

Noise bias

Shape measurement methods depend non-linearly on the presence of noise. The importance of this dependence varies according to the method. This is called noise bias ([Hall and Taylor 2017](#)). It is present for second-order moments ([Hirata and Seljak 2004](#); [Melchior and Viola 2012](#)) and maximum likelihood fitting methods ([Refregier et al. 2012](#)). Shape measurements are biased because the images are noisy. An important non-zero multiplicative factor arises. A simple toy example of noise bias is the following. Draw pixelated and noisy 2D Gaussian profile and try to fit the size parameter. The profile of the distribution and centring are both perfect and yet there is a significant bias in the value of the size parameter. The effect can affect the shear measurement up to the order of 1 to 10 percent if uncorrected (e.g. [Kacprzak et al. 2012](#); [Kacprzak et al. 2014](#)). To remove this effect, methods must be calibrated by correcting individual shape measurements or the measured shear. The pixelisation of the image can also be a sub-dominant source of biases, that translate in different biases for the pixel directions (ellipticity component 1) and the diagonals (component 2, [High et al. 2007](#)). Some methods, when based on a rigorous Bayesian treatment, avoid the noise bias, but require a large number of galaxies (e.g. [Bernstein and Armstrong 2014](#); [Viola et al. 2014](#)).

Shape noise

The variance of the observed ellipticity is the sum of the variance of the intrinsic ellipticity and the variance of the reduced shear. A large number of galaxies are required to push the variance down. The community refers to this variance of the intrinsic ellipticity as shape noise and is a large contributor to the uncertainty in weak lensing (see e.g., [Kacprzak et al. 2014](#); [Niemi et al. 2015](#)). The main driver of the size of the survey is the level of shape noise, which decreases as $N_{\text{gal}}^{-1/2}$. A common technique for shape noise cancellation in simulations is to draw several times the same galaxy, with a different position angle.

Another issue linked to shape noise is the intrinsic ellipticity distribution. Some methods depend on the knowledge of intrinsic shape distribution as the prior for ellipticity ([Viola 2014](#)). As in all Bayesian analyses, the prior distribution is an important quantity. Galaxy ellipticities distributions were the subject of a number of studies, which provide the basis for those priors (e.g. [Lambas et al. 1992](#); [Brainerd et al. 1996](#); [Ebbels et al. 1999](#); [Bernstein and Jarvis 2002](#); [Rodriguez and Padilla 2013](#); [D'Eugenio et al. 2015](#); [Rodriguez et al. 2016](#)). There is a dependence of the shape on the redshift of the galaxies. Other parameters, like brightness and colour are important. Probing galaxy shapes to the depth required to cover the galaxies imaged by surveys like *Euclid* is a challenging task. Ellipticity distributions will have to be estimated using the weak lensing data itself. In some surveys, like *Euclid*, there will be deep observations. These deep exposures will be the basis of the ellipticities distributions. When building the distributions, one should observe in very different line-of-sights to avoid being corrupted by the lensing correlation on small scales.

Model bias

Model-fitting methods (using maximum-likelihood and Bayesian formalisms alike) are affected by yet another bias. Simply put, trying to fit a simple model is not sufficient to capture the complexity of real light profiles. Modern (model-fitting) methods represent galaxies using two components: a bulge and a disk. There can be significant differences in number of free parameters (e.g. is the centroid from the bulge the same as the disk's?). Both the galaxies and the PSFs must be well modelled to remove model bias (Lewis 2009; Voigt and Bridle 2010; Melchior et al. 2010).

Object blending

What is captured on an image is a 2D projection of a 3D distribution of sources. The light of two objects can therefore be blended if they are too close-by. As the image registers fainter objects, systematic blending errors increase. Blending is a relatively new worry, as previous surveys were limited by other systematic errors (like shape noise). If detected, the blended objects can be rejected. The effect on weak lensing surveys is to reduce the number density of galaxies and thus impacting the S/N (e.g., Chang et al. 2013; Miller et al. 2013). In addition, flagging and filtering of blended or close-by neighbours biases the redshift distribution and correlates background and foreground sources. If no correction is applied, this leads to multiplicative biases in the cosmological parameters of the order of the percent (Hartlap et al. 2011). There will be problematic blends that cannot be rejected or treated, i.e. objects change shape and it cannot be detected. Blending objects can take the form of a fainter, undetectable, object subtly influencing the shape measurement method. Blending could not be detected when two objects are so close on the sky that their blend is ambiguous. The distribution of ellipticities of ambiguously blended objects differ from unblended galaxies, with a significantly offset for the mean and a large variance. This increases the shape noise bias, especially from space because of the smaller PSFs (Dawson et al. 2016).

Broad-band observations and chromatic effects

To increase the number of observed galaxies in typical cosmic shear surveys, and instead of exposing for a longer time, broad-band filters are used. The resulting image is the integration of the light within a large wavelength range. The colour of galaxies varies spatially across their profile. They depend of the spectral energy distribution (SED) (e.g., Barbera et al. 2005; Welikala and Kneib 2012). These colour gradients across the profile block a clear definition of ellipticities as values may vary with colour. The mitigation of this potential bias in future cosmic shear surveys is hampered by the fact that the PSF is also integrated in the same large wavelength range. There is thus a chromatic dependency. The exact profile of the PSF is determined by the spectral type of the stars (see Sect. 5.2). Cypriano et al. (2010) proposed a correction scheme based on the generation of a PSF whose SED matches the SED of the galaxy. This is done by weighting PSFs of different spectral types. Eriksen and Hoekstra (2017) present a scheme that estimates the effective PSF size, that is the size of the SED-weighted PSF, from broad-band images using machine learning. This correction is sufficient when there is no

colour gradient, but when there are gradients, biases arise which will have to be calibrated out (Voigt et al. 2012; Semboloni et al. 2013).

Multi-exposures

In modern weak lensing surveys, the observational strategy is to combine multiple exposures. Multiple exposures allow to increase the detection limit and sample different region of the CCDs. Finally, they are important as there are gaps between the detectors. Gaps are filled by multiple exposures of the same object with some translation, or dither, between the frames. There are basically two options to treat multiple exposures: (i) co-add them and analysis them as one image or (ii) analyse them separately. Co-addition is not a straightforward task. The images must be rescaled, oriented and interpolated such that they astrometrically match. This leads to a potentially complicated PSF, and thus its correction is challenging. The advantage, of course, is a better S/N image.

3.2.3 Computational challenges

The measurement of weak lensing signal requires a large number of lensed galaxies. This high number of galaxies is required because the intrinsic, unlensed, image is physically inaccessible and several sources of errors corrupt the data. Moreover, to be able to perform cosmological model selection and method calibration, numerous highly demanding simulations must be performed. A weak lensing analysis is therefore not trivial from a computational point-of-view. The computational challenge is three fold: (i) *Euclid* surveys and alike will harvest petabytes of data each, the challenge is thus storing the very large amount of collected data, (ii) data treatment (from pre-processing to shape measurement and computation of the statistics) and (iii) simulations. The second issue, central processing unit (CPU) or graphical processing unit (GPU) computation time demands, is at the core of this thesis. The different shape measurement methods do not require the same amount of time to measure a galaxy stamp. Expansive operations (such as interpolation, denoising, iterative processes,...) are typical of some classes of methods. These must be kept to a minimum and still produce robust and reliable shape catalogues. These computational challenges lead to the development of intricate networks of processing functions and the creation of multiple data centres (see Dubath et al. 2016, for a discussion about the *Euclid* data processing plan and the Swiss Data Center organisation). They are responsible to develop and roll out runtime-optimised algorithms on the basis of the shape measurement methods cited in Sect. 3.2.1, which were usually coded by astronomers. To be included in a pipeline of any mission (be it in space or on the ground), the algorithms must meet the requirements placed on their output and validated. Weak lensing surveys, and most of astrophysics in general, have really entered the world of big data. Some members of the community now speculate that the pressure on the computational infrastructure will outweigh the observation time pressure in the future and that data specialists and computer scientists will become increasingly important (see reflections in, e.g., STScI Science Definition Team 2016).

3.3 Weak lensing statistics

In this section, we turn to the different possible statistics used to infer information from shape catalogues. We look at two-point statistics (also known as second-order statistics) and higher-order statistics that improve the constraints on the cosmological parameters.

3.3.1 Second-order statistics

Extracting information from shear catalogues, in particular cosmological information from the large-scale structures, is not a straightforward task. The most widely used statistics are second-orders statistics, which include two-point correlation function and power spectrum.

Two-point correlation function

Even when the deformations of the image of a distant source by the lensing of the large-scale structure cannot be predicted because the distribution of matter is unknown, distortions of objects located nearby should be similar. This angular correlation is expected to be higher as the separation of the objects is smaller. This property can be exploited to extract information by inspecting the average angular correlation. Angular correlations functions are the simplest tools to derive meaningful statistics. The shear two-point correlation function (2PCF) is the average of the multiplication of the ellipticities of galaxy pairs (Miralda-Escude 1991; Schneider et al. 2002a). Shear components are decomposed into a tangential γ_t and cross component γ_\times ,

$$\gamma_t = -\Re(\gamma \exp(-2i\varphi)), \quad \gamma_\times = -\Im(\gamma \exp(-2i\varphi)), \quad (3.8)$$

with respect to a given direction on the sky θ and angle φ . Tangential (radial) alignments around overdensities (underdensity) result in a positive (negative) γ_t . From these components, three two-points correlators can be built: $\langle \gamma_t \gamma_t \rangle$, $\langle \gamma_\times \gamma_\times \rangle$ and the vanishing $\langle \gamma_t \gamma_\times \rangle$. The two remaining correlators are the basis for the 2PCF ξ_\pm ,

$$\begin{aligned} \xi_+(\theta) &= \langle \gamma_t \gamma_t \rangle(\theta) + \langle \gamma_\times \gamma_\times \rangle(\theta) \\ \xi_-(\theta) &= \langle \gamma_t \gamma_t \rangle(\theta) - \langle \gamma_\times \gamma_\times \rangle(\theta). \end{aligned} \quad (3.9)$$

In practice, the 2PCF can be averaged as follow. Select pairs of galaxies (i, j) whose angular separation is $|\theta_i - \theta_j| \approx \theta$. Then, average the galaxy pairs' ellipticities taking into account a weighting scheme w which translates the shape measurement uncertainty. In other terms,

$$\hat{\xi}_\pm = \frac{\sum_{ij} w_i w_j (\varepsilon_{t,i} \varepsilon_{t,j} \pm \varepsilon_{\times,i} \varepsilon_{\times,j})}{\sum_{ij} w_i w_j}. \quad (3.10)$$

This estimator is of course sensitive to intrinsic alignment (Sect. 2.4.7). This can be seen by injecting the weak lensing regime relation between ellipticity and shear ($\varepsilon \approx \varepsilon^s + \gamma$), into the

above equation. In surveys, there are many masked areas (e.g., in the vicinity of bright or saturated stars). The 2PCF allows for a simple treatment of those regions.

Lensing power spectrum

The lensing power spectrum P_κ is the two-point correlation function of the Fourier transform of the convergence $\tilde{\kappa}(\ell)$, where ℓ is the wave vector. We assume the flat-sky approximation (Blandford et al. 1991; Miralda-Escude 1991; Kaiser 1992) to compute the power spectrum of the convergence (see LoVerde and Afshordi 2008, for a treatment of the power spectrum without the flat-sky approximation). The power spectrum of the convergence can be directly linked to cosmological parameters and to the density power spectrum. Being a two-point statistics, the 3D information gets lost or at least diluted, which makes the lensing power spectrum less sensitive to cosmology than other probes, e.g. the CMB anisotropy power spectrum. The lensing power spectrum is most sensitive to the degenerate parameter $\sigma_8 \Omega_m^\alpha$, and to some extent other parameters like to the zeroth order of the dark energy parametrisation w_0 and to h . This degeneracy in the power spectrum weakens the utility of the tool, but it can be combined with other information to lift it. The power spectrum of the shear is equal to the power spectrum of the convergence (this holds also in real space). There are many ways to estimate the power spectrum, even directly from observed ellipticities, but the most robust estimate is to take the 2PCF of the shear in real space, i.e. what was described in the previous paragraph.

Further two-point statistics

Other second-order statistics can be derived from the 2PCF, they are widely used in cosmic shear surveys. A common motivation to use another set of statistics is to separate E- and B-modes (see Sect. 2.4.5). We remind the reader that B-modes should essentially be null and thus serve as a test of the systematics in the measurement.

The aperture-mass dispersion is denoted as $\langle M_{\text{ap}}^2 \rangle(\theta)$. The name comes from relation between this correlation and the over-density. It is based on the concept of aperture mass, i.e. the integral over the tangential shear around the centre, weighted by a function Q_θ , or equivalently, the integral of the convergence and its weighting function U_θ , (Kaiser and Squires 1993; Schneider 1996; Schneider et al. 1998; Crittenden et al. 2002). By construction, the mass aperture dispersion is sensitive to E-modes. By replacing the tangential shear by the cross term, one can be sensitive to B-modes. The variance of the mass aperture is the two-point statistics. It can be interpreted to describe the strength of the lensing between different line-of-sights. This statistics is sensitive to masking issues, and to mitigate them, one can, e.g., inpaint data (Pires et al. 2009a), or express the mass aperture dispersion in terms of the 2PCF, which is almost insensitive to masking issues. However, like other two-point statistics, the mass aperture dispersion suffers from leakages, thus undergoing E-/B-mode mixing. One can either fit theoretical predictions to the data to extrapolate the statistics with the risk of biasing the results or simply ignoring the scales at which the leakage occurs, but to the cost of losing information. The mixing of the modes can be mitigated by the choice of filtering functions

(e.g., [Schneider et al. 2010](#); [Fu and Kilbinger 2010](#); [Eifler et al. 2010](#)).

3.3.2 Higher order statistics

The convergence power spectrum is only sensitive to the Gaussian part of the mass distribution for the large-scale structures. The non-Gaussianities, which arise particularly on the small and to some extent at intermediate scales, are not captured. They degrade the quality of the cosmic shear measurement (e.g., [Takada and Jain 2004](#); [Hilbert et al. 2012](#)). To measure them, higher-order statistics should be used.

Lensing bispectrum

The most natural extension is the bispectrum which measures the three-point correlation of the convergence in Fourier space. This corresponds, in real space, to the extension of the 2PCF to the three-point correlation function (3PCF). [Fu et al. \(2014\)](#) showed that a combined analysis of the 2PCF and 3PCF improves constraints on cosmological parameters as they do not have the same dependency the cosmological parameters and allow to lift the $\Omega_m - \sigma_8$ degeneracy. Other statistics exist, like the aperture mass skewness $\langle M_{\text{ap}}^3 \rangle$ ([Pen et al. 2003](#); [Jarvis et al. 2004](#)).

Peak counts

Regions of overdensities will create high convergence (and also important tangential shear). Identifying these overdensities or “peaks” in a convergence map or aperture mass map is a powerful cosmological probe as it depend on the halo number and mass functions, which in turn, depend on the cosmology (e.g., [Kaiser 1986](#); [Kratovichil et al. 2010](#)). They probe the non-Gaussianities in the large-scale structures. There are however complications: the peaks are not directly linked to the halo mass profile. This is due to the presence of noise and the projection onto a 2D image (e.g., [Fan et al. 2010](#)). The former can produce false peak detections, and biases the measurement of peaks of low mass. Small masses distributed along the line-of-sight can reproduce the effect of a single large mass.

A major advantage of peak statistics is that it is a non-linear probe, but constructed from the (observed) shear only, and thus is less sensitive to systematics than other non-linear statistics. It can be jointly used with other probes to probe degeneracies on cosmological parameters (e.g. [Dietrich and Hartlap 2010](#)). Obtaining simulated theoretical peak statistics is complicated, even if significant progress has been made in recent years (see e.g., [Lin and Kilbinger 2015](#); [Peel et al. 2017](#)).

3.4 Weak lensing analysis pipeline

A weak lensing analysis pipeline includes a series of modules (also referred to as “blocks”). Each of these modules can be extremely sophisticated, or pretty straightforward. A module might be very different for different sources of data. After basic pre-processing tasks (flat-fielding, de-biasing, cosmic rays removing, CTIs corrections, ...), the first of such module performs object detection and classification into stars, galaxies and others.

Then, a PSF determination and reconstruction block runs on the stellar objects to determine the convolution kernel at the position of the galaxies (See Fig. 3.3). For example, in the *Euclid* mission pipeline, this PSF block will have to include a colour dependence to the PSF interpolation. The next module is the galaxy shape measurement method to provide, depending on the method, the shear or ellipticities of individual galaxies or their distribution for an ensemble of galaxies. This module should treat the galaxy images in their full complexity, and with possible multi-exposures. Weak lensing statistics are computed in the following block. The goal here is to turn shape catalogues into cosmologically meaningful information. At this point, the purely data analysis ends, and we enter the realm of interpretation of the results. The next module is simulations and their statistics, which are compared to the data-based statistics. Goodness of fit analyses are conducted to perform model selection. Finally, a

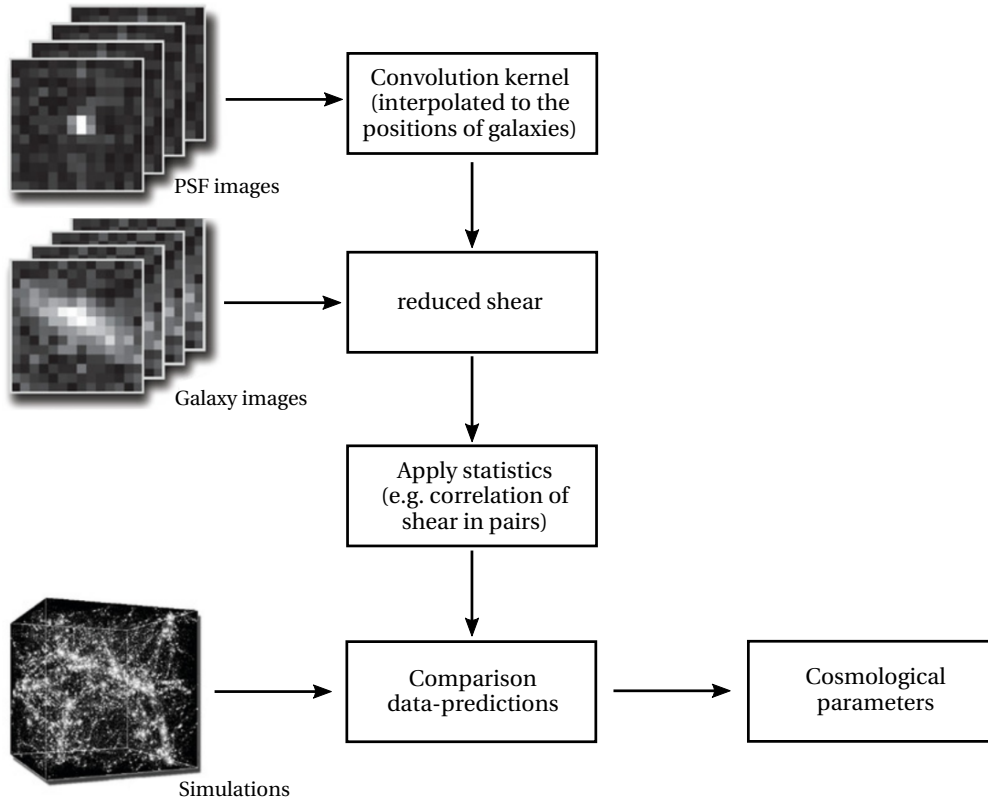


Figure 3.3 – Sketch of a typical weak lensing analysis pipeline. Adapted from [Bridle et al. \(2009\)](#).

cosmological parameters box allows to draw conclusions and constraints on the parameters.

The details of the typical pipeline described in the above is of course highly dependent on the survey, but those are common steps that must be made in all of them. Each of the block is important and can bias the end results, i.e. the constraints on the cosmological parameters.

3.5 Weak lensing surveys

We have examined the path of light bundles from emission to registration on a detector. We also described the difficulties of the inverse process, i.e. from the galaxy images to a meaningful measurement of the shape. We highlighted a few common practices to turn shape catalogues into cosmological constraints. This section is dedicated to the data gathering surveys. We will first make an important detour: since 2004 and the STEP1 Challenge (Heymans et al. 2006), the community has made important progress in designing shape measurement algorithms. The credit of this intense effort goes, of course to the researchers, but also to the community-wide challenges which concentrated the resources to solving increasingly complex issues. Then, we will overview the major (cosmic) shear weak lensing surveys, observed from the ground and orbit. We dedicate section (Sect. 3.6) to the *Euclid* mission as it is the foundation of this thesis.

3.5.1 Challenges

The STEP initiative

The first instance of the competition was set up in 2004, and published its first results in Heymans et al. (2006). The Shear TEsting Programme (STEP)², aimed at providing a set of blind and realistic simulations to any team interested in participating. The term blind in this context means that the simulations parameters were known only to a restricted number of organisers and were not communicated to the contestants. At the heart of the first edition was the wish to improve the credibility and recognition of weak lensing as a cosmological probe. STEP was launched with the aim to improve all aspects of the weak lensing technique. First and foremost, improve the accuracy and reliability of shape measurement. This was done (and is still done so today in comparable challenges) by scrutinising shape measurement algorithms and testing their sensitivity to different parameters. Another aspect is sharing technical expertise and theoretical understanding. The authors of this STEP 1 challenge tried to provide a realistic population of galaxies while making a few simplifying assumptions, sheared by a few different values, constant across images. The images were simulated observations from the ground. The images were thus convolved with simple Gaussian and Gaussian-with-spikes PSFs. The data consisted in survey-like images, mixing stars and galaxies. This was designed to be an easy test of the methods as only smooth and elliptical profiles were used. The results of the challenge showed that the methods of that time were capable of measuring relatively simple smooth galaxy profile with an accuracy of the order of the percent and removing the

²http://www.roe.ac.uk/~heyman/step/cosmic_shear_test.html

anisotropy of the PSF with relatively good performance. However, the KSB/KSB+ family of methods showed a large variance in its performance, i.e. a dependence on the implementation details was highlighted.

The following STEP programs included more sophistication in the simulations. In the STEP 2 challenge, the focus was on more complicated galaxy morphologies (Massey et al. 2007a). The number of galaxies was also increased to compute more robust statistics. The sensitivity of methods to a number of parameters was estimated, in particular for galaxy morphology, size, PSF parameters and shear direction with respect to the pixel grid. The increased complexity of the STEP 2 data, however, did not prevent the methods from improving.

The third STEP challenge, dubbed STEP space reproduced the same format, but for space data with complicated PSFs. The methods did perform as well in STEP space than in STEP 2. At that point, it was becoming clear that shape measurement was extremely challenging. It was then decided to take a step back and scale the complexity down to understand the sensitivity of the methods in a simpler environment in a fourth and last challenge dubbed STEP data.

The GREAT initiative

A follow-up initiative, the GRavitational lEensing Accuracy Testing (GREAT), continued this simplification process. Complications not directly linked to shape measurement (such as blending) were not taken into account. The image format also changed: instead of natural-looking images postage stamps of galaxies were aligned on a grid, with an increased amount of data, to better assess the performance. The first of these challenges, the GREAT08³ began in late 2008 and came to an end six months later. This first edition also blinded its data, except for a training dataset that could be used for calibration. There were two branches in the challenge: low and realistic noise. Further major simplifications are a constant shear and PSF across fields (Bridle et al. 2009). With the advent of the GREAT challenges, the community became less oriented towards astrophysics and more towards statistics. Contributors with non-cosmological background were thus encouraged to participate. The discovery of model bias was one of the results of this first instance. New techniques were introduced, including stacking e.g. galaxy images or two-point statistics to increase S/N. It showed also that different methods were performing better in certain conditions. Those good performance conditions varied as a function of the method. The objectives of the surveys at the time were to push the multiplicative bias to 2% and the additive bias to 0.1%. GREAT08 demonstrated that it was possible on intermediate complexity dataset (Bridle et al. 2010).

The next edition, GREAT10⁴, was launched in late 2010 and ran for 9 months. For this edition, both the shear and the PSF were spatially variable, in an effort to propose more realistic simulations (Kitching et al. 2011). The main goal of this challenge was to reconstruct either the shear for each galaxy, the shear power spectrum or correlation function. In addition to this galaxy shape measurement challenge, a “star challenge” was put in place. This star challenge

³<http://great08challenge.info/>

⁴<http://great10.pbworks.com/>

consisted in interpolating the PSF at non-star positions. GREAT10 tested more aspects of the typical weak lensing pipeline by testing the PSF determination and reconstruction block. As in each edition of the challenge, the data volume increased significantly. The symbolic volume of the terabyte was reached, for over 50 millions galaxy stamps. A symbol of a further departure from astrophysics to enter the big data and the statistics worlds. The methods did perform even better than in the GREAT08 challenge and reached sub-percent multiplicative bias accuracy. The demanding *Euclid* requirements ($\mu \lesssim 2 \cdot 10^{-3}$, $c \lesssim 1 \cdot 10^{-4}$), however, were not reached (Kitching et al. 2012, 2013).

The latest variation of these community-wide event is the GREAT3 challenge⁵, which ran for six months between October 2013 and April 2014 and included both constant and variable shear and PSF branches. GREAT3 also provided the contestants with two survey experiments: ground- and space-based. The galaxy images were either simulated or real. The real images were obtained by re-convolving *Hubble* Space Telescope (HST) COSMOS images with an appropriate kernel. Multi-epoch treatment was also tested in dedicated branches of the challenge. Finally, full complexity branches (i.e. including variable PSFs, multi-epochs and real galaxies) were proposed in their space/ground and constant/variable shear options (Mandelbaum et al. 2015). There were in total twenty branches⁶ in this challenge, which made it extremely large from a data standpoint. The six months period was extremely short to answer all of the hardships of this many-fold challenge. All in all, the goal was to test shear estimation in a realistic distribution of size, S/N, ellipticity, morphology with reasonably complicated PSFs. An important result of GREAT3 is that there are a range of different approaches that could reach the demanding requirements of the future so-called stage IV surveys. The inclusion of real galaxy profiles provoked systematic biases of the order of the percent, a significant bias, even if still sub-dominant with respect to noise bias. This challenge also showed that estimating a reliable metric Q of the performance is hard, as systematic biases on Q can arise (Mandelbaum et al. 2014).

The future of shape measurement challenges

The last challenge ran in 2013-2014. The frequency of such events until that point was one competition every two-three years. To the best of the author's knowledge, there is currently no plan for a next community-wide challenge in the near future. This is not due to an over-confidence from lensers⁷. The community members are now involved in different large-scale surveys that significantly differ in their design and their requirements. We already discussed the colour dependence of the shape measurement of *Euclid*, which is a negligible effects for other ground-based surveys. They, however, will have to deal images of lower resolution. The priorities of the lensers are now diverging as we enter an era in which data is plentiful and of high quality. There is a trend to replace community-wide challenges, with survey-specific

⁵www.great3challenge.info

⁶Five experiments: simulated galaxies “control” with constant PSF, real galaxies with constant PSF, simulated galaxies with multi-epoch and constant PSFs, simulated galaxies but variable PSF and full complexity. Each of these experiments contained four options: ground-/space-based and constant or variable shear.

⁷Researchers in gravitational lensing are sometimes dubbed “lensers”.

challenges, to answer dedicated questions. An example of such challenge is the continuation of the GREAT series in the radio domain⁸ (Harrison et al. 2016). Weak lensing the radio domain is currently a burgeoning field as the SKA is entering its first data acquisition phase soon.

3.5.2 Ground-based surveys

In the following, we mention the most important weak lensing surveys. Surveys are often categorised into different stages, depending on their ambitions. The first, pioneering surveys were Stage I in the early 2000s. After the development of weak lensing methods and expanding survey areas, surveys matured into Stage II. We are currently in Stage III and will be moving by the end of the decade into Stage IV surveys.

The SDSS⁹ was the first large-scale survey to use CCDs. It began operation at the turn of century, and observed an impressive 20 000 deg² over the years, in the *ugriz* filters. The 2.5-meter telescope located in New Mexico, USA has been adapted over the years to different uses. Galaxy-galaxy studies have been conducted, and weak lensing analysis were run on some of the deepest SDSS images, in the so-called stripe-82.

The CFHTLS¹⁰ was completed in early 2009 and observed about 140 deg² with a 340 megapixels camera in the Sloane filters (i.e. u', g', r', i', z'). In 2012, building on the CFHTLS (and on its data), the CFHT Lensing Survey (CFHTLens)¹¹ probed 154 deg² for a number density of ~ 17 resolved galaxies per arcmin² and a median redshift of 0.75. The faintest galaxies used in the analysis are $i' \sim 24.5$. The goal of CFHTLS was to probe the dark energy EOS w to 5-10%. The results of this survey lead to cosmological constraints derived from 2D, tomographic and 3D shear analyses (e.g. Kitching et al. 2014).

KiDS¹² is imaging 1 500 deg² in four filters (u, g, r, i) with its 300 megapixels, 1 deg² field-of-view OmegaCAM instrument. It started observing in late 2011. It is smaller in survey area than SDSS, but deeper, by about 2 magnitudes (24-25th magnitude depending on the filter), and has better image quality, with a well-behaved PSF across the field-of-view. KiDS can be seen as a sequel to the CFHTLens survey, but with data taken from the VLT Survey Telescope in Paranal, Chile. Early results show a tension in the degenerate parameter S_8 with the Planck data, but it might be too early to call for new physics just yet (Hildebrandt et al. 2017; Efstathiou and Lemos 2017).

The Hyper Suprime-Cam survey (HSC)¹³ is another survey currently observing. Mounted on the 8.4-meter Subaru telescope, HSC has a large 1.77 deg² field-of-view, with ~ 870 megapixels on Mauna Kea in Hawaii. HSC is a relatively small area on the sky (1 400 deg²), but provides

⁸<http://radiogreat.jb.man.ac.uk/>

⁹<http://www.sdss.org>

¹⁰<http://www.cfht.hawaii.edu/Science/CFHTLS/>

¹¹<http://www.cfhtlens.org>

¹²<http://kids.strw.leidenuniv.nl/>

¹³<http://hsc.mtk.nao.ac.jp/>

deep and high-resolution imaging, thus putting the focus on high-redshift galaxies. The images are taken in five filters (g, r, i, z, y) with a limiting magnitude of 24.5-26. Science operations began in 2013. The footprint overlaps with SDSS', thus taking advantage of data already acquired. This survey serves as a path finder for large-scale surveys.

The last major current survey currently in operation is the DES¹⁴, which uses the 4-meter Blanco Telescope at the Tololo CTIO observatory in Chile (see e.g., [Dark Energy Collaboration et al. 2016](#)). The optical imager covers 2.7 deg² with 61 (working) CCDs for 570 megapixels and five filters (g, r, i, z, Y). DES will image about 5 000 deg² in the South Galactic hemisphere, with ten visits per field (for the g, r, i, z bands). DES started surveying the sky in August 2013, observing for ~ 5 years. The limiting sensitivity of the detector is $S/N = 10$ for magnitude 24.1 objects in the i band. Small patches of sky are observed more frequently and are optimised for time-domain applications. The data obtained in these small patches is deeper, allowing for, e.g., shape measurement methods calibrations.

The Large Synoptic Survey Telescope (LSST)¹⁵ operations are planned over 10 years starting in 2022 to image 18 000 deg², and will generate hundreds of petabytes. LSST aims at answering multiple questions about astrophysics, some being the nature of dark matter and dark energy. LSST has a 8.4-meter telescope located in Cerro Pachón, Chile with an enormous small-car-sized camera of 3.5-deg² field-of-view with 3.2 gigapixels! The survey will be optimised for weak lensing survey. Tight constraints will be placed on cosmological parameters and the growth of cosmological structures, and in turn on what is dark energy (for more information, see [LSST Science Collaboration et al. 2009](#)).

The last survey we highlight is the SKA¹⁶. This radio interferometer telescope will raise many challenges, one of them will be weak lensing in the radio domain. The first phase of scientific operations, starting in mid 2020s will cover 5 000 deg². This will provide constraints of the order of 5% on cosmological parameters and 50% on w . The next phase will increase its sensitivity by a factor of ten, which will place SKA's derived cosmology at the same level as the most advanced, stage IV, optical surveys.

3.5.3 Space-based surveys

COSMOS¹⁷ observed the sky using the HST and combined the visible image with a mapping of the large-scale structures to measure the cosmic shear as a function of redshift. This lead to the first 3D map of the large-scale structures ([Massey et al. 2007c](#)). With 1.7 deg² and 2 millions galaxies, the observed area remains small, but the ACS instrument of the HST has a small field-of-view, and, thus, imaging a large region of space is time-consuming. An extensive sky survey would require dedicated space missions.

¹⁴<https://www.darkenergysurvey.org/>

¹⁵<http://www.lsst.org>

¹⁶<http://www.skatelescope.org>

¹⁷<http://cosmos.astro.caltech.edu/>

To dramatically expand the survey area, the *Euclid* mission was proposed to ESA and accepted. We dedicate Sect. 3.6 to the description of this ambitious mission.

Wide Field Infrared Space Telescope (WFIRST)¹⁸ has a peculiar history. This 2.4-meter telescope, which has a HST-like optics quality, was offered to NASA as it was discarded by another U.S. government department. This mission is designed to be launched in a high orbit, most likely around the Sun-Earth Lagrange Point 2, with a launch date in the mid 2020s. Poised to answer scientific questions about, in parts, dark matter and energy, it will be complementary to ESA's *Euclid*. There will be two instruments onboard, a Wide-Field Instrument (WFI) comprising of a ~ 290 megapixels camera with a 0.28 deg^2 operating in near-IR in multiple bands and a powerful coronagraph, to directly image exoplanets. WFIRST will observe $2\,000 \text{ deg}^2$ to ~ 2.5 magnitude shallower than *Euclid* to provide ~ 380 million galaxy images (Spergel et al. 2015). Also relevant to gravitational lensing, WFIRST will work in the time-domain to spot microlensing events by observing the same patches of sky at a high cadence.

3.6 The *Euclid* mission

The *Euclid* mission¹⁹ is the *raison d'être* of this thesis. To achieve the science goals, extremely ambitious requirements are set. Some of these requirements, which push the understanding of the telescope and observed physical phenomena to new frontiers, open, on their own, new investigations into effects previously unknown or deemed completely negligible. In this section, we will highlight some of the most interesting and relevant features of this space mission. The information provided in this section comes mainly from Laureijs et al. (2011).

3.6.1 Scientific goals

The *Euclid* mission is a survey mission to probe cosmology and fathom the nature of dark energy, dark matter and gravity. The main drivers of the satellite are two powerful probes: weak lensing and BAO. In addition to those main probes, more information can be derived using *Euclid* data such as galaxy clustering, redshift-space distortions (an effect arising in redshift-space, due to the peculiar velocities of the galaxies) and integrated Sachs Wolfe effect (in which CMB photons are gravitationally redshifted, affecting the spectrum). However, during the derivation of the requirements, only WL and BAO were taken into account.

The fundamental puzzle of the cosmological constant and the behaviour of gravity on the largest scales generated a great diversity of ideas. There is a strong need today to distinguish good theoretical models of the Universe. To achieve this model selection, an improvement of at least an order of magnitude in data quality and quantity is needed. This is the driving force behind the *Euclid* proposal. The fundamental questions that will be addressed by the missions are:

¹⁸<https://wfirst.gsfc.nasa.gov>

¹⁹<http://euclid-ec.org/> and <http://sci.esa.int/euclid/>

1. **Dynamical Dark Energy:** is dark energy a constant or is it time-dependant? In other words, what is the EOS of dark energy $w(a)$? *Euclid* will measure $w(a)$ to 1%. This performance was decided without theoretical assurances that it will be sufficient to stress-test the concordance model of the cosmological constant. Any significant deviation from $w(a) = -1$, however, would jolt the community, as it would show that it is not a constant.
2. **Modification of gravity:** is the cosmological principle a good assumption? Does gravity behave differently at large distances? There are a plethora of proposed modified gravity formalisms that potentially can explain the accelerated expansion of the Universe. Answering these questions would hint at whether to revise our understanding of space-time. Most of the modifications impact the growth of structures, by changing a parameter that quantifies the rate of growth.
3. **Dark matter:** what is it? What is the number of relativistic species of neutrinos? Understanding the nature of neutrinos, and in particular their absolute mass scale is important as they impact the structure formation over small scales. This damping imprints the observed matter power spectrum, which depends on the total mass of neutrinos. *Euclid* will measure the profile of dark matter halos and the matter power spectrum, imposing constraints on the self-interaction cross-section of dark matter.
4. **Initial conditions:** is the power spectrum of the primordial density fluctuations described by a Gaussian probability distribution? Answering this question will shine light on the physics of inflation. *Euclid* will characterise the dependency of the power spectrum of primordial fluctuations on its spectral index n_s with high precision, allowing model selecting inflationary physics. This measurement produces a result which is independent from CMB results. *Euclid* will measure the non-Gaussianities in the power spectrum, for which theoretical predictions differ.

To derive exquisite constraints on the cosmological parameters, the survey is optimised for WL and BAO. The former requires high quality images and reasonably accurate redshifts, determined through photometric redshift techniques in the visible and IR. The visible photometry will be acquired *via* ground-based observations. BAO techniques demand near-IR spectrography to accurately measure redshift. The IR photometry and spectroscopy will be provided by *Euclid* observations, with a much better performance than ground telescopes. Both probes are statistical techniques and their results should be derived from surveys covering a large fraction of the sky. These demands can only be met by a space mission with a wide field-of-view. The remaining systematic biases of a space mission such as *Euclid* much smaller than a ground-based collaboration. This is due, in particular, to the stability and small size of the PSF, which, as described in Sect. 3.2.2, is a major source of systematics for weak lensing.

In all four primary science domains, the contributions of *Euclid* will be significant. The accuracy and precision on cosmological parameters will increase dramatically yielding constraints improvements at least ten better than over the current results (see the impressive Tab. 2.2 of Laureijs et al. 2011). Forecasts for the performance of the mission show that all current

constraints will shrink. This shrinkage is due to the precision of probes, their joint predictions, and a very strict systematic control. *Euclid* probes were selected because dark energy and gravity on cosmological scales can best be fathomed with weak lensing and galaxy clustering (e.g. [Peacock et al. 2006](#); [Albrecht et al. 2006](#)). These two techniques can be combined to constrain cosmological expansion, growth of structures, and dark and luminous matter. Moreover, the probes are almost not sensitive to astrophysical effects. A critical advantage of the mission is to combine imaging and redshift surveys in the same regions of sky as the resulting tests are more powerful than individual probes. These improvements are driven by three advantages of *Euclid* ([Laureijs et al. 2011](#)). First, an exquisite PSF, which reduces the magnitudes of the corrections. Second, very deep NIR photometry, down to $m_{AB} \sim 24$, about three magnitudes lower than any ground survey. This allows to provide, in a conjoint effort with ground-based telescopes, much better photometric redshift (see, e.g., [Rhodes et al. 2017](#), detailing the argumentation for collaboration effort, notably on the photo- z front). Third, a gigantic number of NIR spectroscopic measurements. A large number of measurements are required when dark energy became important, $0.7 \lesssim z \lesssim 2.1$, which implies measuring $H\alpha$ lines in the near infrared, which can only be achieved from space due to Earth's atmosphere.

On top of the studies of the main scientific drivers of dark energy, dark matter and general relativity, *Euclid* will create an archive of legacy data, extremely valuable for years, if not decades to come. In short, this legacy database will contain more than a billion galaxies images and several million spectra for objects at high redshifts. At shorter distances, *Euclid* will be able to measure stellar populations of nearby galaxies, provide morphological information, masses and star-formation rates, and find over 10^5 new strong lensing systems.

3.6.2 Implementation of *Euclid*

The *Euclid* satellite was selected as part of ESA's Cosmic Vision 2015-2025, the current cycle of planning space science missions. The current mission plan is a merger of two previous space mission proposal, Dark Universe Explorer (DUNE) which was a weak lensing mission and Spectroscopic All Sky Cosmic Explorer (SPACE) which proposed BAO and redshift-space distortion surveys. Both proposal were selected to an assessment phase and the proposals were merged. After years of preliminary studies and reviews, the *Euclid* mission was definitely selected by ESA's Science Programme Committee as a medium-class mission in October 2011. A large ($\gtrsim 1000$ scientists from over 100 institutes) consortium lead by Yannick Mellier regrouping thirteen European countries²⁰ are now working hard to prepare this mission. The design and manufacturing of payload module (i.e. the telescope and the service module) was awarded in December 2012 to Astrium SAS (now Airbus Defence and Space) and the construction of the satellite to Thales Alenia Space.

The *Euclid* telescope is a 1.2-meter diameter Korsch placed onto a cylindrical bus measuring

²⁰Austria, Denmark, France, Finland, Germany, Italy, the Netherlands, Norway, Portugal, Romania, Spain, Switzerland and the UK, along with a few NASA researchers.

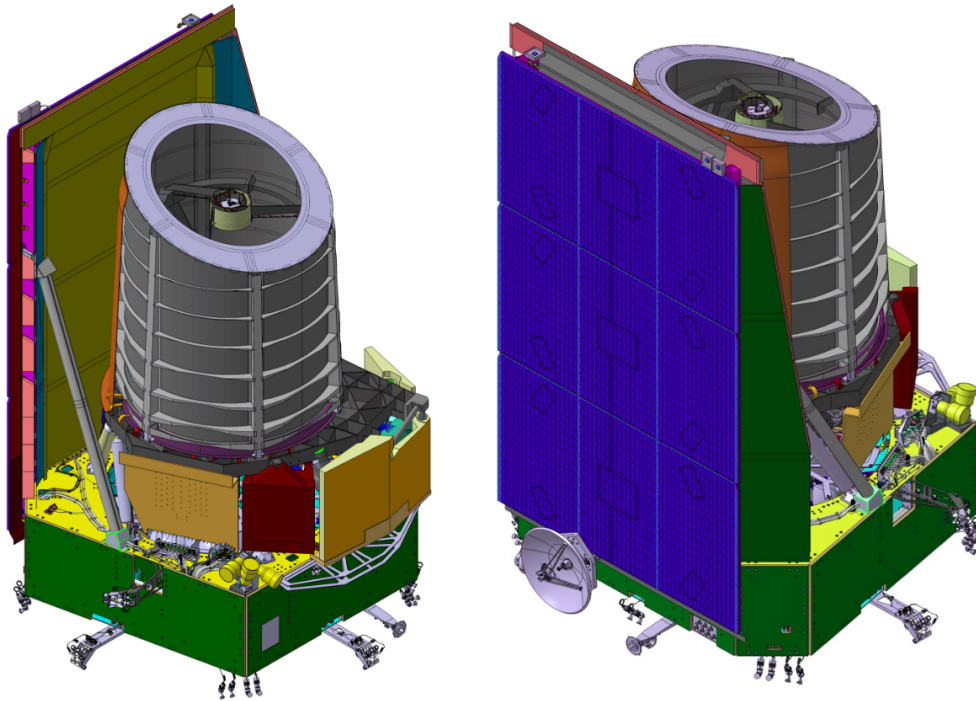


Figure 3.4 – Overview of the *Euclid* spacecraft. Credits: ESA and the *Euclid* consortium.

4.5 × 3 meters (see Fig. 3.4) for a total mass of 2 200 kg and solar panels capable of generating up to 2 400 W. There are two instruments: the high quality optical detector VISible Image (VIS) and a Near-Infrared Spectrometer and Photometer near-IR photometer (NISP-P) with three filters (Y, J and H) and a NISP slitless spectrograph (NISP-S). The images of the weak lensing survey will be acquired by the VIS instrument. VIS will have a $0.787 \times 0.709 \text{ deg}^2$ field-of-view with a broad band filter between 550 and 900 nm and comprises of 36 CCD detectors. The specially designed CCDs are provided by the e2v company are placed on a six by six matrix. There are $4096 \times 4132 \text{ } 12 \mu\text{m}$ pixels per CCD. The first four detectors were shipped in February 2017. VIS images will have a near-HST resolution of 0.1 arcsec per second, with exquisite sensitivity. The performance expectations is to reach a S/N of 10 for a 0.43 arcsec extended source of $m_{\text{AB}} \geq 24.5$ in a 1.3 arcsec diameter aperture from three 565 seconds exposures (Cropper et al. 2016). The two instruments are built by the consortium.

The bus-sized spacecraft is foreseen to launch on a Soyuz ST-2.1B (i.e. with a Fregat third stage) rocket from the European facility in Kourou. The launch is scheduled, for now at least, in May 2021. After a transit of a month, it will reach a halo orbit around the Sun-Earth Lagrange point 2. The mission will last six years, downlinking 850 Gbit/day through its K-band antenna. This major mission will generate an estimated 10 to 100 PB of data.

3.6.3 Weak lensing survey

The imaging of 15 000 deg² with a galaxy number density of ~ 30 galaxies resolved/arcmin² will yield 1 to 1.5 billion galaxy images after the mission lifetime. In addition to this wide field survey, two deep fields of 20 deg² each are planned. The deep survey will be sensitive to objects two magnitudes deeper than the wide field survey. There are two advantages to repeating deep field surveys: (i) they can serve as calibration for shape measurement methods and (ii) they are useful to assess the quality of the survey over time.

The observing strategy is step-and-stare: to observe the same portion of sky several times before moving to another. Each patch of the sky is imaged by four pointings. Each image will be exposed for 565 seconds with a significant dither, optimised such that each source in the field will be detected on at least three of the four images. In each of these fields, there will be an estimated 1 800 to 3 000 stars, mostly depending on galactic latitude of the field-of-view.

Table 3.1 – *Euclid*-derived requirements for the knowledge and reconstruction of the PSF according to science requirement 4.2.1.4 (Duvet et al. 2015; Paulin-Henriksson et al. 2008; Laureijs et al. 2011).

Parameter	Requirement
PSF ellipticity	$e < 0.15$
PSF profile	$R_{\text{PSF}}^2 / R_0^2 < 4$ where $R_0 = 0.2$ arcsec for Gaussian profile
PSF ellipticity stability	$\sigma(e_i) < 2 \cdot 10^{-4}$
PSF size stability	$\sigma(R^2) / \langle R^2 \rangle < 1 \cdot 10^{-3}$

Requirements on the weak lensing algorithms are defined by the cosmological parameter goal constraints and survey design. A few of the important requirements on the PSF are listed in table 3.1. The size of the PSF R^2 in the requirements is defined as

$$R^2 = q_{11} + q_{22}. \quad (3.11)$$

To reach the cosmological parameters, the weak lensing shape measurement algorithms are also constrained with the tightest requirements ever set. The requirements are a multiplicative bias of

$$|\mu| < 2 \cdot 10^{-3}, \quad (3.12)$$

and an additive bias of

$$|c| < 1 \cdot 10^{-4}. \quad (3.13)$$

These were the baseline for the GREAT3 challenge and continue to remain the goal for every new or improved shape measurement algorithm. They were devised in a series of influential papers that inspect the nature of systematics (Massey et al. 2013) and define WL experiment in space (Cropper et al. 2013).

4 Machine-learning techniques

Overview

In the previous chapters, we described the fundamentals of cosmology and gravitational lensing. We also enumerated a number of challenges linked to the measurements. The projects developed in this thesis mainly focus on improving shape measurement techniques. In this chapter, we present a few of the possible tools that are used to meet those challenges.

Correcting the shear for many effects usually involves simple mathematical operations. The difficulty resides in evaluating these corrections, as a function of the telescope optic and properties of the observed objects. The observation process is often simulated to calibrate the shape measurement methods. A promising way to perform best this non-trivial calibration is to exploit the capabilities of machine-learning techniques.

Machine-learning techniques try to achieve goals; for example to minimise prediction errors given a set of input data. Realistic simulations provide the necessary elements that warrants machine learning: the input data and importantly the true parameters (known as the “ground truth”, “truth” or “target”). The adaptability of machine-learning techniques is a great strength: the amount of resources allocated to the methods changes depending on the complexity of the pattern(s) in the data. More resources, i.e. more capacity in the method, makes reaching the optimisation goal more difficult, but it can deal with more subtle correlations. This optimisation process, appropriately called “learning”, usually requires large amounts of data and computation power. And here comes another strength of the approach: flexibility. What the technique learns depends on this objective. Once optimised, or trained, methods process new and previously unseen data fast. Machine-learning algorithms can be tailored for all sorts of applications, by adapting the structure, the objective or the approach.

In this chapter, we start by situating a few milestones of the field in historical context. We then turn to describing selected algorithms such that we will have a solid background to explain the work done in this thesis.

4.1 The turbulent rise of machine learning

The development of machine-learning techniques spans multiple era and fields. Machine learning gained in popularity in recent years as the amount of available data dramatically increased and computer power exploded. It is a computation-time expansive field as learning often requires to evaluate operations over large matrices. Text in this section is based on multiple references, which review the evolution of the field: [Bishop \(2006\)](#); [Schmidhuber \(2015\)](#) and [Goodfellow et al. \(2016\)](#).

4.1.1 Gradual evolution of the machine-learning approach

Machine learning is part of the larger field of artificial intelligence (AI). Today, AI is an extremely active field with reports of many impressive achievements. like self-driving cars (although with sub-human performance, see e.g. [Zhou 2017](#)) or supra-human abilities at many board games (e.g. with a clear Go victory over the best humans [Chouard 2016](#); [Silver et al. 2016](#)). Myths and dreams about what we would nowadays call AI can be traced at least to the Greek antiquity with, e.g., the giant automaton soldier Talos tasked with protecting Europa in Crete. Intelligent machines, capable of human-like reasoning, have a particularly powerful attraction, which prompted several hoaxes¹ and passionate reactions². With the advent of modern computers, AI has become a reality, or at least a very real possibility. Research into AI lead to the rise of smart automation of many processes, including in science³.

In the beginning of computers, applications were limited to solving a task by executing a list of mathematical instructions and did not learn (for example [McCulloch and Pitts 1943](#)). This was the era of rule-based systems, with a rigid rule set that a computer followed. Rules were fixed ahead of time. Rule-based systems can solve a large variety of problems and tasks and tremendously impacted society. They are still relevant today but cannot hold the title of machine-learning techniques.

There are many problems for which rules of behaviour are at least very difficult to write. A well known example in the lensing context would be the automation of lens detection. More general problems include understanding spoken words and finding certain objects in a picture. When humans take decisions or look at a scene, they infer information based on their extensive knowledge of the world. AI researchers therefore explored algorithms that were able to make logical inferences. Information was hard-coded to give context to the input data. However, this so-called “knowledge base” approach did not generate any significant achievement.

This unsuccessful attempt sparked research into machine learning, that is the capability of the algorithm to discover for itself the mapping from input features to the ground truth. Machine learning through its many methods quickly showed its potential. Providing sufficient

¹The most famous story is the “Mechanical Turk”. A fake chess-playing automaton of the late 18th century.

²There was a recent push by many influential individuals to take time to reflect on the consequences of AI research (e.g. for robot soldiers, [Gibbs 2015, 2017](#); [Sample 2017](#)).

³To name a few in no particular order: scheduling, data mining, parameter estimation, ...

relevant information is crucial to obtain the best performance. The basis in which this data is represented is also important to extract patterns.

Selecting the right features is a lengthy and uncertain process. Human-selected features can take months to be properly designed. Learning a representation of some raw data can be faster and usually yields better performance. An application of this representation learning is dimensionality reduction: for example compressing the PSF information encompassed on a pixel grid to a handful of coefficients. The coefficients are expressed in a basis that was learned and best expresses the data.

A further step in machine learning is the rise of deep learning. In complex problems, there generally are several layers of interpretation based in a simplest-to-most-complex hierarchy of concepts. Building a method with different layers of analysis is in the realm of deep learning. The methods first treat the most basic information, then combines the output of the first layer of analysis to feed the information to a second layer which, usually, is more abstract. Deep learning methods take raw inputs and transform them to the required output by executing a series of simple tasks.

It should be noted that the terminology in AI often changes and depends on the field and the sensitivity of the expert. Machine learning should be understood as an umbrella term that includes methods out of machine, representation and deep learning. Machine-learning algorithms devised at an early stage of the field can still be very well adapted to certain tasks and are, in general, conceptually easier to implement.

4.1.2 The early days

Least square fitting, admittedly part of the machine-learning family, was one of the first methods discovered and applied early in the 19th century. Least square fitting can be applied easily, even in certain cases without the help of a computing machine. With the apparition of the machine computers and the advent of the 1940s, the first studies into learning algorithms were conducted. This learning was modelled on the natural world and in particular the brain, thus the artificial neural network (ANN) denomination. Both simple classification and regression algorithms were tested. The conditions were the same as today, that is, given this input data, predict that output after learning on this training set. Maybe, the greatest achievement of that era was the discovery of the learning process known as stochastic back-propagation (see Sect. 4.3.2). The main limitation of those early models was that they were linear models and as such incapable of learning to perform certain tasks (like reproducing the XOR function, [Rosenblatt 1958, 1962](#); [Minsky and Papert 1988](#)). The failure to learn those (non-linear) tasks prompted a 10-year “AI winter”. The mood at that time was pessimism. Founding was slashed and almost no new results were announced until the re-discovery of back-propagation ([Rumelhart et al. 1986](#)). This pioneering era was guided by biology and neuroscience. At its very beginning, prototypes of flying machines were almost back-engineered from birds. Today however, aircraft do not flap their wings. AI also diverged from

its initial goal of copying nature, even if there is still a strong interaction between the two fields.

After the AI winter, new concepts emerged that dramatically improved the performance of the models. One of such concept is the representation of the data: features that effectively describe the data is better than many badly represented samples ([Hinton 1986](#)). Other advances include the Long Short Term Memory model ([Hochreiter and Schmidhuber 1997](#)). This model is now widely used in automatic translation because it is able to deal with long sequences of data, like sentences. An important theoretical result showed that an ANN with the right configuration could approximate any function with arbitrary precision ([Hornik et al. 1989](#)), thus encouraging researchers to dig deeper. In the mid-1990s however, many academia-to-industry transfer programs failed leading to a second decline in popularity of neural networks. At the same time, kernel-based methods started to shine in the machine-learning field. These methods, of which the popular PCA and Gaussian Processes are part, typically find relations in datasets without computation-intensive learning. Kernel methods are based on a mathematical trick that allows for linear techniques to be used in non-linear problems by increasing the dimension of the feature space. Those methods have been applied many times to many different problems in many different fields ever since with good results.

4.1.3 Modern machine learning

Neural networks were eclipsed by kernel methods for another ~ 10 years until, in 2006, methods for efficient training were discovered (see, [Hinton and Salakhutdinov 2006](#); [Larochelle et al. 2008](#)). This was the dawn of the age of “deep learning.” Models consisting of several layers of neurons linking input to output are refereed to as deep in opposition to shallow networks. The new training methods became ubiquitous in ANN research and, coupled with an extraordinary increase in computing power, lead to new highs in performance. Deep learning is now at the core of many almost-every-day technologies like object detection and classification in natural images ([Krizhevsky et al. 2012](#)).

Unsupervised learning was the focus at the beginning of this third wave of ANN research. Until a few years ago, algorithms must learn with a limited number of examples because it was expensive to collect them (because simulation is hard or/and labelling examples is lengthy). A way to improve the performance would be to learn from unlabelled data, thus the interest in unsupervised methods. There is currently an avalanche of data in all fields. They are not only available but need to be treated, and thus deep learning is revamping older machine-learning techniques for mass analysis. Current state-of-the-art methods are trained with many millions samples. The increase in computing power since the last dip in popularity of ANNs in the mid-1990s allowed to increase the size of networks both in terms of layers and neurons per layer. The complexity of the data is also constantly increasing. Natural images in the early 1990s were small and contained only one object. Currently, deep learning algorithms can analyse multi-object and high resolution images. Theoretical understanding of machine

learning and neural network in particular is still weak. Most of the research is empirical, even progress is made (e.g., [Bruna et al. 2013](#); [Bruna et al. 2015](#); [Mallat 2016](#)).

4.1.4 A golden age in the making?

There is an idea rooted at the core of the development of machine-learning: automatic, fast and efficient analysis of data. This objective is very current in the lensing community. With projects like *Euclid*, LSST or SKA, there is a definite need for algorithms with the three above qualities. Research, including this very thesis, demonstrates the power of current machine-learning tools for many applications. With increasing ANN capabilities, more complex tasks can be performed. At the time of writing, supervised learning, that is learning by studying labelled examples, as a whole does not suffer from major limitations, or at least the field has not yet expanded to them. The difficulty is to create realistic or enough data for an algorithm to learn the correct mapping from raw data to the desired output.

The difficulties of machine-learning algorithms to learn by themselves, to tackle so-called unsupervised problems, are real (see e.g., [Långkvist et al. 2014](#); [Karhunen et al. 2015](#)). The public understanding of the field is quite different from the reality. This major discrepancy could lead the field to plunge in another winter. By continued work in the development of new techniques and improved applications of machine learning, for example, but not limited to science, interest in these time-saving and effective methods can be sustained.

4.2 A feel for machine learning

In this section, we explore simple machine-learning techniques that are ubiquitous in today's algorithms and that were also used in this thesis. While some technicalities are described, the main goal when exploring those algorithms is to underline the machine-learning aspects. Many textbooks detail these methods, like [Ivezić \(2014\)](#); [Bishop \(2006\)](#); [Nielsen \(2015\)](#) or [Goodfellow et al. \(2016\)](#).

4.2.1 Least-square fitting

Least-square fitting is, at least conceptually, the simplest method exposed here. It assumes a model to fit the data. Data here is assumed for simplicity to be two dimensional, i.e. $\mathbf{x} = \{(x_1, y_1), \dots, (x_N, y_N)\}$. The model that maps the input x to the output y is $y = f(x)$ where f is an arbitrary function. The physical model $f(x)$ typically has parameters. In the case of a linear dependence, $f(x) = ax + b$, the slope a and intercept b are the parameters. If they are not correctly adjusted, a mapping that correctly models an experiment will produce wrong predictions. The fitting procedure allows to find the parameters such that some quantity is optimised. This quantity, which translates the error between the model and the data, can be regarded as a cost function. There is no cost to a model that perfectly fits the data and a non-zero cost to another that does not. The goal is to find the parameters that have the lowest

cost. The optimisation procedure thus minimises the cost of fitting a model to a given set of data. This cost is typically defined as

$$J(f, \boldsymbol{\theta}, \mathbf{x}) = \frac{1}{2} \sum_{i=1}^N (y_i - f(x_i))^2, \quad (4.1)$$

with $\boldsymbol{\theta}$ the different parameters of the model to be optimised and the $1/2$ multiplicative constant is included to simplify further computations. The problem of finding $\boldsymbol{\theta}$ such that the cost function $J(f, \boldsymbol{\theta}, \mathbf{x})$ is minimised reduces to finding $\boldsymbol{\theta}$ such that

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = 0. \quad (4.2)$$

This problem can be solved analytically for a few simple models. In the linear case, applying the least-square fitting procedure consists in inverting a matrix. In non-linear cases, however, the parameter space must be explored. Optimisation tools, such as gradient descent, are applied to find the approximate value of the parameters. There are different variations of least-square fitting, in particular to account for a weighting scheme of the data.

Least-square fitting is part of a class of methods called maximum likelihood estimators (MLEs). Likelihood in that context can be understood as the probability of the data given a model⁴. The MLE class of methods finds the maximum likelihood, i.e. the optimised model parameters $\boldsymbol{\theta}$.

Least-square fitting does not however extract the patterns in the data to infer a mapping from the input features x_i to the output y_i . It does not provide a model for the data, but given a mathematical model, it learns its parameters such that the cost of the model is minimum, and in that sense it is part of machine-learning methods.

4.2.2 The curse of dimensionality

In the previous section, we dealt with one-dimensional input data. The dimension of astrophysical inputs is typically in the range of a few dozen for reduced data to several thousands for images. This high dimensionality can cause serious complications. To illustrate the challenge, we try to fit a polynomial of order P to a dataset with dimensionality D . The resulting number of coefficients highly depends on the polynomial and grows as $\sim D^P$. There is a risk that the number of parameters far exceeds the number of training samples. Regions of the space spanned by the input data can be devoid of samples. Actually, in a general case, to guarantee that prediction errors made by the model are smaller than an arbitrarily small scale, δ , there should be at least $N = \delta^{-D}$ examples. This number becomes very large very quickly. This is why model fitting becomes extremely difficult in many dimensions. One can argue that, with the right assumptions, some information can be learned. For example, we can assume that the region is smooth and that the value at a given point can be interpolated from the neighbouring samples. This prior is used in many simple algorithms and works well if

⁴The likelihood is not a correctly normalised probability distribution function.

there is a sufficient number of data points. ANNs do not rely solely on the smoothness prior. They exploit the relationships between the different regions of a dataset. Classification and interpolation in high dimensions is difficult, even if ANNs are well equipped to counter the curse of dimensionality (Mallat 2016). A solution to simplify the problem can be to reduce the dimension of the data, by finding a new basis that represent the data more effectively.

4.2.3 Principal Component Analysis

PCA is a popular dimensionality reduction technique in astrophysics. PCA dates back to Pearson (1901). It is classified as an unsupervised learning algorithm, because the target is not known. Least-square fitting needs input features x_i and the ground truth y_i . It is supervised. This algorithm projects the data on a meaningful basis that represent the input data. The dimensionality of the data remains constant in this change of basis. An important property of this newly found basis is its orthogonality and thus the basis vectors are uncorrelated. By retaining a smaller number of coefficients than the dimension of the basis, the dimension of the data is reduced. The cost to be paid is a loss of information. The reconstructed data only approximates the original data. By construction, PCA orders the basis vector, such that the low frequency variations in the data are reproduced by the first coefficients and high frequencies by later coefficients. One can choose a cut-off in the number of coefficients to retain.

A two-dimensional example is shown in Fig. 4.1. The data is drawn from a normal distribution centred on the origin, but with a covariance matrix such that the direction of the largest variance has a non-zero angle with the horizontal. If the basis vectors are rotated such that \mathbf{e}'_1 is aligned with the largest variance direction and \mathbf{e}'_2 with the second largest variance direction, there is more information about the data in the first than in the second component.

The data \mathbf{x} contains N samples of dimension K . The $N \times K$ matrix representing the whole dataset is written \mathbf{X} . Note that this dataset is centred: the mean of the K features has been subtracted to the N samples. The sample covariance matrix is written as

$$\mathbf{C}_\mathbf{X} = \frac{1}{N-1} \mathbf{X}^t \mathbf{X}. \quad (4.3)$$

In a general case, this covariance matrix has off-diagonal non-zero terms which arise from correlations between the features of the data. The goal of the PCA decomposition is to find a projection, \mathbf{P} , such that is aligned with the direction of maximum variance. The first principal component, p_1 is the projection with the maximal variance. This variance is maximised when

$$\lambda_1 = p_1^t \mathbf{C}_\mathbf{X} p_1 \quad (4.4)$$

is the largest eigenvalue of the covariance matrix. The columns of \mathbf{P} are eigenvectors, also known as principal components. Once the projection $\mathbf{Y} = \mathbf{X}\mathbf{P}$ is found and its covariance $\mathbf{C}_\mathbf{Y}$ calculated, all that remains is to write the covariance $\mathbf{C}_\mathbf{X}$ as a function of the projection and its covariance $\mathbf{C}_\mathbf{X} = \mathbf{P}\mathbf{C}_\mathbf{Y}\mathbf{P}^t$, and to order by eigenvalues to find the set of principal

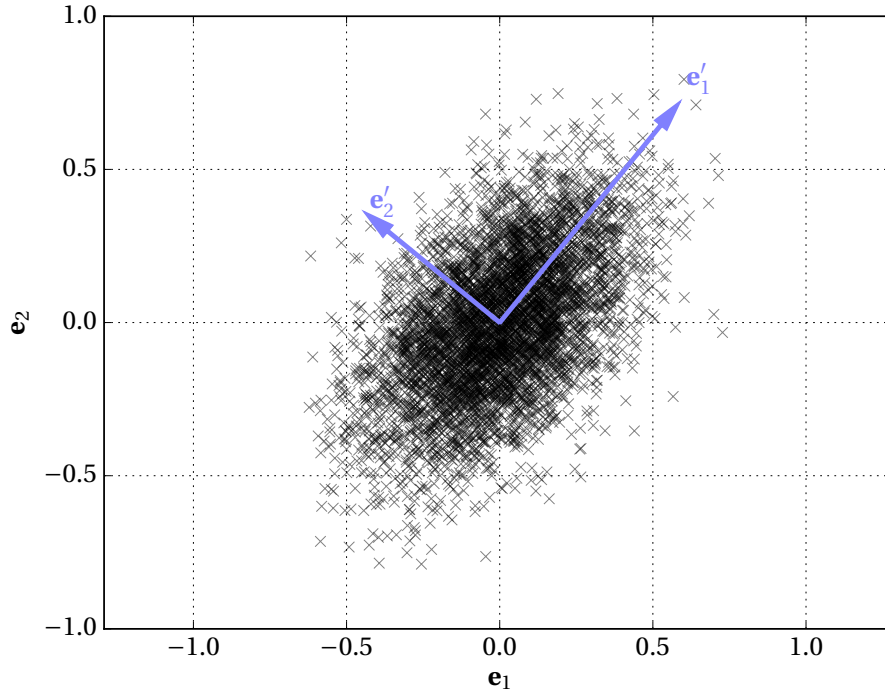


Figure 4.1 – Multivariate Gaussian distribution centred at (0,0) and the eigenvectors of the PCA covariance matrix in blue.

components for X . One of the advantages of the method is its an efficient computation technique. The decomposition of C_X into its eigenvalues can be computed using the singular value decomposition (SVD) directly from the data matrix X . Computation time for SVD scales as $\mathcal{O}(N^3)$ however making it impractical for large datasets.

PCA can be applied to a variety of different types of data. In this thesis, we are particularly interested in the decomposition of images. The following paragraphs introduce a few key points of the PCA-based image analysis. To illustrate its properties, we apply PCA to a very important dataset in machine learning: Modified National Institute of Standards and Technology (MNIST)⁵ (LeCun et al. 1998). MNIST is a large dataset of handwritten digits classically used for method-testing purposes. State-of-the-art methods reach an error rate of the order of 0.2-0.3%, human to supra-human performance. It comprises 60 000 training and 10 000 testing example. This is one of the important good practice of machine-learning testing: the performance of a method is assessed on data that was not used at the training stage. As a rule of thumb of current good practices, modern datasets are divided into a 60% training set and 20% test set. An additional validation set, with 20% of the samples, may be used to optimise hyperparameters (in the case of PCA the number of retained components). An image formed by a grid of $N \times M$ pixels has NM features. The PCA transform yields principal components also referred to as “eigenimages” (or eigenstars for the decomposition of images of stars, ...).

⁵<http://yann.lecun.com/exdb/mnist/>

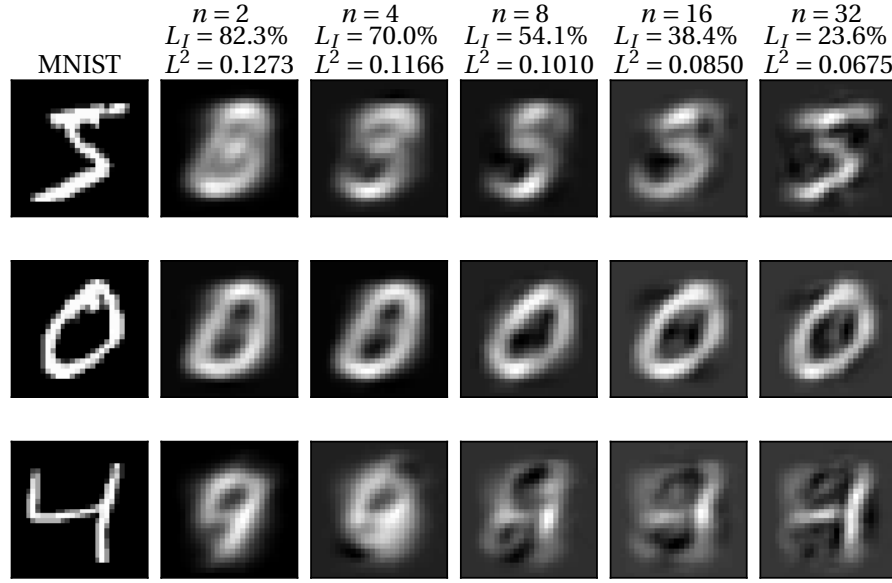


Figure 4.2 – *Left column*: original data. *Next columns*: PCA transform and inverse transform of MNIST data. The number of coefficients is denoted by n_c , the loss of information by L_I and L^2 the pixel-to-pixel root mean square error.

Figure 4.2 shows three MNIST images that are reconstructed using 2, 4, 8, 16 or 32 PCA coefficients. The PCA transforms the $K = 28 \times 28$ -pixel images to n_{coeff} , this can be interpreted as a compression stage. The inverse transform maps the n_{coeff} representation to the original K pixels, the decompression stage, which yields the approximate data $\tilde{\mathbf{x}}$. There is some information loss. This loss can be quantified by the sum of variance of the retained principal components over the total variance. The cut-off threshold can be set according to a metric that quantifies the similarity between the reconstructed data $\tilde{\mathbf{x}}$ and the original \mathbf{x} . That similarity index can be the information loss or for example how well PCA represents the wings of a PSF, or the central region of galaxies. We also introduce the L^2 error, to compare later with other techniques. It is clear that, for a low number of component, the image is very degraded. With a low number of coefficients, we see in Figure 4.2 that some entries are wrongly reconstructed as another digit (5 as a 3 with four components or the 4 in the last row as a 9 with two component and as an indiscriminate digit even with thirty-two components). This can be explained by looking at how the PCA decomposes the digits, see Fig. 4.3.

Expressing a dataset in a maximal variance basis highlights common features between samples. MNIST is a labelled dataset: for every image, the ground truth (a number between 0 and 9) is given. Plotting the data in the two first principal component space as in Fig. 4.3 shows the clustering of samples sharing the same ground truth. A classification tool can then be applied to such a map. In the case of MNIST data, retaining only two coefficients is not sufficient to reliably classify all classes, even if the tendency to cluster is clearly visible. The principal component cut-off threshold can also be selected according to clustering performance.

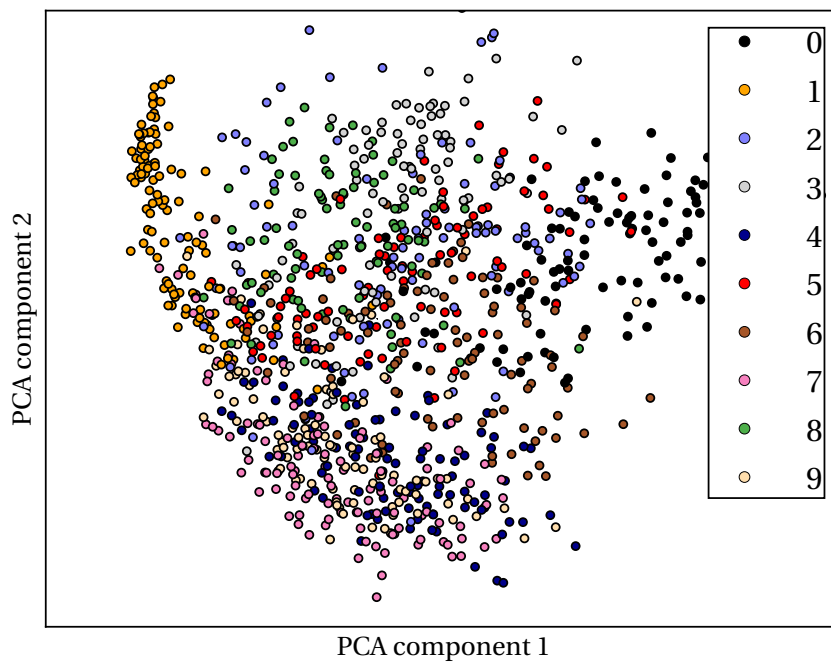


Figure 4.3 – MNIST samples represented in the PCA basis coloured by their ground truth.

We also show the eight first eigenvectors of the decomposition, or eigenimages. As eluded to earlier, part of the PCA procedure is to order the basis vectors by their variance. The first eigenimage will thus be showing low frequency variations of the dataset. Indeed, the first component is based on the mean feature. Further components describe higher frequencies. If the dataset is noisy, removing a small number of principal components can denoise the images. We present the first eight eigenimages of the MNIST PCA transform in Figure 4.4. The eigenimages do not map one to one to the ten digits in the truth set. However, the active regions (i.e. non-zeros pixels) concentrate in small area often looking like parts of numbers. They are actually edge (or shape) detectors, but this is not always the case with PCA. The shape of the active regions depend on the nature of the dataset.

PCA is a common tool to transform data (be it a spectra or an image) in the most meaningful basis. It is a machine-learning algorithm in the sense that it will learn from the data directly a more efficient representation. We will later see two different applications of PCA. First, in section 5.2, a decomposition of the dataset to classify its samples and second, in section 5.3.4 as a way to represent data more efficiently than on a pixel grid. There is a number of such dimensionality reduction techniques. Some are based on extensions of PCA. Other appear from a growing interest in techniques that apply sparsity constraints to data (see e.g. [Starck and Murtagh 2006](#)). That is to express data samples in terms of a much smaller number of non-zero coefficients than the input data. Unlike PCA, sparse methods not requiring the orthogonality of the basis vectors. Other techniques based on ANNs will be described in Sect. 4.5.

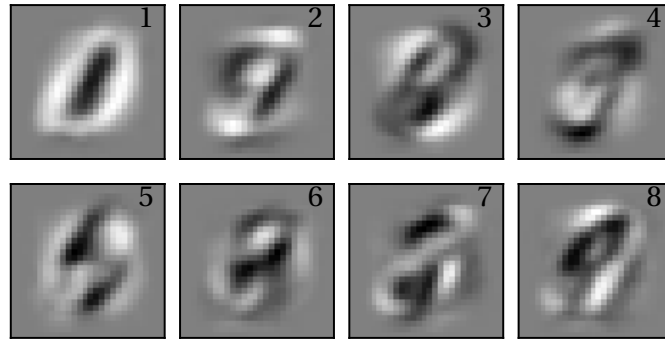


Figure 4.4 – First eight eigenimages of the PCA transform for MNIST. Black pixels are negative and white positive.

4.3 Artificial neural networks

In this section, we turn to ANNs, their basic components, training process and architecture. ANNs are well suited to solve both classification and regression problems. We will also introduce a variation of ANNs called convolutional neural networks, a somewhat recent discovery, that reaches impressive performances on classification tasks.

4.3.1 Fundamentals

The basic unit of neural networks is the neuron, a name that was directly imported from neuroscience. It takes an input \mathbf{x} and computes a scalar output called activation a . For each data sample, there is a number of features and a ground truth, t . When presented with the features \mathbf{x} , the model is expected to produce the target t .

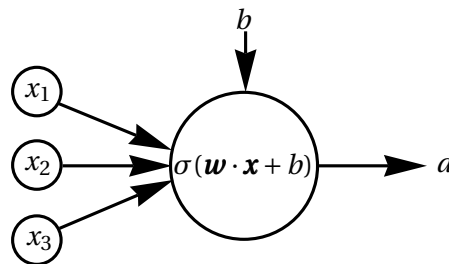


Figure 4.5 – Sketch of a neuron.

A neuron (see Fig. 4.5) accepts any real values for \mathbf{x} and first computes a so-called weighted input z , a scalar, as

$$z = \mathbf{w} \cdot \mathbf{x} + b, \quad (4.5)$$

where \cdot is the scalar product, \mathbf{w} a vector of weights and b the bias. Model parameters consist of these two quantities. In neuroscience, a neuron can be activated, that it yields an on/off

impulse⁶, which depends on input messages. Artificial neurons do pretty much the same. The quantity z is the weighted input for an activation function $\sigma(z)$. There is a wide range of activation functions, which must be monotonic. The simplest is the linear $\sigma(z) = z$, but it does not present non-linear behaviour. Non-linear activation functions include the hyperbolic tangent, $\sigma(z) = \tanh(z)$, the sigmoid and the rectified linear unit (ReLU). The sigmoid function is defined by

$$\sigma(z) = \frac{1}{1 + \exp(-z)}. \quad (4.6)$$

It has been the most commonly used activation function until the recent appearance of ReLU

$$\sigma(z) = \max(0, z). \quad (4.7)$$

Its advantages over the sigmoid function are faster computation and it only sparsely activates the neurons. The result of the above computation is called the activation a and is the output of a neuron. The mathematical model of the neurons is written

$$a = \sigma(z) = \sigma(\mathbf{w} \cdot \mathbf{x} + b). \quad (4.8)$$

The parameters to be tuned in this neural model are $\theta = \{\mathbf{w}, b\}$. In section 4.3.2, we will see how to train the parameters such that they produce the correct outputs.

A network consists of neurons organised in layers that sequentially propagate input features through until the last layer. The different nodes in the same layer do not communicate with each other, they treat the input independently. When all neurons of a layer are connected to all inputs and outputs, they are known as fully-connected, or dense. As their weights and biases are different across the layer, they will produce different results. Individual neurons in a layer are not activated by the same value of the input features. They are sensitive to different characteristics of the data. The number of neurons in layer ℓ is not constrained by the dimension of other layers or the dimension of the input. We denote the number of neurons in a layer by the quantity n_{j_ℓ} . Outputs of a layer, i.e. a vector of the activations, \mathbf{a}_ℓ are the input for the next layer $\ell + 1$ such that for an arbitrary neuron $j_{\ell+1}$,

$$a_{\ell+1, j_{\ell+1}} = \sigma(\mathbf{w}_{j_{\ell+1}} \cdot \mathbf{a}_\ell + b_{j_{\ell+1}}). \quad (4.9)$$

The final layer is known as the output layer. All layers located between input features and the output are called hidden as shown on Fig. 4.6. The number of nodes per hidden layer is arbitrary, while the input layer is, of course, of the same size as the input data and the output's as the target. Finding the correct architecture for the networks is not straightforward, we will touch on that subject in Sect. 4.3.3. A network with several hidden layers is already called deep; the dividing line between shallow and deep being unclear. When an architecture for a network is determined, the next step is to train it.

⁶Note that here we dramatically simplified the picture.

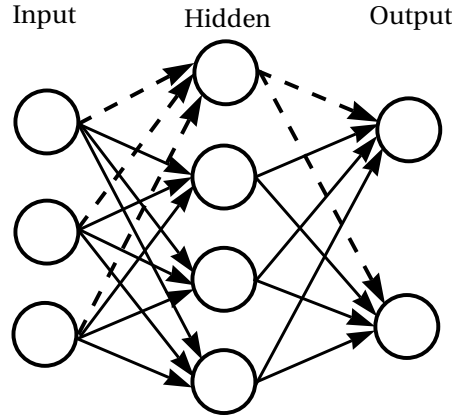


Figure 4.6 – Network of neurons organised in fully-connected layers. The input layer takes features in and propagates them through different layers until the output and the predictions. Dashed lines show all connexions between the input layer, the first hidden node, and outputs. In this particular case, the output has a dimension of two. This kind of networks is sometimes called multilayer perceptrons (MLPs).

4.3.2 Optimisation of model parameters

Training is the process of optimising the model parameters of all neurons θ assuming a cost function J . This cost function is to be minimised during training. A large proportion of machine-learning tools are trained using a maximum likelihood approach. Cost functions are thus derived by taking the log-likelihood of the model distribution. There are multiple cost functions that must be chosen according to the task at hand. A common choice for regression problems is, like in least-square regression, the mean square error (MSE)

$$J(\theta) = \frac{1}{2} \sum_i^N (y(x_i, \theta) - t_i)^2, \quad (4.10)$$

where the input data is \mathbf{x} , the truth or target vector is \mathbf{t} . The function $y(\mathbf{x}, \theta)$ yields the output of a model to the input data sample \mathbf{x} . When the data is to be classified in two classes (i.e. $t = 0$ or $t = 1$), a cost function based on cross-entropy between target and input data is better. Training a classifier based on cross-entropy is faster and usually performs better than one based on MSE. The cost function can be extended to a multi-class error function to be used in mutually exclusive many classes problems, like say, is this object a tennis ball, a football or a basketball? In addition to the error based on the output of a model, one can add regularisation term(s). Regularisation should be understood as a way to minimise the generalisation errors. The performance of a model on a dataset that was not used during training is reduced compared to its training performance. New errors creep in because those new samples may contain new information (or some combination of features) with which the model struggles to deal. Regularisation can reduce these generalisation errors, but will not decrease the training error; it constrains the model further. If too much regularisation is applied, the prediction will underfit the data, that is the learned model will be too simplistic for the data. On the contrary,

if there is no regularisation and the model has too many free parameters, the model risks to overfit: it would be capable of reproduce the data very well but loose all generalisation capabilities. Putting penalties on model parameters is a common regularisation technique. This new term is controlled by a multiplicative factor λ and consists of some norm of the model parameters. A popular choice is L^2 for which the extra term R reads

$$R(\boldsymbol{\theta}) = \frac{\lambda}{2} \|\boldsymbol{\theta}\|_2^2. \quad (4.11)$$

The goal of this regularisation is to push down the weights, and also called weight decay term. Another is the L^1 regularisation where the norm is changed to $\|\boldsymbol{\theta}\|_1$. In contrast to L^2 which tends to minimise all weights, L^1 leads to sparse weights. There are other forms of regularisation: like neglecting some of the connexions in an ANN as to avoid over-specialisation of nodes in an ANN (i.e. dropout, [Hinton et al. 2012](#); [Srivastava et al. 2014](#)).

Now that the error function is established, we turn to the optimisation of the model parameters with back-propagation. We thus aim to find the global minimum of the cost function, which is typically a highly non-linear function. The first step in the training process is to choose a first guess for the parameters $\boldsymbol{\theta}^{(0)}$. The τ th update of the parameters is written

$$\boldsymbol{\theta}^{(\tau+1)} = \boldsymbol{\theta}^{(\tau)} - \alpha \Delta \boldsymbol{\theta}^{(\tau)}, \quad (4.12)$$

where $\Delta \boldsymbol{\theta}^{(\tau)}$ is the weight vector update, a term to be computed, and $\alpha > 0$ the learning rate. A stochastic gradient descent (sometimes abbreviated SGD) conducts a stochastic approximation and is, formally, estimated from a single point. However, in practice, one uses a solution between the purely stochastic gradient and the true gradient: batch optimisation, that is using a fraction of the training data to estimate the update. Usually, the data is separated into different batches. Batches are used iteratively to train the method. Once all batches have been used, a so-called epoch of training is done and the training goes on, further computing predictions errors from the first batch again. The gradient of the cost function, $\nabla J(\boldsymbol{\theta})$, is computed to derive the update vector, such that, as the name tells, the algorithm follows the gradient to find the optimum. There is no guarantee that gradient descent will find the global optimum within a reasonable amount of time. However, it usually converges to a low enough minimum quickly enough to be useful for machine-learning applications. To that end, we will take advantage of the back-propagation procedure.

Back-propagation is based on a few simple steps outlined here for a general ANN with L layers. The goal is to compute the gradient of the cost function with respect to the model parameters. This gradient is the update vector in eq. (4.12). For the last layer L this is $\frac{\partial J}{\partial \boldsymbol{\theta}_L}$. This expression can be expanded to

$$\frac{\partial J}{\partial \theta_{j_{L-1}j_L}} = \frac{\partial J}{\partial a_{j_L}} \frac{\partial a_{j_L}}{\partial z_{j_L}} \frac{\partial z_{j_L}}{\partial \theta_{j_{L-1}j_L}}, \quad (4.13)$$

where we made the simplifying assumption that there is only one training sample (in order

to get rid of the vectorial notation). The index j_ℓ runs over the n_{j_ℓ} neurons of the ℓ th layer. The first left-hand term is the derivative of the cost function with respect to the activation. Carefully chosen activation and cost functions lead to simple expressions for the above and thus limit the computation time and in turn decreasing the training time. For, e.g., MSE, this yields the very simple

$$\frac{\partial J}{\partial a_{j_L}} = a_{j_L} - t_{j_L}. \quad (4.14)$$

The second term is the derivative of the activation function

$$\frac{\partial a_{j_L}}{\partial z_{j_L}} = \sigma'(z_{j_L}) = a_{j_L} (1 - a_{j_L}), \quad (4.15)$$

where we assume the sigmoid function as activation function, σ . We now write, as a shorthand notation the error at the neuron j_L of the last layer as

$$\delta_{j_L} = \frac{\partial J}{\partial a_{j_L}} \frac{\partial a_{j_L}}{\partial z_{j_L}}. \quad (4.16)$$

The update for the last layer and neuron j_L is now

$$\frac{\partial J}{\partial \theta_{j_{L-1}j_L}} = \delta_{j_L} \frac{\partial z_{j_L}}{\partial \theta_{j_{L-1}j_L}} = \begin{cases} \delta_{j_L} a_{L-1} & \text{for weights } w \\ \delta_{j_L} & \text{for biases } b \end{cases}. \quad (4.17)$$

Updating the parameters of the last layers is not sufficient to train the model. Parameters for all layers must be updated as well. The gradient of the cost function for the next-to-last layer $L - 1$ is computed from the error of the last layer δ_L . This can be noticed from the following quantity

$$\frac{\partial J}{\partial z_{j_{L-1}}} = \sum_{j_L} \frac{\partial J}{\partial z_{j_L}} \frac{\partial z_{j_L}}{\partial a_{j_{L-1}}} \frac{\partial a_{j_{L-1}}}{\partial z_{j_{L-1}}} = \sum_{j_L} \delta_{j_L} w_{j_{L-1}j_L} \sigma'(z_{j_L}) = \delta_{j_{L-1}}, \quad (4.18)$$

which is actually the error term for neuron j in the next-to-last layer $L - 1$. The update for that neuron is can be computed

$$\frac{\partial J}{\partial \theta_{j_{L-2}j_{L-1}}} = \frac{\partial J}{\partial z_{j_{L-1}}} \frac{\partial z_{j_{L-1}}}{\partial \theta_{j_{L-2}j_{L-1}}} = \delta_{j_{L-1}} \frac{\partial z_{j_{L-1}}}{\partial \theta_{j_{L-2}j_{L-1}}}. \quad (4.19)$$

In summary, back-propagation to compute the gradient of cost function with respect to the model parameters, is

1. Activations a and weighted inputs z of each of the layers ℓ are computed. This action of propagating the activation through the layers of the network until the output layer is called feed-forward;
2. Error terms of the last layer δ_{j_L} are evaluated using eq. (4.16);

3. Errors are propagated from the last layer to the previous one and so on until the first layer using eq. (4.18);
4. Update terms are computed by multiplying the errors by $\partial z_{L-1} / \partial \theta_{j_{L-2} j_{L-1}}$.

When dealing with more than one training example, the total gradient $\partial J / \partial \theta$ is recovered from taking the average over the samples. It was therefore implicitly assumed in the above that the total cost function can be written $J = \sum_i^N J_i / N$ where the index i runs over the samples. In modern implementations of back-propagation, the algorithm takes advantage of vectorial coding tools that considerably speed up the computation and simplify the code.

A hyperparameter remains to be tuned. The learning rate α has to be set. There are various approaches as to how to determine it from a systematic grid search to more refined methods linked to regularisation terms. Adding regularisation terms does not change the approach of back-propagation and must be integrated in the cost function J to derive appropriate update rules. Another choice that can prove important, and whose importance is sometimes overlooked, is the initialisation of the parameters. They are usually drawn from a normal distribution, but more elaborate solutions try to optimise initialisation of all layers with the same gradient and activation variances (see e.g. [Glorot and Bengio 2010](#)). Other techniques than gradient descent and back-propagation exist to learn the best model parameters given the data.

4.3.3 Building an effective ANN

The activation function does not need to be the same across the network. For classification tasks, the last layer contains often the same number of nodes as there are classes. To translate the non-normalised output of the network to a normalised probability distribution function (PDF), the activation of the last layer is a softmax. In regression tasks, the last activation function is often linear such that any value can be easily reached. If the output were the result of a non-linear function the activation would change fast between zero and one. Intermediary values would require (very) finely tuned parameters.

Architectures of ANNs, that is the number of neurons per layer and how many layers, are not easy to optimise. Even if some advices can be found, mostly in technical publications (see e.g. [Bengio 2009](#)) and blog posts, there is no definite and simple way to determining it. Common practices are to either optimise the structure by testing numerous possibilities or to build on architectures that were successful on related tasks.

Input data should be pre-processed to ease learning and prediction. Depending on the activation function of the last layer, the output can be bounded. So, pre-processing must be applied both to features and targets. Again, there are several possible ways to explore that can result in small difference in performances. An example of pre-processing would be to standardise the data (subtract the mean of the feature and divide by its standard deviation).

ANNs are able to approximate any function with arbitrary precision (this is called the universal approximation theorem, [Cybenko 1989](#); [Hornik et al. 1989](#); [Hornik 1991](#)), but this result does not provide a method to determine the architecture of the network. More neurons and more layers can lead to better results, but this is not always the case. As in most machine-learning techniques, more free parameters in a model lead to a better fitting of the data, but possibly also to overfitting. In this thesis, we take a precautionary route and always try to minimise the size of the architecture. ANNs as they were presented in the above have rarely more than a few layers, but can reach several hundreds of neurons per layer. It is hard to give typical numbers for the number of neurons per layer or even of layers, because it varies much across fields and applications.

The number of hidden layers is limited by a learning issue. The different layers actually learn at different rates. Late layers can adapt to a task much more quickly than layers closer to the input, in effect slowing down the learning process of the network. Note that the inverse (early layers learning while late layers seem stuck) can also happen. Effectively, this means that the parameters of the slow-learning layers are almost constant throughout the optimisation, the network can become seemingly stuck in a non-optimal state. This is due to the magnitude of the gradients that tends to decrease as they are back-propagated through the network. For this reason, this effect is called the vanishing gradient problem (and the inverse is called the exploding gradient problem). Looking at eq. (4.19), it is obvious that if all individual terms are smaller than one (the random initialisation produces weights smaller than one most of the time, and as z initially averages to zero, the mean of $\sigma'(z)$ is $1/4$), gradients are becoming smaller and smaller for layers closer to the input. Classical ANNs, a MSE cost function and sigmoid activation function trained using back-propagation, are inherently victims of unstable gradients. They cannot grow beyond a few (understand three or four) hidden layers before suffering from the vanishing gradient problem. Even if this a fundamental hurdle of the field, many of the modern natural images classifiers are dozen of layers deep (see, e.g., Google's 22-layer deep natural images classifier, [Szegedy et al. 2014](#)). Effects of unstable gradients can be mitigated using a variety of ways ([Bengio 2012](#)): reducing the number of connections, specifying the model further using regularisation, changing the activation function or taking advantage of the computing power of GPUs. The two last propositions aim at reducing the computing time for a single training step such that in the same amount of time, much more training steps can be done. Another way out is to train the network in two steps. First by finding approximate parameters for each of the layers sequentially. This is in effect a greedy layer-wise pre-training step ([Hinton and Salakhutdinov 2006](#); [Bengio et al. 2006](#)). Then fine-tuning them using classical back-propagation of the gradient. Greedy pre-training is based on unsupervised learning: each layer learns a new representation of the input, based on the representation of the previous layer. Combining unsupervised and supervised learning gave rise to the concept of auto-encoders explained in Sect. 4.5.

Another way to improve the performance of the network is to present it with more data: by showing more data, the network learns to generalise better. A simple algorithm fed with a sufficient number of training samples will perform better than a sophisticated method trained

on a small dataset. Increasing the size of the data is relatively easy when it is simulated. In case of real data, however, it is necessary either to label more as yet unlabelled data or to collect more samples. Those two possibilities might be time consuming and expansive. Most of the data generation processes imply some invariance. Taking the example of a handwritten digit reader, a two is still a two if it is moved a few pixels away from the centre of the image, or rotated by some angle. Data augmentation is a technique that takes advantage of these invariances to increase multiple times the size of the dataset. The algorithm learns that the invariance does not influence the prediction as the target is the same. The main drawback is that augmented data is not purely new data. The examples have undergone simple mathematical operations (translations, rotations, symmetries, ...), but stay inherently the same. A machine-learning tool taught with augmented data, when possible and with the previous caveat born in mind, usually produces better results than with the original data set size.

Checking that the network has learned well can be achieved by testing it on a validation dataset, which was not used for training. A large difference in the value of the cost function would indicate a poor generalisation ability. This is an easy diagnostic to assess the quality of the predictions, but one can also visualise the model parameters and interpret them. Early layers are always easier to interpret. They are usually sensitive to clear patterns in the data. Returning to the MNIST examples, correctly trained neurons can be interpreted as, like for the PCA eigenimages (Fig. 4.4), edge detectors. Visualising neurons sensitivity allows to detect dead neurons or duplicates that specialise in the same task. Such features in the neurons behaviours can be addressed by changing the architecture of the network.

Training is a stochastic process due to the random nature of the initial weights and data samples, which are drawn from a given distribution. When two networks, with the same architecture, are trained on the same set of data, their performance and final value of the cost function will likely be different. The quality of training of course depends on the patterns in the data. In classification, when the representation of the data generates well separated classes, it is easy to learn good classification rules, even with the simplest techniques. For more complicated data sets, predictions might differ for a fraction of the samples from one network to another. A way to mitigate this is to apply ensemble learning, that is training several networks and, when predicting the label or value of some new data, polling them to extract some sort of consensus. The pool of network is called a committee. A common way to determine the committee's output, is to take a weighted average of the members' predictions. We refer to a learning run that become stuck in so high local minimum that the performance is severely affected as catastrophic. Members that underwent a catastrophic training can be prevented from voting. Weights are estimated either from their training or validation performances. The downside of the committee approach is that the resources need are multiplied by the number of members. Members, however can have simpler architectures than single-member ANNs for similar performances, thus decreasing individual training times.

4.3.4 Convolutional neural networks

Common, fully-connected, neural networks treat inputs without accounting for their spatial structures. In MNIST images, there is a clear spatial structure. Two features are treated equally by ANNs irrespective of their relative position. To better classify data that is organised on a grid, like images, the network architecture must be revised. A seminal paper by [LeCun et al. \(1998\)](#) introduced the concept of convolutional neural networks (CNNs), also known as convolutional nets, or convnet. Input data is no longer vectors, but tensors and can thus now represent an image with one or multiple channels. A basic example of a CNN is shown on Fig. 4.7. CNNs are built on a few basic blocks, that we describe in the following.

Neurons in a **convolution layer** are not fully connected to the input data. Instead, they are receptive to a small region of the input. The position of this receptive area depends on the position of the neuron in the hidden layer. This creates, in effect, a convolution. The kernel is made up of the model parameters to be learned during training. An arbitrary number of kernels can be learned for each convolution layer. The number of kernels to be learned, its spatial extent, window sliding step, and size of zero-padding are hyperparameters, and, hence, not determined during training. These K kernels, i.e. the weights and biases, are shared by all neurons in the layer. The result of the convolution layer, the activation of the neurons, are K different matrices. A great advantage of these shared weights and biases is that the number of parameters to be optimised is dramatically reduced in comparison to a fully-connected layer. This, in turns, diminishes the importance of the vanishing gradient problem.

It is common to include a **pooling layer** as a follow-up to a convolution. This operation reduces the size of the representation, in an effort to further decrease the number of parameters. There are no parameters to be learned for this block. The extent of the pooling region and the operation are set as part of the architecture of the network. Pooling does not reduce the number of features maps, only their size. Max-pooling, a very common pooling technique, is a simple operation that, in essence, takes the maximum over a region of the feature map.

Another non-parametrised layer is a layer that introduces non-linearity. This layer simply applies the **ReLU** operation to the activation maps.

A **normalisation** layer normalises activations of the previous layer, usually by applying a standardisation operation, with the intent of decreasing training time.

Fully-connect networks are typically the last block after a combination of the above layers. They can be several hidden layer deep and usually end with a softmax layer to classify the learned representations according their respective labels.

Typical layer patterns can be summarised as

$$\text{Input} \rightarrow (\{\text{convolution} \rightarrow \text{ReLU}\} \times N_c \rightarrow \text{pool}) \times N_m \rightarrow \text{Fully-connected} \times N_F \rightarrow \text{output},$$

where N_c , N_m , and N_F indicate repetitions. Designing the architecture of a CNN is complicated

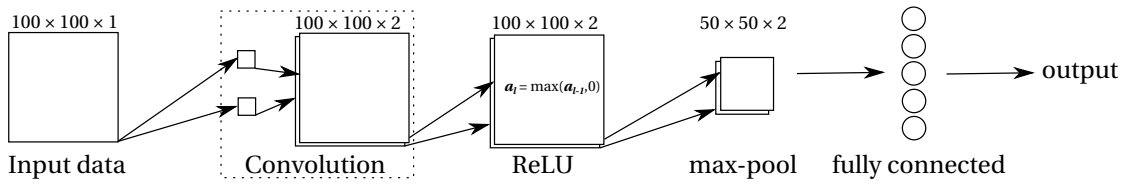


Figure 4.7 – Example of a small CNN architecture containing all the of the important blocks. The connexions of the last fully-connected layer are omitted for clarity. The numbers showed above each layers denote the dimension of the data. Input data is a two dimensional matrix. The convolution layer has two kernels.

and heuristics are few, just like for ANNs. Fortunately, there are examples available. ImageNet⁷ is a database of natural images (understand everyday images) containing millions of samples. Identifications challenges are organised every year using this database, and the best algorithms are, for most of the tasks, CNNs. Winning methods required major computational resources. Their architecture is public (sometimes their trained parameters as well) and non-experts in CNN design are encouraged to reuse the architectures (for some of the most famous and performing designs, see [LeCun et al. 1998](#); [Krizhevsky et al. 2012](#); [Simonyan and Zisserman 2014](#); [Szegedy et al. 2014](#); [He et al. 2016](#)). Weights of good models can be reused, and only the last layer should be retrained. This technique is called transfer-learning.

The training process of CNNs is similar to supervised ANNs'. Enough data must be labelled, the cost function adjusted to the task and enough CPU or GPU resources must be dedicated to training. Small networks with a few hundred thousand examples take of the order of the hour to train on CPU-limited laptops, but deep networks with millions examples need large GPU clusters to be trained in a few weeks. CNNs are popular nowadays because of their relative versatility and the many packages that propose ready-to-use networks.

There are, of course, extensions of CNNs. We will mention here one that will be used in Sect. 6.3: residual CNNs (introduced by [He et al. 2015](#)). The information in a CNN built with the above blocks flows from layer to layer in a continuous and sequential manner. Residual networks add more pathways, and modifies training slightly. Before processing convolutions, the information x is stored and added to the output of the convolution maps F . In a conventional convolution layer, the kernel K is learned from an unreferenced mapping. If the optimum kernel is denoted K^* , the network tries to learn that $K(x) = K^*(x)$. In a residual network, we are in a better situation from the training point of view of training, as we have $K(x) + x = K^*(x)$.

⁷<http://www.image-net.org/>

4.4 Random forests

In the introduction of this chapter, we quickly discussed rule-based methods. We discarded them from the machine-learning technique club, because they do not learn to adapt to the data. Decision tree and its evolution, random forests (RFs), are methods that learn rules to classify⁸ data, according to some pre-determined heuristic. These rules take the form of decision boundaries, which separate the data representation space into the different labels.

4.4.1 Decision trees

Decision boundaries can be applied hierarchically, from a rough to much finer classification. A decision tree learns rules that are applied sequentially, following a tree-like structure. At each node, the data is divided into two subsets according to some rule that was learned on a training dataset. This separation of the data into subsets in further child nodes continues until a stopping criterion is reached. This criterion can take several forms: nodes contain only one class of objects, introducing further nodes do not improve on the performance, or the pre-set depth of the tree is reached.

The important step when growing a tree is to choose the decision, or split, criterion. This entails selecting the feature and the boundary value for each of the nodes, independently. Usually, decision trees split data according to one feature at a time, so the learning procedure must find the best variable. There are several ways to define the splitting criteria. It is usually defined by estimating a measure of the homogeneity of the data in feature space, and then selecting the highest scorer. In this work, every time we used random forest, we chose to work with the Gini coefficient, G . This coefficient approximates the probability of misclassification of a random sample according to the label distribution. For tasks with k classes, $G = \sum_i^k p_i(1 - p_i)$, where p_i is the probability of class i given the data.

There is no theoretical limit to the depth of a tree, but the number of nodes increase exponentially with the depth. This therefore induces an important risk of overfitting in decision trees, and individual trees are known to be less accurate than other techniques and non-robust.

4.4.2 Growing forests

In the previous section on ANNs, we explored the idea of ensemble learning by forming a committee of network members that voted on a consensus. Here again, we use ensemble learning to mitigate the disadvantages of individual trees. Proposed by [Breiman \(2001\)](#), the idea behind RF is to transform weak learners into a group of strong predictors. There is an obvious name for a technique that is based on many trees: a forest.

The classification of a data point by a forest is an average of the predictions from individual trees. These trees are built to work on a few randomly selected features. This makes the RF

⁸There are extensions of RF for regression tasks, but they will be neglected in this manuscript.

robust to noise and outliers. By averaging over many trees built with different rules, and because of the law of large numbers, RFs do not suffer from overfitting. The accuracy of RF is higher than individual tree's accuracy. Random inputs and features result in good classification performance. Their training time is very short in comparison to a method based on ANNs, and can score as high. RF can be applied to unsupervised cluster, or to the detection of outliers (see Sect. 5.1.5 and 5.1.6). Another advantage is that the importance of each feature in the data can be readily extracted out of the tree structure. In addition, trees can be grown in parallel. ANNs still prove to be better at feature extraction than RFs, and deep learning model tend to perform better on competitions like MNIST and ImageNet.

4.5 ANNs for dimensionality reduction: auto-encoders

In the PCA section (4.2.3), we described how a machine-learning technique can be used to convert high-dimensional data to low-dimensional coefficients, or codes. Hinton and Salakhutdinov (2006) propose a non-linear generalisation of the PCA method. This adaptation is a deep neural network subdivided into two stages: encoder and decoder. The middle layer is a bottleneck as it contains the lowest number of neurons, producing the code, i.e. the representation, of the input data. Auto-encoders (AEs) being deep networks are subject to unstable gradients and are thus notoriously hard to train. Hinton and Salakhutdinov showed that deep networks, like AEs, can be optimised using gradient descent if the initial values of the model parameters are already close to a good solution. Learning the compressed representation of the data, in that well initialised case, with AEs is more successful in many instances than with PCA. In this section, we highlight some important AE concepts, for a more gentle and coding-oriented approach, visit the Stanford UFLDL tutorial⁹.

4.5.1 Learning the identity function

The goal of an AE is to learn the task

$$\mathbf{x} \mapsto \text{AE}(\mathbf{x}) = \tilde{\mathbf{x}} \text{ such that } \tilde{\mathbf{x}} \text{ is arbitrarily close to the input } \mathbf{x}, \|\tilde{\mathbf{x}} - \mathbf{x}\| < \delta. \quad (4.20)$$

We denote the encoder by $f(\mathbf{x})$ and the decoder by $g(\mathbf{x})$, such that AEs are described by $\mathbf{x} \mapsto g(f(\mathbf{x})) \approx \mathbf{x}$. The code is obtained by applying the encoding function $f(\mathbf{x})$ to the input.

A typical architecture is shown in Fig. 4.8. The encoding and decoding parts both consist of the same number of layer, and the same number of neurons, mirroring each other's design. AEs that have multiple hidden layers are sometimes referred to as stacked, because of the pre-training technique used to find sufficiently good parameters. The number of neurons per layer decreases from the top to the bottleneck layer. In the example in Fig. 4.8, the number of neurons is monotonically decreasing from the input down to the code layer. The first layers can be, however, overcomplete. Overcomplete AEs do perform well, and do not overfit. This

⁹http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial

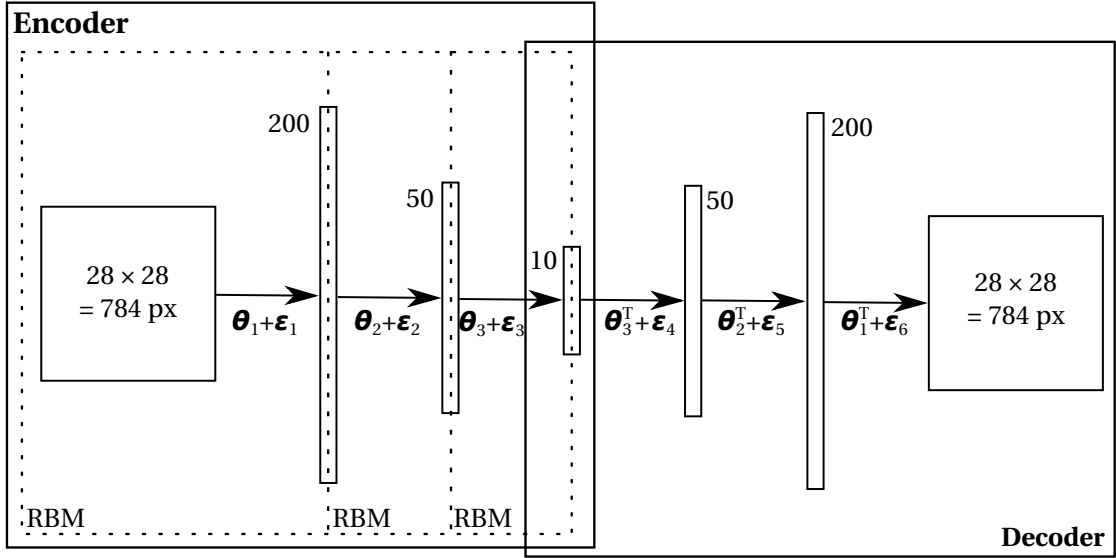


Figure 4.8 – Example of a small AE architecture, taking as input an image stamp of 28×28 pixels, with three layers to the representation layer. First the approximate model parameters of the encoder are learned by separately and sequentially training the RBMs. After this pre-training, the encoder part is built by connecting the RBMs together and inverting the encoder to prepare the decoder. Finally, the AE is finely tuned using a back-propagation algorithm.

somewhat surprising fact is currently not consolidated by theoretical results (see, e.g., [Vincent et al. 2010](#)). Neurons are classical computing units, that linearly combine inputs, add a bias term and apply an activation function.

The incapacitating vanishing problem is mitigated by first finding an approximate solution for the model parameters θ . This implies a pre-training phase, to bring the randomly initialised parameters to a satisfying state for back-propagation. In the pre-training phase, the layers learn, without supervision, one after another using a generalised version of restricted Boltzmann machines (RBMs). RBMs are shallow two-layer networks that are applicable on binary vectors. The terminology for RBMs is a bit confusing as the first layer is called the visible and the second hidden layer. The visible features are propagated to the hidden layer by a forward pass. RBMs learn to reconstruct the visible inputs by undergoing a backward pass. This backward pass is computed with the same weights, but different biases. There is one set of weight for a RBM, but two sets of biases: one with the dimension of the hidden layer, for the forward pass, and the other of the dimension of the inputs for the backward pass. For the \mathbf{v} and hidden, \mathbf{h} , nodes, a joint energy is defined as

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i \in \text{input}} b_i v_i - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_i h_j w_{ij}, \quad (4.21)$$

where b_i and b_j are the biases of the backward and forward pass, while w_{ij} denotes the weights. All these parameters, summarised by the quantity θ , are initialised by drawing from

appropriate distributions¹⁰. For a given training sample, the hidden state is set to 1 with a probability of $\sigma(\mathbf{w}_{ij}v_i + b_j)$, with σ the activation function or zero. The same procedure is repeated to compute a visible state such that an update vector of the weights Δw_{ij} is

$$\Delta w_{ij} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstr}}, \quad (4.22)$$

where $\langle v_i h_j \rangle$ is average and a similar rule can be written for the two sets of biases. The effect of these learning rules is somewhat similar to a gradient descent, and the interested reader is referred to [Hinton \(2002\)](#) for more details. Once the parameters of the layer are optimised, the pre-training of the AE parameters is complete. The activations of the layer are fed to the next RBM as input. The pre-training of the parameters θ_ℓ of layer ℓ cascades down to the code layer, at which point this greedy, unsupervised and layer-wise pre-training of the AE is finished.

The next step is to build the AE. The RBMs are unfolded, and the layers are stacked, keeping weights and the forward-pass biases to construct the encoding part. The architecture of the decoder is inverted with respect to the encoder, while the weights linking a layer to the next are transposed. The biases of the decoder are the RBMs' backward pass biases. In [Fig. 4.8](#), we denoted this transformation of the parameter with the non-formal notation θ_ℓ^T . The weights of the encoder and decoder are still tied, but this constraint is relaxed at that point such that they are allowed to evolve independently.

The fine-tuning of the parameters can now start. This is a supervised procedure as shown in [eq. \(4.20\)](#): AEs must learn the identity function. A cost function penalises poor model parameters, which is typically selected to be MSE, but a cross-entropy-based cost can also be used. The parameters are refined by running a back-propagation algorithm, which yields the minimised parameters $\theta_k + \epsilon_\ell$, where k denotes the index of the layer for parameters determined in pre-training and ℓ the layer index from the input layer to the output layer. When the parameters found at the pre-training stage are good, the fine-tuning time should be limited.

A wealth of regularisation techniques and tricks can be used to improve the performance, reduce the training time or force the representation or the weights to be sparse, which we will describe in [Sect. 4.5.3](#).

4.5.2 Application to MNIST

Let's compare the results of a decomposition by AEs with the PCA technique on the MNIST dataset. AE codes are not ordered, that is there is no ranking of the coefficients in terms of information gain. Unlike PCA, the number of coefficients cannot be chosen post-training. Training time for AE is also substantially longer than PCA. This major disadvantage can be mitigated by the introduction of advanced optimisers and GPUs. We present the results

¹⁰There are diverging opinions. The literature is not unified behind a common distribution.

of a network taking in inputs vector of 784 MNIST pixels to an overcomplete architecture $784 \rightarrow 1000 \rightarrow 500 \rightarrow 250 \rightarrow n_c$, where n_c is the dimension of the code layer. Activation functions in this example are sigmoid for all layers except for the bottleneck layer that has a linear activation function.

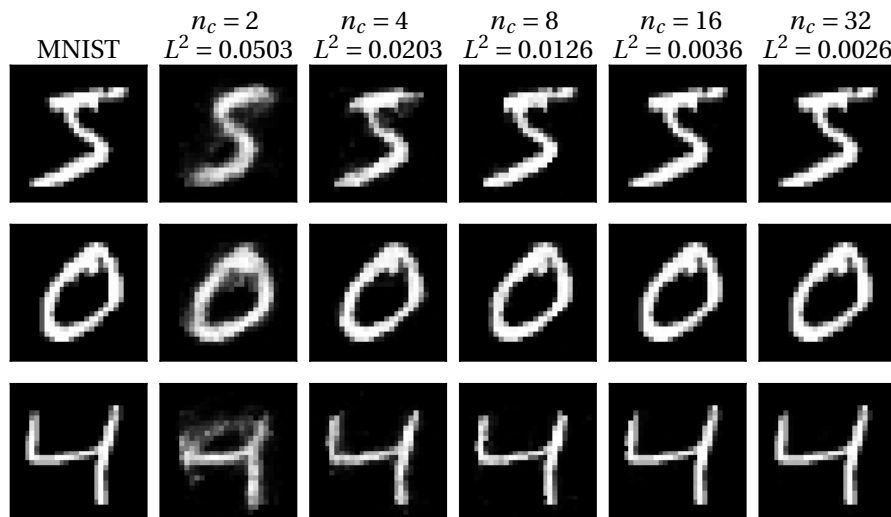


Figure 4.9 – *First column*: original data. *Next columns*: AE transform and inverse transform of MNIST data. The number of coefficients in the code layer is denoted by n_c and L^2 the pixel-to-pixel MSE error.

Figure 4.9 is similar to the MNIST PCA transform for several choice of the number of coefficients (Fig. 4.2). The quality of the reconstruction for three samples is clearly surpassing the PCA performance. The pixel-to-pixel error L^2 with only components is comparable with a PCA result with thirty-two components! The reconstructions at any number of components look much better to the human eye than PCA. Obvious reconstruction errors also occur at a low number of coefficients, like the image of the 4 reconstructed as a 9. A lower number of meaningless reconstruction is reported by the AE algorithm. Changing the architecture of the network do not change significantly the result when its capacity is sufficient to handle the task. Better reconstruction performances in AEs can be attributed to non-linearities and a larger capacity to adjust to complicated datasets. Note that, contrary to PCA, the training procedure is stochastic, thus slightly different representations are learned for multiple runs.

The MNIST digits now cluster better in an AE two-components space using a $784 \rightarrow 400 \rightarrow 100 \rightarrow 2$ encoder. Representing the MNIST data with only two components yields a clearer picture (see Fig. 4.10) than its PCA counter-part (Fig. 4.3). For this reason, classifying a representation that was learned by an AE is usually better than with PCA. The relative performance will of course depend on the nature of the images. There are regions in Fig. 4.10 where the distinction between two classes of data is unclear. The digit 5 seems particularly hard to detect with this architecture as its footprint covers other clusters. To improve this representation, the architecture can be changed (by, e.g., adding more neurons in the bottleneck layer) or the cost function rethought. Note that at this point, we did not constrain the representation, the cost

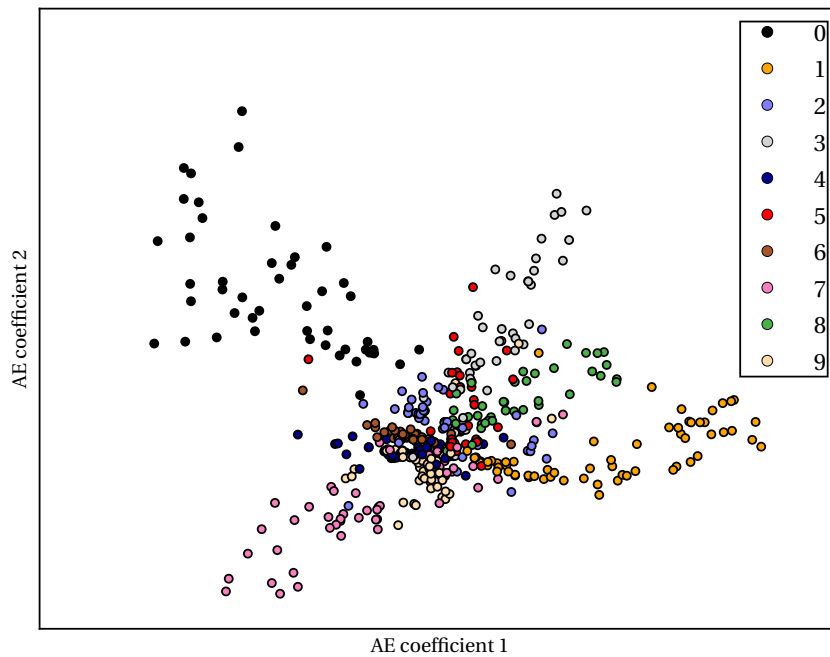


Figure 4.10 – MNIST samples represented in the AE basis coloured by their ground truth. The AE that found this representation has only two neurons in the coding layer.

function relates only the reconstruction data $\tilde{\mathbf{x}}$ to its original version \mathbf{x} . In the next section, we will introduce new terms in the cost function that will put constraints on the representation.

AEs do not have eigenvectors, but their filters can be visualised. There are as many filters as neurons and they are made up of their weights. They are usually visualised in forms of images, when the dimension of the layers allows it. When neurons produce an activation a different than zero, they are called activated. Weights arranged in images like in the filters of Fig. 4.11 show to which region of the image a neuron is sensitive, and with which effect. As it is seen in Fig. 4.11, typical AE filters are somewhat random without structure, even if some blobs seem to appear. The filters in the figure come from the AE that learned the two-code representation; random filters do not imply bad performance. Again, we will see in the next section that AE filters can be forced to learn structures and that it increases the performance. Note however, that the only easily interpretable filters are those of the first layers. Deeper layers are more abstract, thus much more complicated to understand.

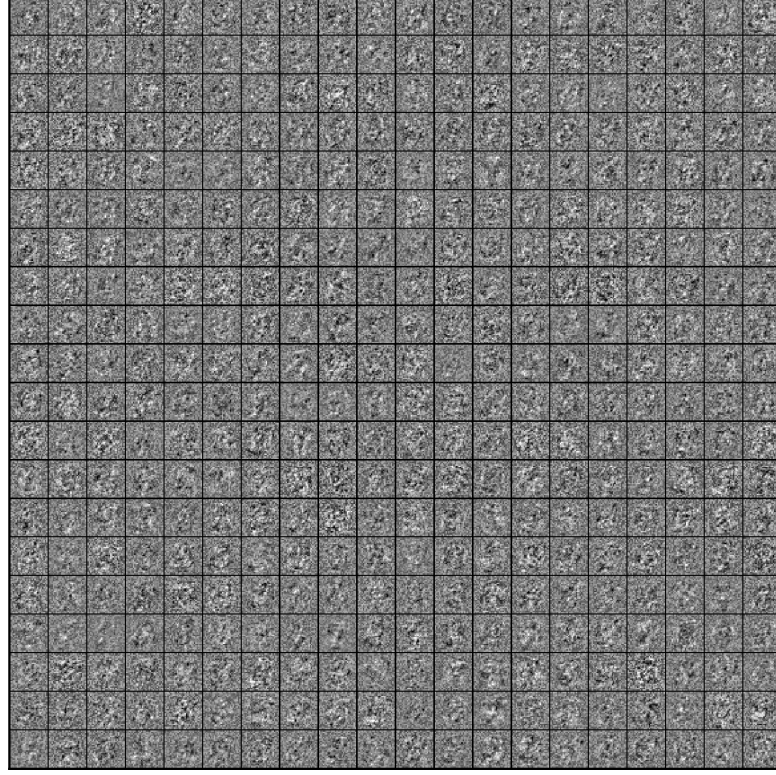


Figure 4.11 – First-layer filters of the $784 \rightarrow 400 \rightarrow 100 \rightarrow 2$ encoder for MNIST. There are 400 filters as there are 400 neurons in that first layer and each filter is 28×28 pixels. Dark pixels are negative weights and white pixels positive.

4.5.3 State-of-the-art improvements

The previous sections treated the case of a classical AE with tied weights, and a MSE cost function. There are however plenty of variations in the AE approach. We describe a few interesting AE flavours in the following.

Denoising AEs (dAEs) are designed to be robust to the partial destruction of the input (Vincent et al. 2008, 2010). The representation learned by the network should capture stable features in the data, i.e. the algorithm should be able to reconstruct the image from only a fraction of the original data. Working with corrupted or low S/N data is a recurrent theme in astrophysical image processing, and spawned research into denoising algorithms (see, e.g., Starck et al. 2010). Denoising AEs work by learning with training samples, $\hat{\mathbf{x}}$, that are degraded (e.g. either by applying a mask or by adding Gaussian noise). The truth target remains however the unperturbed sample, \mathbf{x} . DAEs are tasked with minimising the cost function

$$J_{\text{dAE}} = \sum_{i=1}^N \mathbb{E}_{\hat{\mathbf{x}}_i \sim q(\hat{\mathbf{x}}_i | \mathbf{x}_i)} \{C(\mathbf{x}_i, \tilde{\mathbf{x}}_i)\}, \quad (4.23)$$

with $q(\hat{\mathbf{x}}_i|\mathbf{x}_i)$ the distribution of the corrupted samples given the original data \mathbf{x}_i and C an arbitrary cost function, be it MSE, cross-entropy or something else. The minimisation runs over the expectation \mathbb{E} of the cost function C . In other words, dAEs try to reconstruct the original data from a compressed representation of a corrupted version. At each iteration of the training procedure, the network sees a different realisation of the noise. dAE achieve impressive reconstruction results, even in the presence of a high level of noise. An interesting effect is the distribution of the activation in the filters. Filters learned without noise are basically random, whereas filters of dAE show coherent activations in the form of feature detectors. Different topologically close pixels in the filters activate in neurons, that is activations are organised in blobs that are clearly distinct from one neuron to another.

Contractive AE (cAE) architecture is classical, the difference to an AE is the introduction of a regularisation term in the cost function (Rifai et al. 2011). The Frobenius norm,

$$\|J_f(\mathbf{x})\|_F^2 = \sum_{i,j} \left(\frac{\partial y_j(\mathbf{x})}{\partial x_i} \right)^2, \quad (4.24)$$

where the Frobenius norm is taken over the Jacobian $J_f(\mathbf{x})$ of the output y , which is of dimension $j \in M$. The goal of this penalisation is to make the representation robust to small variations of the input. The difference is that a dAE promotes a robustness of the reconstruction $\tilde{\mathbf{x}} = g(f(\mathbf{x}))$, whereas cAE make sure that the representation $f(\mathbf{x})$ is invariant. dAE does not guarantee an invariant representation, and thus, according to Rifai et al. (2011), cAE might be a better choice than dAE to extract features. In addition, cAE takes an analytical approach whereas the robustness introduced by dAE is built on a stochastic corruption process.

Sparsity penalisation terms can also be applied. Again, the main difference is in the cost function. When most neurons are not activated most of the time (i.e. they return a value close to zero when using an activation function like sigmoid and ReLU), such that only a small number of neurons fire, the network is said to be sparse. The average activation of a given neuron j is defined by

$$\hat{\rho}_j = \frac{1}{N} \sum_{i=1}^N a_j(\mathbf{x}_i). \quad (4.25)$$

The sparsity of the activation can be controlled *via* the constraint $\hat{\rho}_j = \rho$ where ρ is a sparsity parameter, close to zero. This is enforced by adding the penalisation term J_S of the Kullback-Leibler (KL) divergence between the sparsity parameter and the activation mean,

$$J_S(\hat{\rho}, \rho) = \sum_j \text{KL}(\rho \| \hat{\rho}_j) = \sum_j \left\{ \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \right\}. \quad (4.26)$$

A clear advantage of sparse AE is that its filters usually are edge detectors similar to Gabor filters (Lee et al. 2008; Nair and Hinton 2009). As for dAE, the effect of adding constraints is to have more useful filters. Enforcing sparsity in AEs with only linear activation functions can,

for MNIST classification, reach state-of-the-art classification (Makhzani and Frey 2013).

Convolutional AE provide layers that, like in CNNs (Sect. 4.3.4), convolve the input with a learned kernel. Encoders are based on the same building blocks as CNNs. The decoding part consists of the inverse modules: upsampling operators and un-convolution operation¹¹. Classical AEs do not use the spatial relations between the pixel, even though there is some topological structure in natural images that was left unused (Ranzato et al. 2007). Ignoring the structure of the image introduces redundant parameters such that there cannot be a very localised feature (Masci et al. 2011).

Until now, all of the AE flavours we presented were symmetrical: the input and output were of the same dimension. We now include a small **asymmetry** in the network. There are two obvious types of asymmetries: (i) different numbers of layers in the encoder and decoder and (ii) different dimensions of the input and output. The former has very recently demonstrated similar to slightly better results than dAE (Majumdar and Tripathi 2017). We will explore further the latter option, which is also a recent development in the field (Lee et al. 2016). The input layer is no longer of the dimension of the data, but the data plus a small number of features n_f . An example of an asymmetric AE (aAE) would be, when the data is, say, $28 \times 28 = 784$ pixels, $784 + n_f \rightarrow 200 \rightarrow 50 \rightarrow 10 \rightarrow 50 \rightarrow 200 \rightarrow 784$. The idea behind this extension is to include additional information. An example (that will re-emerge later in this manuscript) would be a stamp extracted from an image as the main input and the spatial coordinates of the stamp in the larger image. The cost function is still computed between the reconstructed and original input.

4.6 Evaluating performances

When the parameters of machine-learning algorithms are optimised and validated through, e.g., evaluating the cost function on a validation set, the technique is trained. Throughout this chapter, we used the term performance to describe how well a technique would reach the objective of its task. For ANNs, that objective would be to minimise the cost function, but this error function does not necessarily represent the final objective. In this section, we introduce a few important concepts and metrics that serves as performance measures in this thesis both for regression and classification tasks.

Some simple diagnostics tool help improve a model, like tracing the evolution of the cost function and, also possibly, some measure of performance with training steps. Plotting training and validation performances reveals the generalisation gap that must be driven down to the minimum, and might help identify dead ends in the training (if, e.g., the value of the cost function stays stable with training epochs).

¹¹Deconvolution is not formally the inverse of convolution, but is commonly used in the AE literature to denote the inverse of the convolution layers, which are gradients of the convolution.

4.6.1 Classification metrics

The cross-entropy and even MSE are used to train classification techniques. There are many useful measures of performance in classification, which are sensitive to different effects in the output distribution. To describe these metrics, there are sometimes confusing and multiply-defined terms. We will restrict their use to the minimum, but we need to introduce a few of them now, in a visual format for clarity in a confusion matrix in Tab. 4.1. Note that in this confusion matrix, positive can be interpreted as “the output of the method is 1” while negative translates into “the output is 0.”

Table 4.1 – Confusion matrix for a binary classification and its terminology.

		Ground truth	
		<i>true</i>	<i>false</i>
Prediction	<i>positive</i>	true positive (TP)	false positive (FP)
	<i>negative</i>	false negative (FN)	true negative (TN)

Accuracy describes the localisation of the distribution of samples. It is given as,

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{total number of samples}}, \quad (4.27)$$

so, simply put, the number of correctly predicted samples to the total number of samples. It ignores the repartition of false negatives and false positives. Accuracy is not an appropriate test for applications that require their level of false negative or false positive to be under control. In precedent sections (3.5 and 3.6), we argued that high accuracy was required in shear measurement algorithms. We stress that accuracy in ML and WL does not mean the same thing! In WL, it means that there is no bias in the measure, i.e. that the errors in estimation, averaged over a large number of realisations are consistent with zero. The precision is thus the width of the distribution, defined in the confusion matrix as

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false negative}}. \quad (4.28)$$

The false-positive rate (FPR) describes the probability of rejecting samples,

$$\text{FPR} = \frac{\text{false positives}}{\text{false positives} + \text{true negatives}}, \quad (4.29)$$

while the true-positive rate (TPR) expresses probability of detection,

$$\text{TPR} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}. \quad (4.30)$$

The two-dimensional space in which the FPR and TPR are the two axis makes up a great visual aid to selecting the discriminative criterion. Binary classification methods do not return directly return a binary figure labelling the class. The output is a real (floating point)

value representing the confidence of the method that the sample is belonging to class 0 or 1. Probability is a term that cannot be used here as some of the methods are not properly normalised, so a method that classifies objects must have a decision threshold. When a classifier is represented in a plot with FPR encoding the x -axis and the TPR the y -axis, it forms a curve. This is known as the receiver operating characteristic (ROC), and is a graphical representation of the performance when the decision threshold is varied. A monkey-like algorithm that would randomly classify object lives on the diagonal FPR=TPR. The perfect classifier would be so without the need of a decision threshold (i.e., its predictions would be 1 or 0). This curve would go through the point (FPR= 0, TPR= 1). ROC plots allow for a quick and reliable comparison between different classifiers, marginalising over the detection threshold. The choice of the separating threshold can be based on the ROC by trading off the TPR for the FPR (e.g. [Kleinbaum and Klein 2010](#)).

Area under the curve (AUC) is the integral of the ROC over the FPR. It summarises the ROC as a single scalar. A random algorithm would score AUC = 0.5, while an ideal method would reach AUC = 1. For this metric to be high, the true positive rate and the precision must both be high. For a more in-depth treatment of ROC and its related quantities, we refer to [Fawcett \(2006\)](#).

The F_1 -score is a metric which summarises the performance to one scalar value. For a binary classification, the F_1 -score is defined by

$$F_1 = \frac{2 \cdot \text{true positives}}{2 \cdot \text{true positives} + \text{false negatives} + \text{false positives}}. \quad (4.31)$$

This metric is not only sensitive to the number of correctly classified objects, but also to the number of wrongly classified objects (false negatives and positives). An error-free classification corresponds to $F_1 = 1$, while a completely random decision-making algorithm would have an F_1 score that tends to zero. In an exclusive multiclass scheme, F_1 of the system can be taken as the average over individual F_1 -scores. Individual scores are computed by considering that the classification is binary: the class label or anything else. The confusion matrix can also be expanded to include the correct numbers of entries.

4.6.2 Regression metrics

MSE is the typical cost function for many ANN regressions and measures the quality of the regression and, of course, be used as a basis to gauge quality. This is not the only way, specialised metrics can be design to probe all aspects of the regression. This metric is often denoted by Q .

Developing a method to measure shape of galaxies requires to compare predicted values to the ground truth. This can be done at the level of each prediction by using a MSE error function, but it does not give an overall performance. The overall performance can be estimated, e.g., by comparing the ellipticity components in terms of μ and c factors in constant shear fields (see Sect. 3.2.2). The comparison can be done simultaneously by computing a Q metric or by

Chapter 4. Machine-learning techniques

studying the shear linear bias factors. This measure does not drive the training, but is used to describe the performance of the method. For variable shear, the metric is computed on aperture mass dispersions.

The quantities exposed in the above are the performance metrics used in this thesis, but others are possible, depending on the application. One example would be a penalty when catastrophic outliers are detected.

5 Weak lensing measurements

Overview

In the last chapters, we reviewed the subjects of this thesis from top, cosmology, to bottom, weak lensing data analysis techniques and machine learning. We will now present the work undertaken in detail. The common theme of the different contributions is weak lensing data reduction. The weak lensing analysis pipeline requires many modules. Each of them consists in several algorithms interfaced to one another. As an example, let's take the PSF module. It must first detect point sources in images, extract stamps, test the objects against a multitude of rejection criteria, find a good PSF reconstruction and store it in an useful form for the next block. The present work focuses on two of the blocks: (i) the determination and reconstruction of the PSF, and (ii) the determination of the reduced shear.

We begin by exploring one of those rejection criteria: binary (and multiple) stars. The PSF should be reconstructed from images of observed point sources. For all intended purposes, a (unresolved) star is a point source. Multiple stars are multiple point sources whose signal blend in the final CCD image. The observed object is no longer a point source, and introduces biases in the measurement and reconstruction of the PSF. We describe this effect, which becomes non-negligible in surveys like *Euclid* and propose mitigations in Sect. 5.1.

We then will turn to the problem of finding the spectral class of a star quickly and reliably. As we explained in Sect. 3.2.2, broad-band observations and chromatic effects generate biases. To avoid those biases, the spectra of the galaxy and the PSF should match as best as possible. Determination of the spectrum of a point-source object in the context of *Euclid* will take time as it needs to be observed in multiple bands (see Sect. 3.6). We propose a method to determine the spectrum (or the colour) using only VIS images in Sect. 5.2. In addition to proof-of-concept experimentations, we will show the reconstruction of a colour-magnitude diagram using real HST data.

Once a point-source object has been vetted, it can be used for PSF reconstruction. The observed imaged of a point source must be modelled such that it can be interpolated to the

right image coordinates and spectrum. We choose to infer a PSF model directly from the data. In Sect. 5.3, we teach AEs (see Sect. 4.5) to extract a better representation of the PSF, which includes a spectral component.

After reconstructing the PSF from reliable sources, we can measure galaxy shapes to detect the reduced shear. As discussed in Sect. 3.2.1, this is a hard problem that can be tackled by very different approaches. We propose in Sect. 5.4 a machine learning technique to predict the ellipticity or the shear of galaxies. We constrain the system to minimise biases, which implies using cost function different than the typical MSE.

5.1 The effect of unresolved binaries on PSF determination

Multiple stellar systems are ubiquitous and common in the Galaxy. Typical estimations gauge that 30-40% of stellar systems are binary stars, or multiple-star systems. This fraction increases with the mass of the main star, such that about half of the solar-type stars reside in binary systems. It is inevitable that multiple systems will be imaged by *Euclid*. In this section, we explore whether we can accept images of *unresolved* multiple systems to reconstruct the PSF. In a simplification effort, we restrict ourselves to the very common binary systems, excluding only a small fraction of the systems. We make another assumption: all stars observed are field stars and in their main sequence (MS) phase. We derive expected deviations and reconstruction errors for typical systems and infer error budget contributions as a function of the apparent magnitude of the binary. In addition, we research mitigation techniques and propose methods to flag binaries based on a statistical study of the shape of the source objects.

5.1.1 Multiple stellar systems

In the introduction to this section, we stated that multiple systems were ubiquitous. Star formation occurs by fragmentation and collapse of molecular clouds to the local Jeans mass. Molecular clouds fragment to create regions of higher density. This leads to the formation of protostars by accretion of gas and dust to central regions of the fragments. Stellar formation is illustrated in Fig. 5.1. We recommend the outstanding [Maeder \(2009\)](#) textbook for a much more in-depth approach.

Simple numerical hydrodynamical studies (e.g., [Bate 2009](#)) of the collapse of unstable and turbulent molecular clouds yield the correct fraction of multiple systems. If the multiple stellar population is reproduced with simple codes, it hints at a very common process ([Reipurth et al. 2014](#)). Close binaries – i.e. with semi-major axis $a \lesssim 100$ AU – tend to form from orbital decay of wider binaries through dynamical interaction, accretion and interaction of with the proto-stellar disks ([Bate 2012](#)). The orbital decay of the wide binaries to closer-in systems theoretically suggests that very wide binaries (separations $a \gtrsim 1000$ AU) represent a small proportion of the multiple systems population.

5.1. The effect of unresolved binaries on PSF determination



Figure 5.1 – Sketch of stellar formation. (a.) Molecular cloud in which gravity acts as a contracting force while gas pressure, turbulence, magnetic fields and radiative pressure oppose the collapse. (b.) The cloud fragments in multiple clumps under the effect of gravity. Due to turbulence and the initial density distribution there is a hierarchy of clumps. (c.) Individual clumps collapse to form clumps, some of the clumps merge and thus contain multiple proto-stars. (d.) A supernova explosion terminates star formation. Some of the lowest mass stars were ejected through N -body interactions and multiple systems (in red) emerge.

Because of the gravitational interactions between the stars, ejections or capture of system members are possible. Systems with large mass ratio are less dynamically stable than systems with similar mass members. Close systems tend to exhibit very similar masses, because low mass members would have been ejected from the system by the time they become T Tauri stars (White and Ghez 2001). T Tauri are young, pre-MS, stars with too cold of a core to induce hydrogen fusion and are powered by energy released from the collapse.

Multiples systems are important in many aspects of astrophysics such as stellar formation, stellar evolution or type Ia supernovæ to name a few. Their central role motivated many studies and surveys. To determine the fraction of binaries, one has to systematically survey a restricted volume of space. Solar-type stars are sufficiently both frequent and easy to observe to be prime candidates for early studies. Early searches were restricted to low numbers of stars, hence leading to incomplete samples (e.g. Abt and Levy 1976; Marcy and Benitz 1989; Duquennoy et al. 1991). Systematic searches demand very long telescope time, and thus, even if recent searches collected more data, the number of well characterised stars and multiple systems is still limited to a few hundreds (Raghavan et al. 2010; Milone et al. 2012; Cummings et al. 2014; Nardiello et al. 2015; Riddle et al. 2015; Rodriguez et al. 2015). Duchêne and Kraus (2013) compiled the available data and summarised the current knowledge on the multiplicity properties. The Duchêne and Kraus (2013) review will be the reference for our binary population simulations.

There are a few important quantities to describe multiplicity properties. The first is the frequency of multiple systems, MF , a number below one, but the second, the companion frequency, CF , that is the average number of companions to the main member, can exceed one. The orbital semi-major axis, a , can be interpreted as the physical separation; a relates to the orbital period, P . An important quantity that derives from the semi-major axis is the angular separation, α , i.e. the separation measured off the image. It depends on the semi-major axis, on the distance to the binary system and on the orientation of the system with respect to the line of sight. The distribution of semi-major axis is modelled by a log-normal with parameters pivot period \bar{P} and $\sigma_{\log P}$. There are other models to predict the distribution of semi-major axis, and distinguishing between them would favour certain formation models. The interpretation

of the observations, however, is complicated by the dynamical evolution of the systems, which impact the orbital period. The mass ratio, q , is defined as $q = M_{\text{companion}} / M_{\text{main}} \leq 1$, whose distribution can be modelled on a power law with index γ . A quantity that we will neglect in the following studies is the eccentricity of the system, e . Contrast is the difference of magnitude Δm between the companion (or secondary) and the main (sometimes also called host and primary): $\Delta m = m_{\text{companion}} - m_{\text{main}} \gtrsim 0$. Note that we use m for magnitude and M for mass. We summarise the main properties of multiple systems as:

1. The more massive the primary star, the larger the probability of a companion.
2. The more massive the primary star, the lower the mass ratio power-law index γ . Massive stars tend to host companions of lower relative masses than low-mass stars.
3. The more massive the primary star, the broader the physical separation.
4. The mass ratio q is, on average, close to one, because systems with low q are less stable.

5.1.2 Unresolved binaries-induced bias in PSF determination

PSF determination and reconstruction require observations of point sources. A solar-like star observed from a distance of 1 pc subtends an angular size of about $5 \cdot 10^{-3}$ arcsecond, while the *Euclid* VIS pixel size is 10^{-1} arcsecond, so its surface is completely unresolved. The closest star to the Sun is located 1.33 pc, but to reconstruct *Euclid* PSFs, stars located several thousand pc away will be observed, with completely negligible angular sizes (the saturation limit for the survey exposures is $i(AB) \sim 18$ mag!, [Cropper et al. 2016](#)). Single stars are thus ideal candidates.

Multiple stars appear as superposition of several point sources. Their apparent separation is small, but non-null, thus the image formed on the CCD is no longer due to a point source, but two point sources located very close to each other. The resulting image is a blend of light of the two stars and, mathematically speaking, no longer the PSF. Multiple system images should be excluded, otherwise there is a risk of reconstructing a biased PSF. Say that we can identify all binary objects if their angular separation is at least half a VIS pixel. It would mean that typical systems (i.e., $a = 50$ AU and contrast of $\Delta m = 0$ mag) could be identify and flagged if their distance to the Earth is $d \lesssim 2.5$ kpc. That relatively small distance unfortunately leaves a large number of systems undetectable. The model of a definite detection threshold at half a pixel is an assumption that will be revisited in Sect. 5.1.5 and 5.1.6. The errors that can be tolerated in *Euclid* PSF reconstruction are specified by the scientific requirements and reproduced in Tab. 3.1. In an effort of clarity, they are repeated here:

$$\begin{aligned} \sigma(e_i) &< 2 \cdot 10^{-4} && \text{(PSF ellipticity stability),} \\ \sigma(R^2) / \langle R^2 \rangle &< 1 \cdot 10^{-3} && \text{(PSF size stability).} \end{aligned} \tag{5.1}$$

These figures are the maximal total errors, but they are broken up into error budgets (to be

5.1. The effect of unresolved binaries on PSF determination

consulted in [Massey et al. 2013](#); [Cropper et al. 2013](#)). The contribution of the unresolved binaries was overlooked when the budgets were established on the grounds that multiple systems would be easily identified and removed. Estimations of the number of available objects (e.g. in [Cropper et al. 2013](#), and other PSF *Euclid* conferences contributions) take into account that $\sim 30\%$ of stars cannot be used owing to blending with galaxies light, multiplicity or other disqualifying artefacts. Multiple systems that are identifiable are part of that estimate, but not unresolved binaries. We will demonstrate in the few next paragraphs that unresolved binaries can indeed lead to significant biases.

To highlight the induced bias, we play with a toy-model. We simulate observations of a binary system on an image stamp with the GalSim library ([Rowe et al. 2015](#)). We assume a Gaussian PSF. We place two point sources in the stamp: one at the centre, representing the main star, and another at position at most half a pixel away to simulate unresolved binaries. The resulting image is measured using a weighted second order scheme ([Hirata and Seljak 2003](#)), implemented in GalSim (see Sect. 5.4.4 for more details). The effect on the shape and size of the profile varies with contrast and relative position. The variation with contrast is straightforward to grasp: a higher contrast means that the companion star is dimmer, thus the deformation of the profile is smaller. A white dwarf (WD) in a binary system with a solar-type star essentially does not influence the measured profile as the ratio of its luminosity to the solar-type star is very small. Large separations have a larger effect than smaller separation. The profile measured off a binary is always larger than the PSF. The image of unresolved binaries can be up to a few percent different than the original size, i.e. $\Delta R^2 / R_0^2 \lesssim \mathcal{O}(1\%)$. We use the subscript 0 to denote the intrinsic value of a parameter.

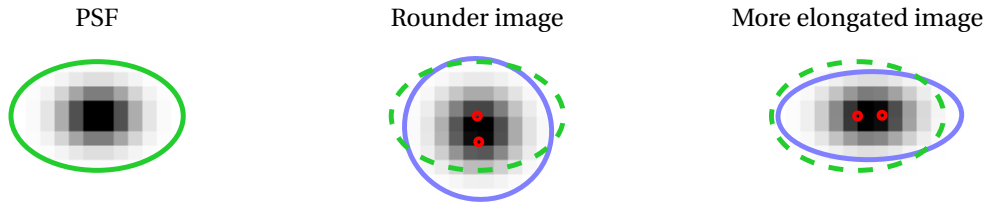


Figure 5.2 – Effect of null-contrast binaries on the measured profile. The ellipses highlight the light profile (but slightly exaggerated the ellipticities). Green is for the PSF and blue for the binary images. Red dots show the position of the primary and secondary stars. Note also that the separations are larger than the 0.05 arcsec detection threshold.

The resulting shape depends on the anisotropy of the PSF. The image can be rounder or more elongated as illustrated in Fig. 5.2. When the two stars are perpendicularly aligned with the anisotropy of the PSF, there is a relative position that induces a round image. The ellipticity of the binary image first reduces to a round profile and then increases as the separation between primary and secondary stars increases. When the stars are aligned with the anisotropy, the image is more elongated than the PSF. Both the ellipticity and the orientation of the image can thus significantly change. The possible configurations of relative position star-companion for a contrast $\Delta m = 0$ are systematically tested.

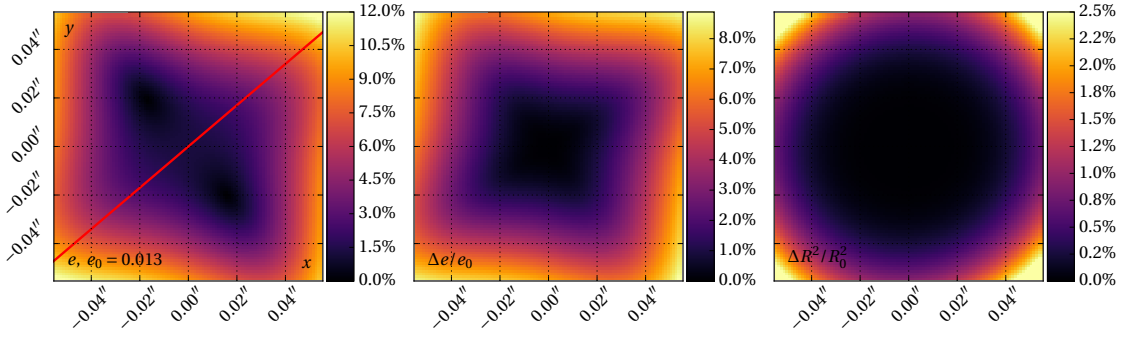


Figure 5.3 – Reconstruction errors for a binary star. The primary star is centred on (0,0) and its companion stars is at the relative position shown in the axis. The contrast between two stars is zero magnitude. The PSF is elongated in the e_2 direction as shown by the red line. (*Left.*) Ellipticity modulus e . (*Middle.*) Relative error $\Delta e/e_0 = e/e_0 - 1$. (*Right.*) Relative size error.

Figure 5.3 presents the result of the systematic exploration of the relative position space for a PSF aligned with the e_2 direction. The PSF for this experiment is a *Euclid*-like VIS PSF, that exhibits complex structures as shown in Fig. 5.15. The measured ellipticity of the system clearly shows the effect discussed previously: the ellipticity decreases perpendicularly to e_2 , i.e. to the anisotropy, until zero and then increases to values larger than the original PSF. The largest errors are detected in the direction of e_1 , because true e_1 of the PSF is small; it changes fast when a companion is added in its direction. The size increases with the separation irrespective of the relative position. The magnitude of the change in size is small relative to the effect on shape. We will re-use this fact when we describe the method to detect the unresolved binaries, in sections 5.1.5 and 5.1.6.

Errors, both in shape and in size, are larger than the requirements (eq. 5.1) when reconstructing from one profile. The next step is now to study how do the binaries affect a profile averaged over many systems. That is, we ask the question whether these worrying conclusions can be dismissed by averaging the observed profiles.

We explore this question by designing another simple experiment. Many realisations of the same binary system are generated. The relative position of the two stars is changed, but the separation is kept constant. The contrast between the two stars is null. Mock binary images are aligned on their centroid and co-added. This co-addition of point sources to increase the signal over the noise can only be performed in a localised region, otherwise the resulting PSF will be biased. Here, we co-add 5 000 binary systems and compute the error on the reconstructed profile. We return to a simple Gaussian PSF and extreme values for the ellipticity to artificially increase the magnitude of the effect.

In Fig. 5.4, we present two examples. A perfectly round ($e_1 = e_2 = 0$) PSF will not change its shape when measured from binary profiles. However, the size increases. The size of the profile increased by $\sim 30\%$ in this round PSF example. The second example is a very elongated PSF

5.1. The effect of unresolved binaries on PSF determination

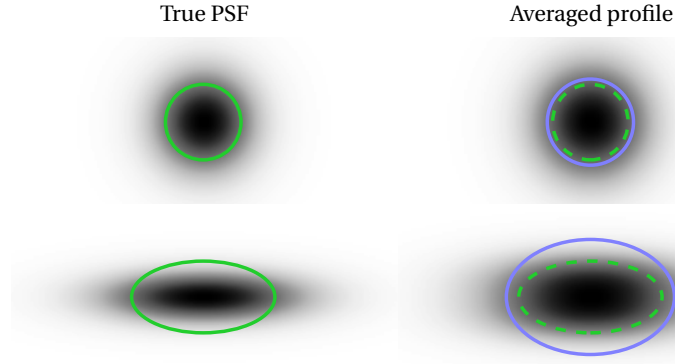


Figure 5.4 – Observed profiles for (*Left.*) the PSF and (*Right.*) averaged profile over 5000 stars. The resolution is increased for visibility. The green ellipses show the contour of the PSF, while blue ellipses represent the averaged profiles.

($e_1 = 0.6, e_2 = 0$). The resulting profile is larger, but also rounder. The ellipticity component e_1 is decreased by $\sim 30\%$. The relative position of the binaries is independent of the anisotropy of the PSF. When companions are randomly distributed about the host star, the mean separation is a constant, and does not depend on the relative angular position. This independence of the binary tends to make the recovered profile rounder. Going back to the question: is the bias induced by unresolved binaries averaged out by co-adding objects?, the qualitative answer is no. We relax the two assumptions made in this paragraph (constant separation and zero contrast) to study averaging light profiles across different binary configurations.

An estimate of the bias can be derived analytically for idealised cases, with a large number of systems. The light profile of the companion star can be written as function of the main star *via* $I_c(x, y) = f I_m(x - \delta x, y - \delta y)$ and total intensity of $I_c = f I_m$. The quantity f is therefore the ratio of the intensities. The centroid $\bar{\mathbf{x}}$ of the binary star is

$$\bar{\mathbf{x}} = \frac{I_m \bar{\mathbf{x}}_m + I_c \bar{\mathbf{x}}_c}{I_m + I_c} = \bar{\mathbf{x}}_m + \frac{f \cdot \delta \mathbf{x}}{1 + f}. \quad (5.2)$$

We now compute the second order moments using the above centroid as the reference point in the xx direction:

$$\begin{aligned} q_{xx} &= \frac{\int \left(x - \bar{x}_m - \frac{f \delta x}{1+f}\right)^2 I_m dx dy}{\int I_m dx dy + \int I_c du dv} + \frac{\int \left(u - \bar{x}_m - \frac{f \delta x}{1+f}\right)^2 I_c du dv}{\int I_m dx dy + \int I_c du dv} \\ &\stackrel{u \rightarrow x - \delta x}{=} \underbrace{\frac{\int \left(x - \bar{x}_m - \frac{f \delta x}{1+f}\right)^2 I_m dx dy}{I_m(1+f)}}_{\text{}} + \frac{\int \left(x - \bar{x}_m - \frac{f \delta x}{1+f} - \delta x\right)^2 f I_m dx dy}{I_m(1+f)} \\ &= q_{xx,m} + \frac{f \delta x^2}{(1+f)^2}. \end{aligned} \quad (5.3)$$

Similarly, for the yy and xy direction, one derives,

$$q_{yy} = q_{yy,m} + \frac{f\delta y^2}{(1+f)^2}, \quad (5.4)$$

$$q_{xy} = q_{xy,m} + \frac{f\delta x\delta y}{(1+f)^2}. \quad (5.5)$$

We now co-add N binary profiles. Using the relation between the ellipticities and the moments (eq. 3.2), we can derive the expectation value (denoted by $\langle \cdot \rangle$) of the ellipticity component e_1 ,

$$\langle e_1 \rangle = \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy} + \langle f r^2 / (1+f)^2 \rangle}, \quad (5.6)$$

with $r^2 = \delta x^2 + \delta y^2$. Note that as the relative positions of the companion star are random, thus for large enough number of stars in the stack N_\star , $\langle \delta x^2 \rangle \sim \langle \delta y^2 \rangle$. The term $\langle f r^2 / (1+f)^2 \rangle$ is much smaller than the rest of the denominator, we can safely neglect it. The expected bias, i.e. the difference between the measured and intrinsic ellipticities $\delta e_i = e_i - e_{i,0}$, for $i = 1, 2$, yields

$$\langle \delta e_i \rangle \simeq -e_{i,0} \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle, \quad (5.7)$$

where R_0^2 is the intrinsic size of the PSF defined as $R_0^2 = q_{xx,0} + q_{yy,0}$. These results confirm the observations made in the toy examples in the previous paragraphs. As binary stars are introduced, the measured ellipticity components shrink as a function of the intrinsic ellipticity component, the size of the PSF, R_0^2 , and parameters of the binary population: the relative intensity, f , and the separation, r . This tends to make the profile rounder than the PSF. Large PSFs will be less affected by unresolved binaries. Effects on ground-based PSFs are therefore negligible. The bias on the size is written

$$\frac{\delta R^2}{R_0^2} \simeq \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle. \quad (5.8)$$

We confirm the result that the size of the profile increases. In the above derivation, we worked in the $N \rightarrow \infty$ limit. When the number of systems is small, the bias cannot be described by the above equations, and is dominated by the shot-noise error.

In this section, we demonstrated that unresolved-binary induced biases can be significant when determining the PSF from a single object. We also verified that the effects cannot be averaged out on simple numerical experiment and with an analytical approach. We now turn to quantify this binary bias arising from *Euclid*-like observations.

5.1.3 Unresolved-binary bias in *Euclid* observations

To simulate *Euclid*-like observations of stars, a realistic population of single and multiple systems is required. The Besançon model of the Galaxy (BMG) (Robin et al. 2003) produces a realistic catalogue. We simulate multiple *Euclid* pointings in the Galaxy and select stars in the magnitude range $16 \leq i(AB) \leq 24.5$. This model does not however include multiple systems, we had to add them at the catalogue level. We implement the empirical data summarised in Duchêne and Kraus (2013) by evaluating the probability of each star in the catalogue to have a companion. The evaluation is restricted to one companion as the number of stars that could potentially be in systems with more than two members is negligible, $\mathcal{O}(0.2\%)$. The catalogue of stars now contains an additional column: the probability of the star being the host member of a system. For each star, we add a companion with probability given in the catalogue. A semi-major axis a and a mass ratio q are generated from the probability distributions described in Duchêne and Kraus (2013). The mass ratio is transformed to a contrast using a mass-luminosity relation (Habets and Heintze 1981) and the object flux are normalised such that the combined magnitude matches the catalogue. We assume random orbital plane orientations with respect to the observer and random phases. This consideration reduces the measured angular separation. At this stage, the catalogue lists a realistic stellar population in terms of masses (thus luminosity) and of binary population. Images are generated using GalSim, and a *Euclid* VIS-like pixel scale. We make the following simplifications: the PSF is Gaussian, with *Euclid*-like ellipticities and size and the images are noiseless. This simplifies the analysis and reduces the number of simulations as there is no noise bias. Simulating more complicated PSF models does not influence the binary bias significantly (see Sect. 5.1.4). The images of the objects ($\sim 65 - 70\%$ of the objects are single stars) are co-added on their centroid. The outputs of a simulation run are (i) a catalogue of stars for a given pointing with both single and binary systems, and (ii) a stacked image of all the stars. Many runs are performed to average over the binary population and to test different PSF ellipticities. In summary, we follow this procedure:

1. Load a BMG realisation with *Euclid*-like constraints on the field and magnitude cuts;
2. *for each star*: evaluate the probability of being in a binary system;
if the object is a binary system:
 - (a) Draw a mass ratio q , derive the contrast, and normalise the magnitude;
 - (b) Draw the semi-major axis a , and assuming a circular orbit, draw a orbital plane, and a phase angle. Compute the relative position in units of pixels.
3. Draw noiseless images of the systems, and co-add them;
4. Measure the resulting profile, compute the biases.

We emulate stacking, a common method to increase the signal-to-noise (S/N) of the image without taking longer exposures. There is an important caveat when stacking point sources:

the PSF spatial variability. If all the objects across the field-of-view are co-added, the final object is the mean PSF profile. This reconstruction would, without doubt, not be compliant with the stringent *Euclid* requirements. Stacking can only be a local procedure. When the PSF is also stationary in time, objects from different exposures can be co-added, increasing the number of point sources significantly. In the stacking simulations, we assume that all the objects are observed at the same image coordinates¹. We are now in a situation where we can study the unresolved binary-induced bias in a realistic population, unlike the simple experiments of the previous section. The binary distribution functions are crudely known, however. We deal with the large uncertainties by artificially varying the binary fraction.

At this point, we highlight the subtle difference in meaning between the error and the bias on the reconstructed shape parameters. For the ellipticity components, they are

$$\begin{aligned}\sigma(e_i) &= \sqrt{\frac{1}{N_s - 1} \sum_{s=1}^{N_s} (e_{i,s} - e_{i,0})^2} \rightarrow \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle e_{i,0} \quad (\text{ellipticity error}), \\ \langle \delta e_i \rangle &= e_i - e_{i,0} \simeq -e_{i,0} \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle \quad (\text{bias on the ellipticity}),\end{aligned}\tag{5.9}$$

where N_s is the number of simulated stacks, each containing N_\star stacked systems. Note the sign difference between bias and reconstruction error. The reconstruction error only tends to the value of the bias. We thus have to average over many stacking runs (and their many objects per stack) to obtain the error. Similarly,

$$\begin{aligned}\sigma(\delta R^2)/R_0^2 &\rightarrow \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle \quad (\text{error on the size}), \\ \frac{\delta R^2}{R_0^2} &\simeq \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle \quad (\text{bias on the size}).\end{aligned}\tag{5.10}$$

This distinction was done because the *Euclid* constraints on the PSF are given in terms of stability, that this in terms of the errors. We make experimentations with three PSFs: (i) a round profile, (ii) anisotropies aligned with e_1 , and (iii) anisotropies aligned with e_2 . We measure the reconstruction errors and biases for different fractions of binary stars and different numbers of objects per stack ($40 \leq N_\star \leq 10^4$) to analyse the shot-noise effect. We set the number of realisation of the stack to $\mathcal{O}(10^5)$.

We detail the results in Paper I (Sect. 5.1.4). Here, we only highlight some of the conclusions. First, the biases derived from analytical computations (eqs 5.7-5.8) are confirmed by the simulations, and depend on binary population parameters, and the intrinsic PSF shape and size. Second, the reconstruction error is very much affected by shot-noise for a low number of stars in the stack, N_\star . The reconstruction error can be predicted by the analytical results only when $N_\star \sim \mathcal{O}(10^3 - 10^4)$. These analytical predictions are thus optimistic values for the reconstruction error. Third, assuming $N_\star \rightarrow \infty$, we measure the error as a function of the

¹To be exact, the centroid is always at the same coordinates, not the host or secondary stars.

5.1. The effect of unresolved binaries on PSF determination

apparent magnitude. Bright stars, $i(AB) \lesssim 19$ mag, cause the most bias as they are nearby binaries with, on average, a large angular separation. Objects in the range $20 \lesssim i(AB) \lesssim 21$ mag still contribute up to $\sim 10\%$ of the budgeted PSF reconstruction errors. These statements lead to the conclusion that PSF calibration fields for VIS should be carefully selected, and taken preferably at faint magnitudes.

Unresolved binaries-induced biases are by no mean limited to PSF determination. In a recent article, [El-Badry et al. \(2017\)](#) highlighted a bias when fitting stellar spectra. In spectroscopy, two sources are unresolved when they both enter the same spectrograph fiber. The observed spectrum is thus the sum of the individual spectra. Close systems can be detected through their relative velocity Δv . It can be sufficiently high to be detectable with radial velocity techniques or by comparing multiple epoch spectra. Wide binaries have large semi-major axis which translate into long periods and small $\Delta v \lesssim 1$ km/s. Such small line-of-sight offset velocities cannot be detected. Fitting a single star spectrum on a binary system spectrum leads to a underestimation of the temperature and other stellar parameters, which can reach up to 10%. Unresolved binaries could bias stellar evolution diagrams and be, at least partly, for a unphysical upturn in the MS as seen in certain stellar parameter spaces and observed in spectroscopic surveys (see e.g., [Sharma et al. 2018](#)). Concerns over the biases induced by multiple systems are not new. When discussing the [El-Badry et al. \(2017\)](#) paper with Lastro's Pierre North he recalled that forty years ago worried researchers already explored similar effects at Geneva Observatory.

5.1.4 *Paper I: Evaluating the effect of stellar multiplicity on the PSF of space-based weak lensing surveys*

Unresolved binary-induced biases are important in the context of the *Euclid* weak-lensing survey. What we exposed in the above was the core material for Paper I, ([Kuntzer et al. 2016a](#)). The goal of this paper was to make the *Euclid* weak-lensing community aware of this bias and to stir the development of strategies to avoid or mitigate its effects on the PSF reconstruction. The material of this paper and the mitigations were presented at several *Euclid* meetings. We reproduce this Astrophysics & Astronomy research article in this section.

(See next page.)

Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys

T. Kuntzer, F. Courbin, and G. Meylan

Laboratoire d'astrophysique, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny,
 1290 Versoix, Switzerland
 e-mail: thibault.kuntzer@epfl.ch

Received 10 July 2015 / Accepted 25 November 2015

ABSTRACT

The next generation of space-based telescopes used for weak lensing surveys will require exquisite point spread function (PSF) determination. Previously negligible effects may become important in the reconstruction of the PSF, in part because of the improved spatial resolution. In this paper, we show that unresolved multiple star systems can affect the ellipticity and size of the PSF and that this effect is not cancelled even when using many stars in the reconstruction process. We estimate the error in the reconstruction of the PSF due to the binaries in the star sample both analytically and with image simulations for different PSFs and stellar populations. The simulations support our analytical finding that the error on the size of the PSF is a function of the multiple stars distribution and of the intrinsic value of the size of the PSF, i.e. if all stars were single. Similarly, the modification of each of the complex ellipticity components (e_1, e_2) depends on the distribution of multiple stars and on the intrinsic complex ellipticity. Using image simulations, we also show that the predicted error in the PSF shape is a theoretical limit that can be reached only if large number of stars (up to thousands) are used together to build the PSF at any desired spatial position. For a lower number of stars, the PSF reconstruction is worse. Finally, we compute the effect of binarity for different stellar magnitudes and show that bright stars alter the PSF size and ellipticity more than faint stars. This may affect the design of PSF calibration strategies and the choice of the related calibration fields.

Key words. gravitational lensing: weak – methods: data analysis – binaries: close

1. Introduction

Weak gravitational lensing is one of the main cosmological probes to study dark energy and dark matter (e.g. Weinberg et al. 2013; Frieman et al. 2008). In practice, this statistical method requires measuring the shape of billions of faint and small galaxies with the least possible contamination by the instrumental point spread function (PSF). Space-based wide field surveys provide the highest quality check to perform these challenging measurements. Two missions are either under construction or in project to conduct these wide field optical and near-IR surveys: the ESA *Euclid* satellite¹ (Laureijs et al. 2011) to be launched in 2020 and the NASA Wide Field InfraRed Survey Telescope (WFIRST; e.g. Spergel et al. 2015).

Even with the high spatial resolution of a space telescope, any weak lensing experiment must account for the small but non-negligible instrumental PSF, which needs to be removed accurately. The weak lensing distortions are an order of magnitude smaller than the PSF effects on the shape of the observed galaxies (e.g. Kilbinger 2015). To achieve the required accuracy, the PSF must be characterised using stars in the field of view. However, in practice, the high spatial resolution needed to resolve faint galaxies also means that double or multiple stars may significantly affect the PSF measurement, i.e. its size and ellipticity. The goal of the present work is to assess the bias and error introduced by stellar multiplicity on the PSF determination, which is crucial for weak lensing experiments (e.g. Cropper et al. 2013; Massey et al. 2013).

Binary and multiple stellar systems are very common (see review by Duchêne & Kraus 2013, hereafter DK13); they formed

in and their characteristics depend on the initial environment and mass (Reipurth et al. 2014). Several volume and magnitude-limited surveys were dedicated to the search of binary stars, allowing a rigorous construction of catalogues with reliable distribution inferences beginning with the work of Duquennoy & Mayor (1991) and compiled by DK13. Nevertheless, large uncertainties remain in the distribution of the multiple system parameters, even in the solar neighbourhood, reflecting the difficulty of such surveys in accounting for the properties of multiple stellar systems in detail. The frequency of multiple system varies drastically depending on the mass of the primary object, with almost 100% of the most massive stars being multiple systems and on average 35–40% of main-sequence stars being at least binary (DK13).

In this paper, we simulate the error in the reconstruction of the PSF arising from stellar multiplicity. The information on the distribution of multiple systems and their properties (separation, contrast) is fairly scarce but remains sufficient to simulate a space-based survey. For this purpose, we use the Besançon model of the Galaxy (BMG; Robin et al. 2003) in combination with the binary star fraction characteristics drawn from observed stellar distributions summarised in DK13. These data allow us, first, to build mock catalogues with realistic stellar properties and, second, to simulate space-like images of stars. With these simulations, mocks, and images, we derive the bias and reconstruction errors on the PSF width and ellipticity. We also derive analytical predictions and show that the analytical predictions agree very well with the results of the simulations, even for a realistic space-like PSF with complex structures and spikes.

We first study the effect of one single companion as a toy model of the problem in Sect. 2. In Sect. 3, a formalism is

¹ <http://www.euclid-ec.org/>

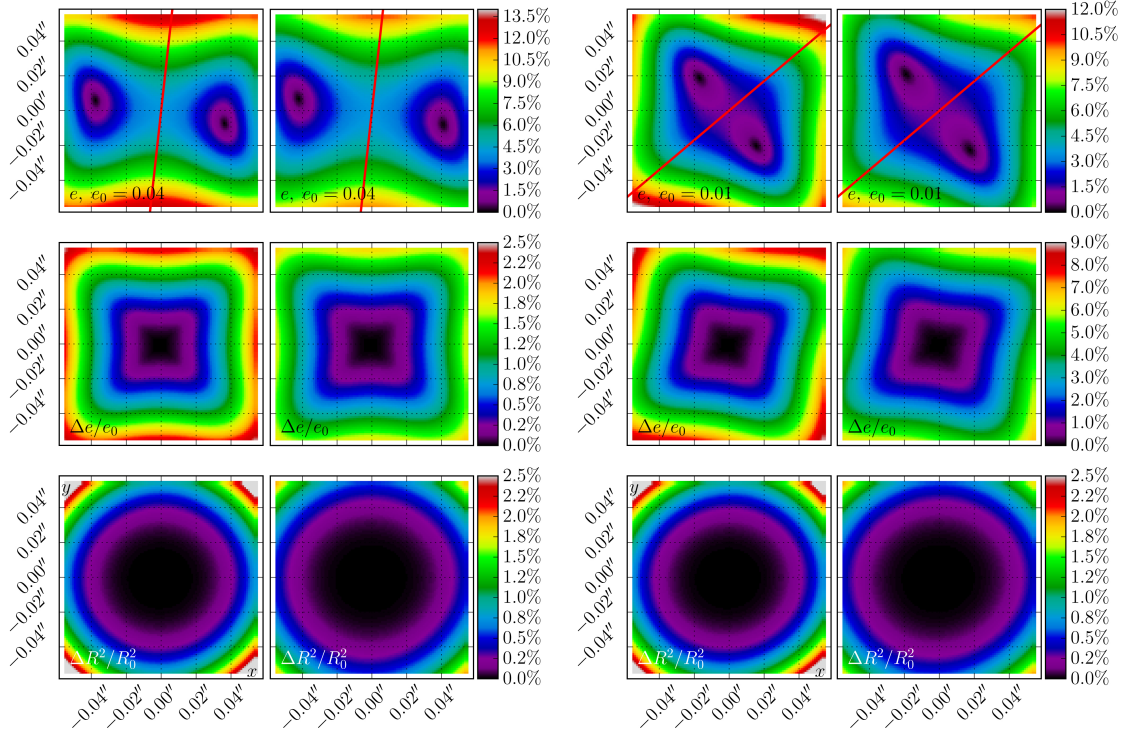


Fig. 1. Reconstruction errors for a single binary star with the main component centred on (0,0). Each point on the map shows the measurement of the quantity labelled in each panel when a single companion star is located at that position. The labels are in arcsecond. The *first column on the left* shows the effect for a PSF aligned along e_1 and a contrast of $\Delta m = 0$ between the main star and its companion. The *next column* is for the same PSF shape and orientation, but for a contrast of $\Delta m = 1$. The block of *six graphs on the right* is for the same two contrasts, but for a PSF aligned along e_2 . The red line represents the orientation of the PSF. *From top to bottom*, the total ellipticity (star + companion) of the PSF, the relative error on the ellipticity due to the companion star, and the relative error on the size are shown.

developed to analytically compute the expected PSF reconstruction error and bias. In Sect. 4, we discuss the current knowledge of the parameters of multiple systems in the Milky Way, describe how we construct the stellar population, and detail the simulations that support our analytical findings. We present the results of the simulations and their parameters in Sect. 5 and conclusions are drawn in Sect. 6.

2. Reconstruction error of PSF for a single binary star

The typical physical separations in binary stars are of the order $a = 50$ AU and typical light contrast is between 0 and 1 mag. With the resolving power of a *Euclid*-like telescope, $0.1''$, binary stars can be resolved up to distances of $d = 2.5$ kpc. These resolved binaries can be identified and removed from samples of stars to be used for the PSF reconstruction. However, most binaries cannot be explicitly identified and still affect the PSF determination, which is an effect that we illustrate in this section.

We generate a simple simulated image that uses a space-like PSF with the GaLSim software (Rowe et al. 2015). This simulation is composed of a main star located in the centre of the image to which we add a companion star at a given position within the central pixel. We then measure the error in the reconstruction of

the PSF of the double star and compare its ellipticity and size to the values found for a single star.

The resulting errors are presented in Fig. 1 for two different ellipticities and for two different contrasts in magnitude, $\Delta m = 0$ and $\Delta m = 1$. Each PSF has its major axis along e_1 and e_2 , respectively. The colour maps in Fig. 1 show the ellipticity of the reconstructed PSF, the relative errors on the ellipticity, $\Delta e/e_0$, and on the size, $\Delta R^2/R_0^2$, where e_0 and R_0^2 are the intrinsic ellipticity and size of the PSF in the absence of a companion.

Two effects are striking. First, the relative errors can easily reach 1% in ellipticity when the companion star is as close as $0.01''$ from its host. This figure is reached when the contrast in magnitude is $\Delta m = 0$. Even for a larger contrast of $\Delta m = 1$, the error remains of the same order of magnitude. Similarly large effects are seen on the size measurement. Second, the shape of the error maps for the ellipticity depend on the PSF orientation and ellipticity, while the error maps for the size show no obvious dependence on the intrinsic PSF ellipticity.

The above test suggests that the presence of binary stars can have an important impact on the PSF reconstruction. They either must be avoided or their effect must be accounted for in the error budget of the mission. While in principle the *Gaia* catalogue is expected to be complete up to $V_{\text{lim}} \sim 20\text{--}25$ mag (Robin et al. 2012), binary and multiple stars catalogues will not be complete for the small separations in which we are interested in this work

T. Kuntzer et al.: Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys

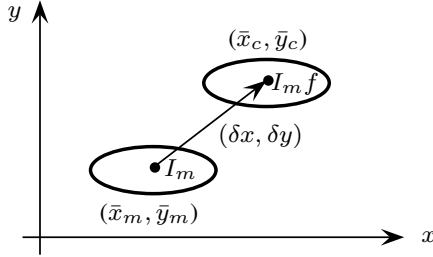


Fig. 2. Sketch of the system with the main object at \bar{x}_m, \bar{y}_m , with total intensity I_m and the companion located $(\delta x, \delta y)$ away with a total intensity of $I_c = I_m f$.

(de Bruijne et al. 2015). In the following, we investigate, both analytically and with image simulations, the effect of unidentified binary stars on the PSF reconstruction for a *Euclid*-like telescope.

3. Expected deviations and reconstruction error

3.1. General case

We investigate in an analytical manner how the PSF is altered by the presence of multiple stellar signals. We show that (i) the bias on the complex ellipticities is a function of both the characteristics of the multiple stars distribution and of the intrinsic complex ellipticities, and that (ii) similarly, the bias on the size is a function of the multiple stars distribution characteristics and of the intrinsic size.

Let q_{xx} be the quadratic moments along the x axis of the image of a star, q_{yy} along the y axis and q_{xy} along the x - y diagonal,

$$q_{jk} = \frac{\int (j - \bar{j})(k - \bar{k}) I(x, y) dx dy}{\int I(x, y) dx dy}, \quad (1)$$

where j, k are either x or y . The parameter \bar{a} denotes the position of the centroid for the axis a . The image is centred on the main star, which is at (\bar{x}_0, \bar{y}_0) . In a noise-free image, the complex ellipticity components are obtained via (Kaiser et al. 1995):

$$[e_1, e_2] = \left[\frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}}, \frac{q_{xy} + q_{yx}}{q_{xx} + q_{yy}} \right] = \left[\frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}}, \frac{2q_{xy}}{q_{xx} + q_{yy}} \right]. \quad (2)$$

We compute the expectation values for the quadratic moments to mimic the result of the stacking of many binary systems. We restrict this analysis to the case of the same intensity for all main stars, while each companion may have a different intensity ratio f_k . All companion stars are randomly distributed around the main star. To get the expectation value of the quadratic moments, we use the limit case when the number of systems in the stack $N_* \rightarrow \infty$. We assume that the light profile of the companion star is given by $I_c(x, y) = f I_m(x - \delta x, y - \delta y)$ and total intensity of $I_c = f I_m$.

For a binary system, the centroid in the x axis is located at

$$\bar{x} = \frac{I_m \bar{x}_m + I_c \bar{x}_c}{I_m + I_c} = \bar{x}_m + \frac{f \delta x}{1 + f}. \quad (3)$$

The computations are analogous for the y axis. We compute the total quadratic moment of the system k , $q_{xx,k}$, in the xx direction, via the parallel axis theorem. The quadratic moment of the binary system k along xx is denoted $q_{xx,m}$, i.e.

$$\begin{aligned} q_{xx,k} &= \frac{q_{xx,m} + (\bar{x}_m - \bar{x})^2 + f_k q_{xx,m} + f_k (\bar{x}_c - \bar{x})^2}{1 + f_k} \\ &= q_{xx,m} + \frac{f_k^2 \delta x^2}{(1 + f_k)^3} + \frac{f_k \delta x^2}{(1 + f_k)^3} \\ &= q_{xx,m} + \frac{f_k \delta x^2}{(1 + f_k)^2}. \end{aligned} \quad (4)$$

The derivation for $q_{yy,k}$ is similar.

3.1.1. Biases for complex ellipticities

The position of the companion stars are random, which implies that $\langle \delta x^2 \rangle = \langle \delta y^2 \rangle$, where the notation $\langle \cdot \rangle$ denotes the expectation value. The expectation value for the complex ellipticity component e_1 is

$$\langle e_1 \rangle = \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy} + \langle f r^2 / (1 + f)^2 \rangle}. \quad (5)$$

As the term $\langle f r^2 / (1 + f)^2 \rangle$ is close to zero in Eq. (5), its Taylor expansion is taken and we get the value of the bias for the system k ,

$$\delta e_{1,k} = e_{1,k} - e_{1,0}, \quad (6)$$

for $i = 1, 2$ and $e_{i,0}$ the i th component of the complex ellipticity, such that the expectation value for the bias is

$$\langle \delta e_1 \rangle \simeq -e_{1,0} \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1 + f)^2} \right\rangle, \quad (7)$$

where R_0^2 is the intrinsic size of the PSF defined as the sum of the quadratic moments in the xx direction plus the moments in the yy direction. This quantity is referred to as T in Kaiser et al. (1995). Similarly, for e_2 ,

$$\langle \delta e_2 \rangle \simeq -e_{2,0} \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1 + f)^2} \right\rangle. \quad (8)$$

The above set of equations shows that the bias, δe_i , is a function of the intrinsic ellipticity of the PSF, $e_{i,0}$, and of the companion distribution through the quantity $\langle r^2 f / (1 + f)^2 \rangle$. It also shows that the magnitude of the bias is the same both in e_1 or e_2 components. The intrinsic size of the PSF R_0^2 can mitigate the effect of the binaries. If we use the same binary population for small and large PSFs (i.e. ensuring that the same binaries are excluded in both cases), the bias is smaller than for small PSFs. Another conclusion drawn from the presence of a minus sign is that the presence of binaries in the stack tends to make the reconstructed PSF rounder than the intrinsic PSF shape.

3.1.2. Bias for size

The size of the PSF is $R^2 = q_{xx} + q_{yy}$. The bias on the size is written $\delta R^2 / R_0^2 = (R^2 - R_0^2) / R_0^2$. Computing the bias gives

$$\frac{\delta R^2}{R_0^2} \simeq \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1 + f)^2} \right\rangle. \quad (9)$$

Thus changes in size depend on the binary star fraction distribution and not on the intrinsic PSF ellipticity.

5.1. The effect of unresolved binaries on PSF determination

A&A 586, A74 (2016)

Table 1. Multiplicity properties for population I main-sequence stars used for this study (adapted from [DK13](#)).

Mass range	Mult./comp. frequency	Mass ratio distribution	Orbital period distribution
$M_\star \lesssim 0.1 M_\odot$	$MF = 22^{+6}_{-4}\%$ $CF = 22^{+6}_{-4}\%$	$\gamma = 4.2 \pm 1.0$	Unimodal (log-normal?) $\bar{a} \approx 4.5 \text{ AU}$, $\sigma_{\log P} \approx 0.5$
$0.1 M_\odot \lesssim M_\star \lesssim 0.5 M_\odot$	$MF = 26 \pm 3\%$ $CF = 33 \pm 5\%$	$\gamma = 0.4 \pm 0.2$	Unimodal (log-normal?) $\bar{a} \approx 5.3 \text{ AU}$, $\sigma_{\log P} \approx 1.3$
$0.7 M_\odot \lesssim M_\star \lesssim 1.3 M_\odot$	$MF = 44 \pm 2\%$ $CF = 62 \pm 3\%$	$\gamma = 0.3 \pm 0.1$	Unimodal (log-normal) $\bar{a} \approx 45 \text{ AU}$, $\sigma_{\log P} \approx 2.3$
$1.5 M_\odot \lesssim M_\star \lesssim 5 M_\odot$	$MF \geq 50\%$ $CF = 100 \pm 10\%$	$\gamma = -0.5 \pm 0.2$	Bimodal $\bar{P} \approx 10 \text{ d}$ and $\bar{a} \approx 350 \text{ AU}$

Notes. Multiple systems are characterised by the frequency of multiple system MF and the companion frequency CF , i.e. the average number of companions per star. The latter can exceed 1. The two last columns present the parameters for mass ratio distribution and for orbital period distribution. See [DK13](#) for more details.

3.2. Ideal reconstruction error

The reconstruction error can be bounded by a requirement for future weak lensing surveys. We write the reconstruction error as

$$\sigma(e_i) = \sqrt{\frac{1}{N_s - 1} \sum_{s=1}^{N_s} (e_{i,s} - e_{i,0})^2}, \quad (10)$$

where $s = 1 \dots N_s$ runs over the N_s stacked images of stars and $e_{i,s}$, $e_{i,0}$ are the ellipticity components of the resulting stacked image and of the ideally reconstructed PSF, respectively. If N_s is large enough, we can write

$$\sigma(e_i) = \langle e_i \rangle - e_{i,0} \approx \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle e_{i,0}. \quad (11)$$

Similarly, for the size

$$\sigma(\delta R^2)/R_0^2 \approx \frac{1}{R_0^2} \left\langle \frac{r^2 f}{(1+f)^2} \right\rangle. \quad (12)$$

3.3. Suboptimal reconstruction error

The above equations are derived for an image that contains a very large number of co-added stellar systems N_\star . If this number gets too small, shot-noise error dominates, introducing an extra source dispersion in the reconstruction error σ . When N_\star is too low, the expectation value $\langle r^2 f / (1+f)^2 \rangle$ will no longer be representative of the stellar binary distribution in the image. We discuss the minimal value of N_\star to escape this suboptimal regime in Sect. 5.3.

The separation r and ratio of the intensity f distributions are typically log-normal as described in [DK13](#). The bias δe is the multiplication of those two log-normal distributions, thus the bias is also an exponential-like distribution. The standard deviation of such exponential-like distributions has a $\sigma_N \sim 1/\sqrt{N_\star}$ dependence (e.g. [Hoogenboom et al. 2006](#)). At low N_\star , this σ_N is more important than the reconstruction error $\sigma(x)$, where x is either the ellipticity component e_i or the size R^2 . The parameter $\sigma(x)$ is the binary PSF reconstruction error signal, while σ_N is a statistical artefact. The total measured error σ_{meas} is a combination of σ_N and $\sigma(x)$. For low N_\star or low binary PSF reconstruction errors, the statistical error σ_N dominates the reconstruction error. On the other hand, when enough binaries are stacked or when the PSF properties exhibit a large value of the size or the ellipticity, the binary PSF reconstruction error $\sigma(x)$ dominates.

4. Binary stars in the Milky Way

We now test our predictions for the PSF reconstruction errors. To accomplish this, we draw a realistic population of stars from the Besançon Model of the Galaxy (BMG; [Robin et al. 2003](#)). To this population of single stars, we add companions according to the properties of binary stars described in [Duchêne & Kraus \(2013\)](#), hereafter DK13), including the orbital and mass ratio parameters. We simulate the PSF reconstruction process by stacking many observations of a mix of stellar systems with and without companions. We then measure the ellipticity and size of the stacked light profile and compare them with the intrinsic values in the absence of any stellar companion.

In the following section, we present the properties of the multiple systems in the Milky Way and discuss the demographics of binary stars with respect to their position on the sky before we describe the simulations themselves.

4.1. Characteristics of multiple stellar systems

[Duchêne & Kraus \(2013\)](#) found that multiple systems are characterised by two main quantities: (i) the frequency of multiple system, MF , i.e. the number of occurrences of a binary system in a given population; and (ii) the companion frequency, CF , the average number of companions per star, which can exceed one. We restrict ourself to binary stars and only use the frequency of multiple systems, MF , to draw populations of binary stars. The physical separation distributions follow a log-normal curve. The detailed representation of the log-normal distribution function suggests different formation processes for multiple systems (e.g. [Heacox 1996](#)). The mass ratio of the companion mass to the primary mass, $q \leq 1$, can be determined from the observation of the flux ratio assuming mass-luminosity relations. The distributions of the mass ratio follow an empirical power law. Table 1 summarises important trends in the population of binary systems. Some of the main trends are:

1. The more massive the primary star, the larger the probability of a companion,
2. The more massive the primary star, the lower the mass ratio power-law index γ . This implies that massive stars tend to host companions of lower relative masses than low-mass stars,
3. The more massive the primary star, the broader the physical separation with the companion.

T. Kuntzer et al.: Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys

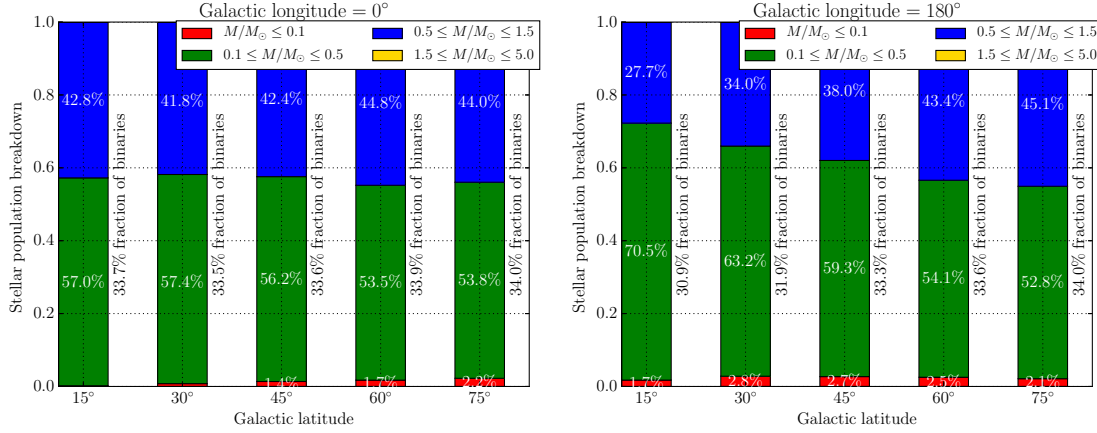


Fig. 3. Main-sequence stellar population breakdown for different Galactic latitudes. The Galactic longitude on the *left panel* is $\ell = 0^\circ$ (towards the centre of the disk) and the *right panel* $\ell = 180^\circ$ (Galactic anti-centre). The bin of massive stars ($>1.5 M_{\odot}$) is not visible as it represents less than 0.2%. We stress the different behaviour of the binary fraction for massive stars (blue vs. green bins) for the opposite directions in Galactic longitude.

4.2. Spatial distribution of binaries

The demographics of stellar populations depend on the Galactic coordinates and most notably on the Galactic latitude, and since different regions of the Milky Way have different stellar populations, they have different fractions of multiple systems. This fraction is mainly controlled by the fraction of massive stars. An informed choice of the stellar population used for the PSF reconstruction can improve the purity of the sample by minimising the fraction of multiple systems.

We illustrate this effect in the following with several realisations of the BMG main-sequence stars at different Galactic coordinates. For each simulation, the number of binary stars is evaluated. Figure 3 shows two different examples for Galactic longitude $\ell = 0^\circ$, near the Galactic centre and $\ell = 180^\circ$, in the direction of the Galactic anti-centre.

Towards the Galactic centre, the stellar population demographics remain approximately constant for latitudes between $b = 15^\circ$ and $b = 75^\circ$. Thus, there is no significant variation in the fraction of binary stars. On the other hand, pointing close to the anti-centre results in an increasing fraction of massive stars with increasing Galactic latitude. Stars of masses between $0.5 M_{\odot}$ and $1.5 M_{\odot}$ represent less than 30% of the stellar population at $b = 15^\circ$, while they account for almost half of the population at $b = 75^\circ$. When this observation is combined with the binary star information of Table 1, this leads to a statistically significant increase of the fraction of binary stars with increasing Galactic latitude, as shown in Fig. 3. This means that the impact of the multiplicity of stars on the PSF determination is not the same at different locations on the sky, not only because of the projected number density of stars but also because of the varying stellar populations. It also means that PSF calibration fields must be carefully chosen and the fraction of multiple stars in the selected fields must be accounted for. In the following, we choose the pointing $\ell = 180^\circ$, $b = 45^\circ$, as the population demographics and fraction of binary stars seems to be representative of all available data and because this pointing is located at a portion of the sky surveyed by *Euclid* (Laureijs et al. 2011).

4.3. Synthetic images of single and binary stars

Our goal is to measure the PSF reconstruction errors for a given number of stars that can be co-added but that are sometimes binaries. To this end, several simulations with different fractions of binary stars are generated. The stars potentially useful for the PSF reconstruction of future space-based surveys have magnitudes in the range $i(\text{AB}) \sim 16\text{--}24$, depending on the exposure duration and detector saturation levels. Thus, we only keep stars falling into this range in the sample. Main-sequence stars more massive than 1.5 solar masses are discarded as well owing to the lack of data about their multiplicity. However, these stars represent less than 0.2% of the whole dataset. We also discard stars that have more than one companion. All our simulations are therefore on the optimistic side in this respect.

The stellar populations generated by the BMG do not contain stellar companion information. We therefore add companions to the BMG stars using the information in DK13 and summarised in Table 1. We replace missing values with the value in the lowest bin of mass, which represents a conservative estimate.

A mass-luminosity relation for low-mass main-sequence stars is used to turn the mass of the companion into a magnitude (Habets & Heintze 1981). For simplicity, we assume that all the stars have the same flat spectral energy distribution. In other words, our simulated images are for flat SEDs but the allocation of companions to the stars follows the prescription of DK13 applied to the stellar populations of the BMG.

The position of a companion star along the assumed circular orbit is drawn from a uniform distribution. The angle between the plane of the orbit and the line of sight, as well as the longitude of the ascending nodes, are also uniformly distributed. The light profile of the companion is normalised according to the mass-luminosity relation, and interpolated and shifted by the required vector. Magnitudes of both objects are modified such that the total apparent magnitude of the system is the original apparent magnitude given in the BMG catalogue. We account for the large uncertainties in the stellar multiplicity parameters and explore the behaviour of the PSF reconstruction at different binary star fraction, by randomly modifying the total binary star fraction of

the stellar population between each realisation of the catalogue to yield a fraction of binary stars between 0 and 1.

When building images of stars and their companions, we adopt a spatial sampling 12 times better than the expected survey resolution. We simulate *Euclid*-like images that have the same resolution as the VIS camera, using Gaussian PSFs with a FWHM of 0.137". In doing so, we adopt a pixel size of 0.0083" and a filter bandpass that matches the VIS instrument of *Euclid*. Down-sampling of the images to the same pixel size as the actual data is not necessary and has little impact on the results in the absence of noise. The results we present in the rest of the paper do not depend on the details of the shape of the PSF, i.e. Gaussians give the same results as more realistic diffraction-limited PSFs generated with the GalSim software (Rowe et al. 2015). This is mainly because the effect of unresolved stellar companions affects the very centre of the PSF.

When dealing with real (noisy) data, the PSF is best recovered by simultaneously considering many different point sources to improve the signal to noise. To mimic this process, we stack many images of stars (some containing binaries) generated as described above. In the present case, all images are (i) noise-free and (ii) centred on the measured centroid of the binary system or the centroid of the single star. Stars are only stacked within a narrow magnitude range. Binaries separated by 0.05" are removed from the sample, i.e. we assume that in a real survey these binaries can be identified as such and removed from the stars used to build the PSF. Each image of a stack is composed of a mix of single and binary stars with realistic separations and contrasts. Measurements of the ellipticity and the size of the stacks are computed with adaptive moments by Hirata & Seljak (2003), as implemented in GalSim (Rowe et al. 2015).

A large number of stacks are measured and divided into bins of similar binary fractions. The biases and random errors on the reconstructed size and complex ellipticities are then estimated by comparing with the input values in the absence of binaries. For each group of stars with a similar binary fraction we measure the standard deviation for the sizes and ellipticities of all objects in the group. This gives an estimate of the precision achieved on a given parameter. The bias is the standard deviation of the difference between the ellipticity of the stack and the input values in the absence of binaries. In the rest of the paper we use the following terminology:

- A stack of stars is an ensemble of stars that all have the same PSF shape and are drawn from the BMG using the binary properties of DK13.
- N_* is the number of stars in each stack; a binary system is counted only once.
- N_s is the number of realisations of the stacks generated for each simulation. Each stack contains N_* stars drawn N_s times. N_s is always large enough to make the numerical errors on the simulation negligible in front the parameters that are reconstructed, i.e. the size and ellipticity of the PSF. Typical values for N_s are hundreds of thousands of realisations.
- PSF reconstruction refers to the measurement of the PSF profile of the stack (size and ellipticity).
- The error on the reconstructed parameters are defined in Eqs. (10) and (12).
- The bias on the reconstructed parameters are defined in Eqs. (7)–(9).

Table 2. Shape of the three PSFs used in the simulations.

ID	e_1	e_2	e	FWHM (")	PA (°)
PSF0	+0.000	+0.000	0.000	0.137	+0.0
PSF1	+0.054	−0.003	0.054	0.137	+88.4
PSF2	−0.021	+0.045	0.050	0.137	+145.5

Notes. For each PSF, we give the ellipticity components and module, FWHM, and position angle.

5. Results

In the following, we analyse the simulations described in Sect. 4.3. We focus on how the error in the PSF reconstruction due to multiple star blending varies with the intrinsic PSF ellipticity, i.e. in the absence of blends, and we show our results as a function of the total fraction of binary stars. We carry out the experiment both for Gaussian PSFs and for diffraction limited PSFs. We then explore the impact of the number of stars, N_* , available to reconstruct the PSF. Finally, we show the effect of changing the stellar populations, i.e. the magnitude distribution of the stars contained in each stack.

5.1. Varying the fraction of binaries for different PSF

We design a simple numerical experiment using three different Gaussian PSFs with the properties shown in Table 2. For each PSF, we generate ensembles of stars with different fractions of binaries and we produce noise-free stamp images on which we measure the ellipticity and size using a simple moment calculation.

Our results are summarised in Fig. 4, where we choose $N_* = 40$ as an example. In practice, N_* can change with different survey characteristics, in particular, the temporal and spatial stability of the PSF. In other words, N_* reflects the number of stars that can be used simultaneously to reconstruct the PSF.

Figure 4 indicates the reconstruction error for the size, $\sigma(R^2)/R_0^2$ and for the two ellipticity components, $\sigma(e_1)$ and $\sigma(e_2)$, as a function of the fraction of binary stars. We also show the biases on the size and on the ellipticity. In each panel, the figure compares the errors and biases, as derived from our simulations (symbols), to the predictions (solid lines) derived in Sect. 3.

The two upper panels of Fig. 4 show that the error and bias on the size as a function of the binary star fraction is independent of the PSF ellipticity, as suggested by Eq. (9). As one would expect, increasing the fraction of binary stars in the stacks increases the size of the PSF. The increase with the binary fraction is linear as the characteristics of the binary stars are the same in each bin. The only difference between the bins is the probability of observing a binary system. The figure also shows excellent agreement between the theoretical prediction and the simulation, even for the small number of stars in the stacks, here $N_* = 40$.

The reconstruction errors $\sigma(e_1)$ and $\sigma(e_2)$ on the complex ellipticity, as obtained from the simulations, appear very similar. They seemingly do not depend on the value of the intrinsic ellipticity as Eq. (11) suggests. However, with $N_* = 40$ stars, we fall in the suboptimal reconstruction scheme described in Sect. 3.3. Those cases are shot-noise-limited because of the small number of stars in the stack, i.e. the PSF reconstruction error of the ellipticity is dominated by noise due to the small number of random positions for the companions. We show in Sect. 5.3 that, as is intuitively expected, increasing N_* reconciles the errors found from the simulations and theoretical prediction: $\sigma(e_1)$ and $\sigma(e_2)$

T. Kuntzer et al.: Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys

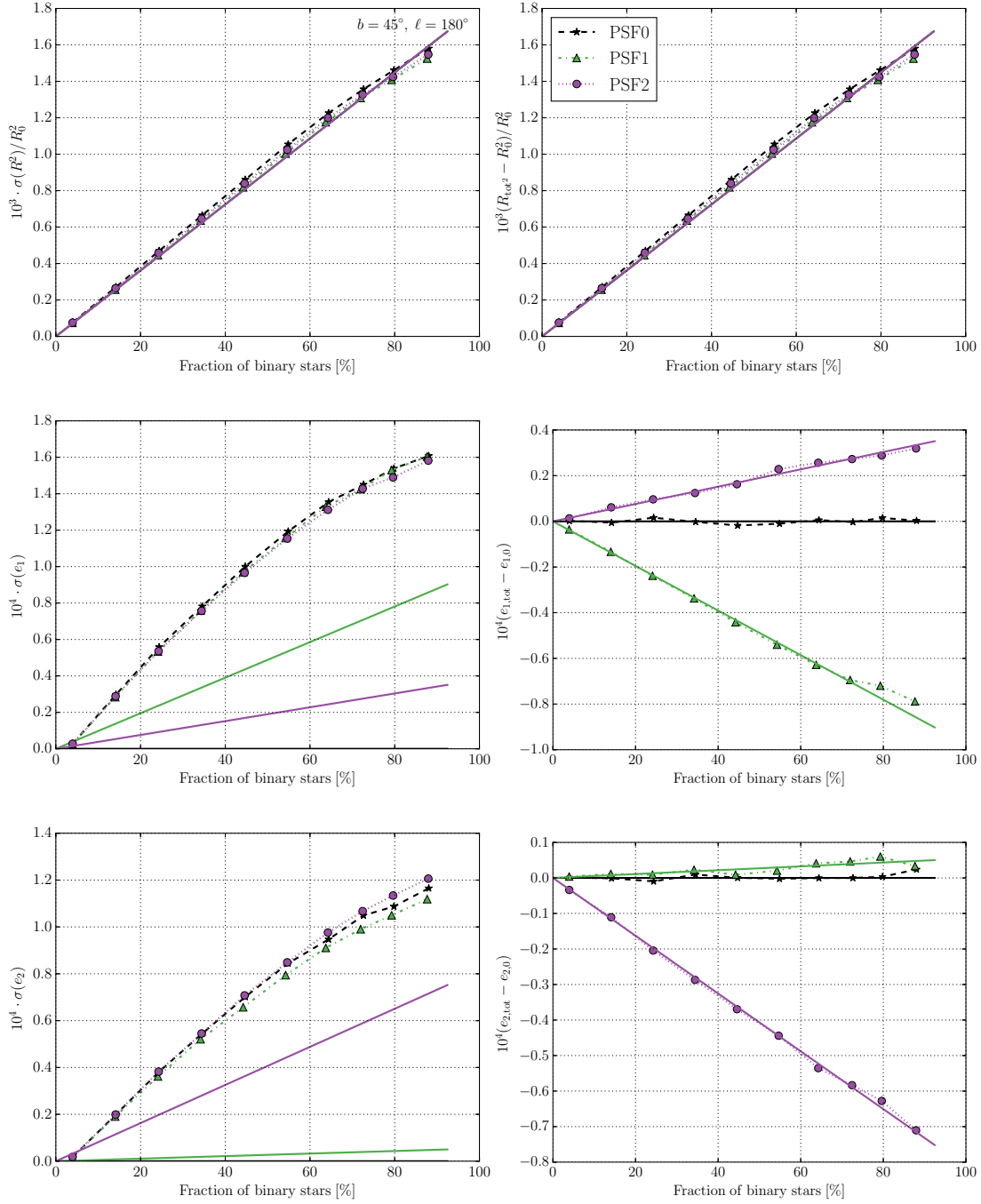


Fig. 4. PSF reconstruction errors (*left column*) and biases (*right column*). From *top to bottom*, plots for the size and two ellipticity components, e_1 and e_2 , are shown. In each panel, the values are given as a function of the binary fraction in the stellar population considered and for the three different PSF shapes described in Table 2. The solid lines represent the predicted values using the formalism in Sect. 3. The latter is valid for $N_* \rightarrow \infty$ while the simulations used in this figure are for $N_* = 40$ stars with apparent magnitude in the range $18 \leq i(\text{AB}) \leq 19$.

do depend on the ellipticity of the PSF following Eq. (11) in the limit of large values for N_* .

We now turn to the biases due to stellar binarity. The bias on the size follows the same curve as a function of stellar binarity as the error on the size. The equations describing the bias and the error, as reported in Sect. 3, are

$$\frac{\delta R^2}{R_0^2} \simeq \frac{\sigma(\delta R^2)}{R_0^2}, \quad (13)$$

which is also well illustrated with our Gaussian simulations in the upper right panel of Fig. 4. The bias, both on the size and ellipticity, is easily measured from simulations with $N_* = 40$ in contrast to the measurement of the errors, which require much larger values for N_* .

The behaviour of the ellipticity is not the same as for the size. For a round PSF with $e_1 = e_2 = 0$, both the bias and reconstruction error equals zero in the limit where $N_* \rightarrow \infty$, as the net effect of adding round PSFs at random angular positions around a star is null. The remaining bias for a round PSF, both on e_1 and e_2 , is of the order of 2×10^{-6} . This can be attributed to the effect of a large but finite number of stacks N_s and to numerical errors.

For non-circular PSFs, the bias on e_1 and e_2 shows a linear correlation to the intrinsic complex ellipticity of the PSF, as does the error. The bias and reconstructions on the ellipticity have the same analytical form with opposite signs, and both are proportional to the intrinsic ellipticity of the PSF, i.e.

$$\langle \delta e_1 \rangle \simeq \frac{e_1}{|e_1|} \sigma(e_1), \quad (14)$$

$$\langle \delta e_2 \rangle \simeq \frac{e_2}{|e_2|} \sigma(e_2). \quad (15)$$

The simulated measurements fall very close to the predictions in all panels of the right column of Fig. 4, i.e. the predictions are validated by simple Gaussian simulations.

The slight departure from linearity, for very large binary star fractions, is due to a changing value of $\langle r^2 f / (1 + f)^2 \rangle$, i.e. for large binary fractions the stellar population is modified with respect to populations with a lower fraction of binaries.

The simple linear relation between the bias and intrinsic ellipticity of the PSF hints at a possible calibration to reduce the effect of binary stars on the PSF shape or, at the very least, on a reliable control of the bias. However, that the lack of knowledge of the distribution of the binary star fraction ($\langle r^2 f / (1 + f)^2 \rangle$) in each specific field used to reconstruct the PSF may complicate the task in practice.

5.2. Synthetic space-based PSF

Up to now, we have only considered Gaussian PSFs. In order to test the effect of a more complex PSF, we use the GalSim software to produce a synthetic PSF similar to future space-based telescopes. More specifically, our PSF is the same size as the *Euclid* VIS PSF and has six struts. No PSF aberration such as coma or trefoil is applied. The strut angle, with an arbitrarily chosen orientation, is kept constant.

Similar to the simulations in Sect. 5.1, we build three PSFs with the characteristics shown in Table 2. We choose the strut-rotated validation case to have the same e_1 and e_2 input as the PSF along e_1 .

The results show the same behaviour as for the Gaussian toy model. No significant departure from the predictions are detected. This is likely because we consider here very narrow angular separation binaries, with a companion star well within the

central pixel of the PSF. It is therefore not surprising that the main change in shape, due to the binarity, occurs in the very centre of the PSF. The details of the PSF shape on larger scales do not matter much. This suggests that (i) simple Gaussian models are sufficient to assess the impact of binarity of the PSF shape; and (ii) our analytical predictions based on Gaussian PSFs are a good approximation of the change in the PSF shape. These predictions can therefore be safely used to predict the PSF reconstruction error as a function of the binary fraction and mean angular separation between the stars and their companions.

5.3. Effect of the number of stars N_* in the stack

As seen in Sect. 5.1 and illustrated in Fig. 4, varying the number of stars in the stack N_* used to reconstruct the PSF changes the reconstruction error. Increasing the number of stars increases the sampling of the coordinates for the companion stars. Eventually, for very large values of N_* , the predicted errors as computed in Sect. 3 are realised by the simulations.

For $N_* = 40$ stars, the reconstruction error on the size, as well as the values for the biases on the size and ellipticity, are compatible with the predictions, as in Fig. 4. However, the errors on the ellipticity components require larger values for N_* , which also depends on the intrinsic ellipticity of the PSF.

This is tested in Fig. 5 for $N_* = 40, 750, 2100, 10\,000$ and for two PSF ellipticities. For low values of N_* , the error is large and independent of the ellipticity of the PSF and the error is dominated by the shot noise due to the number of companions. As can be expected, it takes a much larger N_* to converge to the theoretical predictions, and the smaller the ellipticity, the larger N_* needed. For example, $N_* = 10\,000$ stars in the stack are barely sufficient to converge to the theoretical estimate of the error for PSF2, which has a small e_1 component of -0.021 .

Figure 5 shows that when N_* is very large, the theoretical values for the error estimates on the ellipticity can be modelled by simulating binary populations and the equations in Sect. 3. However, the figure also shows that if for any reason the prediction we make in Sect. 3 with Gaussian PSF does not hold true, then the simulations required to replace the theoretical estimates involve tens of thousands of companions. Given that we also need large numbers of realisations N_s , estimating the effect of binary stars on the PSF reconstruction may quickly become computationally expensive.

5.4. Dependence on the apparent magnitude of the stars

Our ability to resolve a binary star system depends mostly on its physical distance to the Earth, i.e. on its apparent magnitude. The PSF reconstruction errors due to the blend of binaries are correlated with the distance to the binary system through the $\langle r^2 f / (1 + f)^2 \rangle$ parameter. We now turn to the impact of stars in different magnitude bins on the reconstruction of the PSF. In doing so, we adopt a binary star fraction of 35%, which is representative of the mean binary star fraction on the sky (DK13). The resulting reconstruction errors as a function of the $i(\text{AB})$ magnitude are shown on Fig. 6. In the figure, we also indicate, as a reference, the total reconstruction error budget for a stage IV weak lensing experiment: $\sigma(e_i) \leq 2 \times 10^{-4}$ for the ellipticity components and $\sigma(R^2)/\langle R^2 \rangle \leq 1 \times 10^{-3}$ for the size (Paulin-Henriksson et al. 2008). These values are the total error budgets allowed for these parameters.

For the brightest stars with $16 < i(\text{AB}) < 18$, the error in reconstruction takes the full error budget both on the size and

T. Kuntzer et al.: Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys

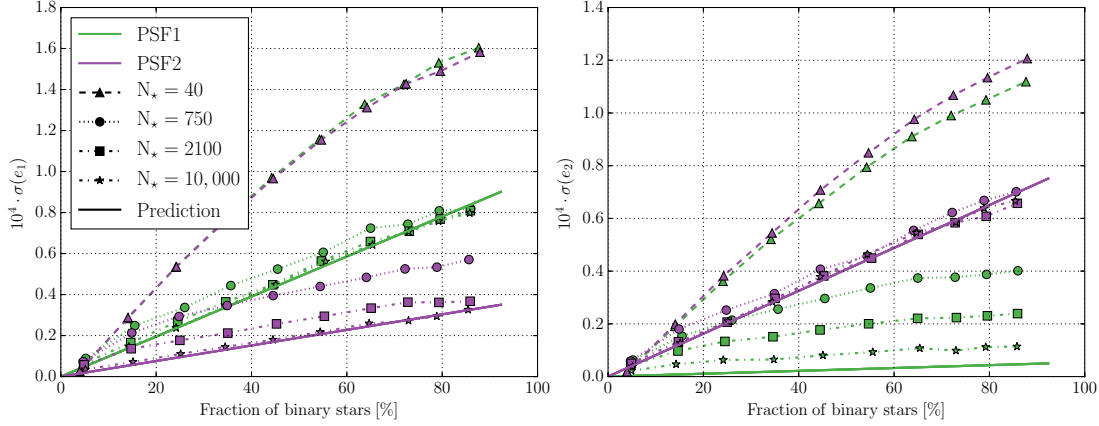


Fig. 5. PSF reconstruction errors on the two ellipticity components, e_1 (left) and e_2 (right). These plots correspond to the two panels in the lower left corner of Fig. 4, but we now show the effect of changing the number of stars, N_* , used in the stack. Depending on the intrinsic ellipticity of the PSF (see Table 2), thousands of stars are needed to avoid being dominated by the shot noise.

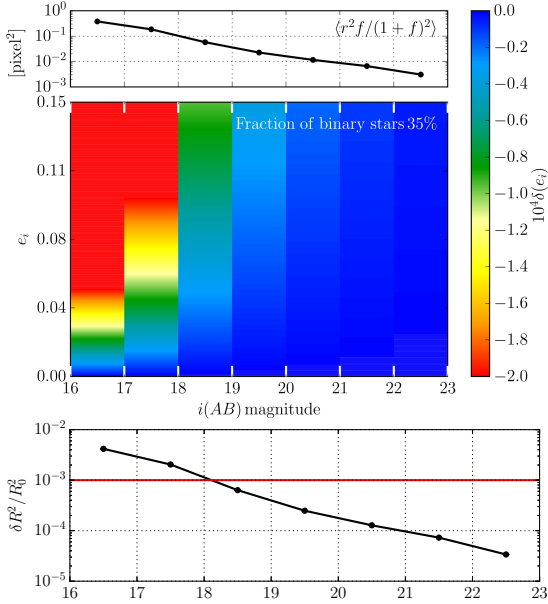


Fig. 6. Theoretical reconstruction error according to the formalism developed in Sect. 3. The stellar distribution is drawn from realisations of the BMG for pointing coordinates to $\ell = 180^\circ$, $b = 45^\circ$. *Top*: value of the $\langle r^2 f / (1 + f)^2 \rangle$ parameter as a function of apparent $i(\text{AB})$ magnitude. *Middle*: reconstruction error as a function of the apparent magnitude $i(\text{AB})$ of the stars in the stack and as a function of the complex ellipticity components e_1 or e_2 (colour code). *Bottom*: reconstruction error on the PSF size. The red line indicates typical requirements for stage IV experiments.

ellipticity, even for small intrinsic PSF ellipticities. Stars in the magnitude range $18 < i(\text{AB}) < 19$ still suffer from a relatively large blending bias, taking up to 60% of the total error budget on the size and ellipticity. Requiring stars fainter than magnitude 20 lowers the blending bias to $\leq 10\%$ of the error budget, which is still large enough to be accounted for.

Stars brighter than roughly $i(\text{AB}) \approx 19$ may not be the best choice for PSF calibration strategies, as brighter stars are much more likely to alter the PSF both in size and ellipticity. Instead, longer observations of faint and most distant stars are preferred and minimise the effect of binary stars. Finally, we emphasise that the impact of binary stars on the PSF shape becomes more important with higher spatial resolution if the stellar population remains the same. Surveys like WFIRST may therefore be more impacted by binary stars than the *Euclid*-like telescope used in this work. On the other hand, because WFIRST has deeper limiting magnitudes, more faint stars can potentially be used to reconstruct the PSF.

6. Conclusions

Multiple stellar systems are ubiquitous in the Milky Way, with a total fraction of multiple stars of about 35% (DK13). These multiple stars, especially when they are not resolved by the instrument, may significantly impact the PSF measurement given the strong requirements on the modelling of the PSF.

In this paper, we present analytical predictions for the PSF reconstruction error introduced by binary stars based on stellar multiplicity parameters and on the intrinsic PSF size and ellipticity. We then verify our predictions by means of realistic numerical simulations. To do so, we use the BMG (Robin et al. 2003) to generate mock stellar distributions. We then add companion stars to produce binary systems following the most recent observational data in DK13. We produce noise-free images used to measure the error and bias on the PSF ellipticity and size.

In our simulations, the parameters of the detector and of the PSF are taken to be *Euclid*-like, but the PSF profile is Gaussian. We also check that our results are valid for more realistic diffraction limited PSFs. Binary stars with a separation larger than $r = 0.05''$ are assumed to be easily identified as multiple and eliminated from the star catalogue to build the PSF. In other words, we only consider double stars that cannot be identified as such but that statistically affect the PSF reconstruction. Our main findings can be summarised as follows.

- Binary stars can significantly alter the reconstruction of the PSF. They introduce both a bias and an error on the size and ellipticity of the reconstructed PSF.

A&A 586, A74 (2016)

- The PSF binary reconstruction errors can be predicted analytically as a function of characteristics of the stellar population considered, i.e. the separations and intensity ratio of the stars, and as a function of the intrinsic PSF size and ellipticity.
- The effect of the binaries cannot be suppressed by averaging a large number of stars.
- The analytical predictions are supported by numerical experiments. The PSF used in the simulations are Gaussian and diffraction-limited.
- Bright stars of magnitudes $i(\text{AB}) \lesssim 19$ are most affected by binarity. These stars are either massive stars with very likely companions or they are nearby stars, with potentially large angular separations to their companion(s).
- Fainter stars in the range $18 < i(\text{AB}) < 19$ may cause significant biases on the PSF shape. If no mitigation scheme is found, these stars can take on their own up to ~60% of the total budgeted error on the size and ellipticity described in Paulin-Henriksson et al. (2008) and Cropper et al. (2013).
- Blending errors in systems fainter $20 < i(\text{AB}) < 21$ can still contribute up to 10% to the typical error budget on the size and ellipticity.
- If any PSF calibration fields are chosen for space missions, targeting fields of faint stars can be preferable to bright stars to avoid nearby binary systems that have larger effects on the PSF shape.
- The total fraction of binaries varies with Galactic longitude. There are hints that the binary star fraction increases with Galactic latitude. Thus, the variations in the underlying stellar populations can have a varying effect on the PSF shape as a function of Galactic coordinates. This may also imply a bias on the shear power spectrum that must be controlled.
- Different stellar types differ in their fraction of multiple systems and also in the distribution of the parameters of the orbits. Regions of the sky with lower binaries fractions could be exploited to quantify the effect of binary stars on the PSF determination differentially across the sky.

Given the above conclusions, it will be important that future space-based weak lensing surveys account for the effect of binary stars on the PSF determination. In particular, their

contribution to the error budget on the PSF shape (both the error and the bias) must be estimated and any PSF reconstruction method must account for the extra error caused by binarity. The above conclusions depend on the ability to clean star samples from multiple systems. This can be done in many ways and with different efficiency depending on signal-to-noise ratio and stellar type. *Gaia* may provide the spectral type of the stars and, in some cases, identify binary stars from their apparent motion on the plane of the sky. Alternatively, specific image processing techniques using the *Euclid* images themselves may be used, but these remain to be devised.

Acknowledgements. We are indebted to the anonymous referee whose valuable comments increased the quality of this work. The authors would like to thank Laurent Eyer, David Harvey, Henk Hoekstra, Matthew Nichols, Pierre North, and Thomas Kitching for useful discussions. This work is supported by the Swiss National Science Foundation (SNSF).

References

- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **431**, 3103
 de Bruijne, J. H. J., Allen, M., Azaz, S., et al. 2015, *A&A*, **576**, A74
 Duchêne, G., & Kraus, A. 2013, *A&A*, **51**, 269
 Duquenois, A., & Mayor, M. 1991, *A&A*, **248**, 485
 Frieman, J. A., Turner, M. S., & Huterer, D. 2008, *ARA&A*, **46**, 385
 Habets, G. M. H. J., & Heintze, J. R. W. 1981, *A&AS*, **46**, 193
 Heacox, W. D. 1996, *PASP*, **108**, 591
 Hirata, C., & Seljak, U. 2003, *MNRAS*, **343**, 459
 Hoogenboom, J. P., den Otter, W. K., & Offerhaus, H. L. 2006, *J. Chem. Phys.*, **125**, 204713
 Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, **449**, 460
 Kilbinger, M. 2015, *Rep. Prog. Phys.*, **78**, 086901
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, Euclid Study Report, ArXiv e-prints [[arXiv:1110.3193](https://arxiv.org/abs/1110.3193)]
 Massey, R., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **429**, 661
 Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, **484**, 67
 Reipurth, B., Clarke, C. J., Boss, A. P., et al. 2014, *Protostars and Planets VI*, **267**
 Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, **409**, 523
 Robin, A. C., Luri, X., Reylé, C., et al. 2012, *A&A*, **543**, A100
 Rowe, B., Jarvis, M., Mandelbaum, R., et al. 2015, *Astron. Comput.*, **10**, 121
 Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints [[arXiv:1503.03757](https://arxiv.org/abs/1503.03757)]
 Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, *Phys. Rep.*, **530**, 87

5.1.5 Strategies for reducing the effect of unresolved binary bias

Binaries are both ubiquitous and common. It is thus not possible to simply refrain from observing them. In this section, we examine three paths that lead to a sample of single stars. The first technique is to look-up each object in the *Gaia* stellar catalogue, the second is to determine the spectral type of the object and deducing a multiplicity probability, and the third is to exploit the bias induced on the shape by unresolved binaries and identifying deformed PSFs. We start by studying the expected binary population in *Euclid* weak-lensing images.

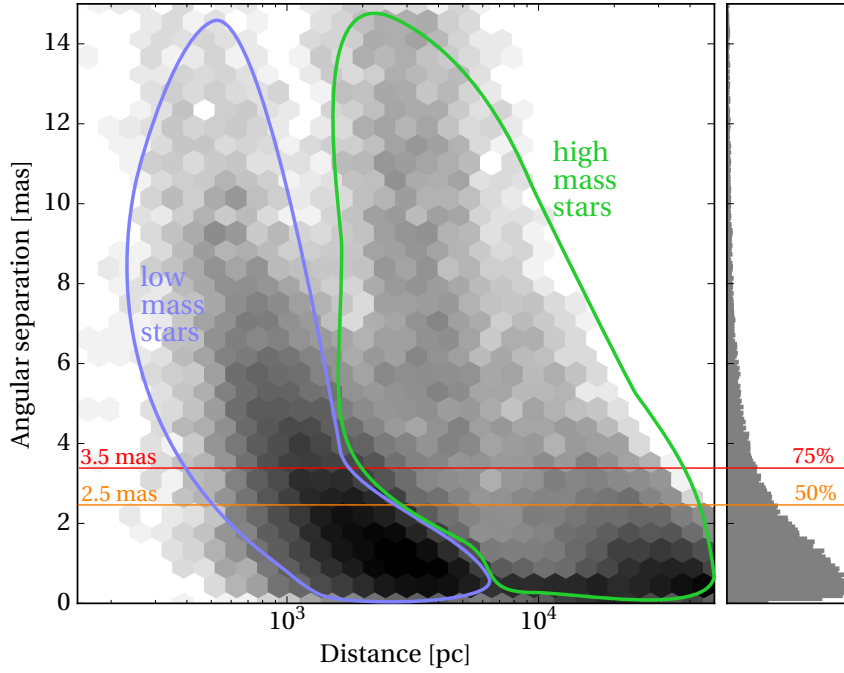


Figure 5.5 – Numerical model of the binary star density as a function of distance and contrast. The grey scale and the histogram are linear. We select stars in the magnitude range $18 \leq i(AB) \leq 24.5$. The plot exhibits a bimodal distribution with low-mass main stars located at shorter distances (blue) and more massive stars (in green). More massive stars are not located in our direct neighbourhood. 50% (orange line) of the binaries are separated by less than 2.5 milli-arcseconds and 75% (red line) by less than 3.5 milli-arcseconds.

Once more, we combine BGM catalogues with the available multiple population knowledge to generate realistic stellar populations with *Euclid* survey constraints. We then draw the companion parameters: mass ratios and positions of the two members relative to the observer. We present the results of these simulations in Figures 5.5 and 5.6, where we show the number density of binaries as a function of the distance in terms of observed angular separation and mass ratio respectively. The results in these figures are averaged over the pointing range foreseen for the *Euclid* weak-lensing survey (with galactic longitudinal bounds of $+15^\circ < \ell < +75^\circ$) towards the anti-Galactic centre. There is a bimodal distribution of the systems in both

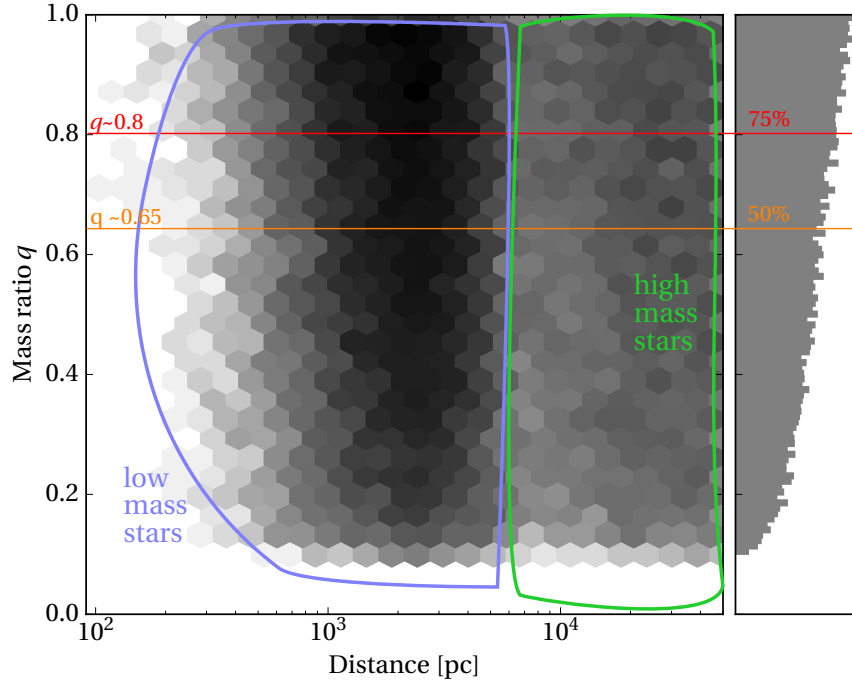


Figure 5.6 – Same as Fig. 5.5, but as a function of mass ratio q . Companions of very low mass stars (down to the brown dwarf limit) are rare with the applied selection cuts. 50% (orange line) of the systems have $q \gtrsim 0.8$ and 75% (red line) have $q \gtrsim 0.65$.

figures. A peak is visible for close-by stars, which are lower mass stars and more common, but intrinsically less luminous. The second peak is for stars that are brighter than the $i(AB) < 18$ cut-off if located close to Earth. In this second region, the low mass stars are too faint to be detected. 75% of the binaries are angularly separated by less than 3.5 milli-arcseconds (mas), and 50% are separated by less 2.5 mas. We previously set a detection threshold at half the VIS pixel scale, that is 50 mas. This assumption removes virtually none of the simulated systems. The mean distribution mass ratio (Fig. 5.6) is unsurprisingly high as low q systems are dynamically unstable during formation. Three quarters of the systems have $q \gtrsim 0.8$, which implies that most stars have similar masses and thus a small contrast. There is a lower bound in mass ratio q : the secondary objects must at least be a brown dwarfs to have a non-negligible luminosity. For low mass host stars, this lower bound on the mass ratio is $q \sim 0.1$. Systems with low q will not influence the reconstruction as their contrast will be small. In summary, there will be two populations of binary stars for *Euclid*: (i) low mass and close-by and (ii) massive stars, but at large distances. Most of the systems are tight-knit as 75% of the stars are separated by an angle of less than 3.5 mas. The mass ratio q is relatively high; half of the systems have at least $q > 0.65$.

We turn to quantifying the bias due to the binary population as a function of their contrast

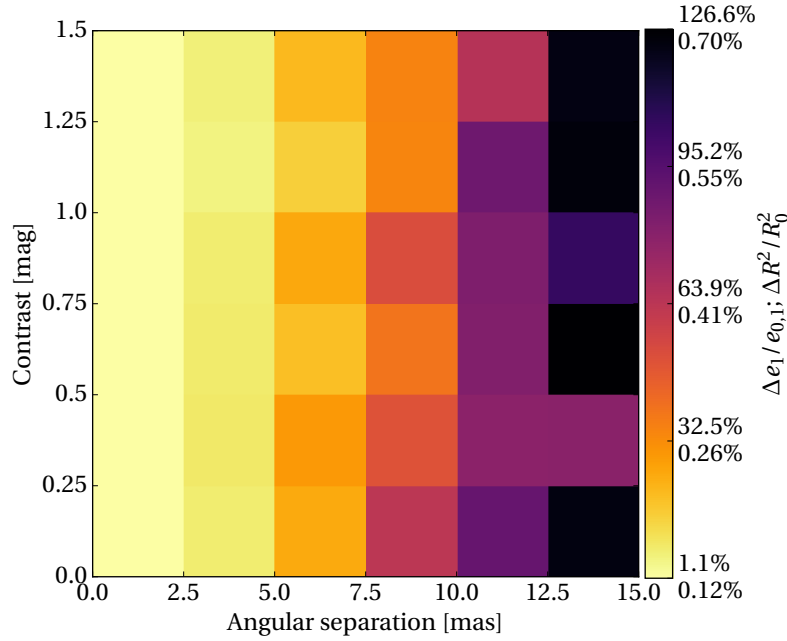


Figure 5.7 – 95-percentile relative errors of the PSF shape parameters e_1 and the size. The relative errors on e_2 are about twice smaller than on e_1 . The values of the relative errors are given in the colour bar for the ellipticity (top number) and for the size (bottom number). The magnitude selection is $18 \leq I(AB) \leq 24.5$.

and angular separation. To that end, we measure the error generated by binaries in a grid of 6×6 bins of contrast and angular separation. We restrict the criteria to a maximum angular separation of 1.5 mas and a contrast of 1.5 mag to avoid poorly sampled bins in the high angular separation high contrast bins. Figure 5.7 illustrates the magnitude of the relative errors due to binaries at small separations and low contrast. We show the 95-percentile value of the expected relative errors in the shape of PSFs. The intuition that larger relative errors occur at larger angular separation and low contrast is confirmed. Note that the first two columns contains $\sim 30\%$ of the global stellar population. This figure is noisy because of the low-number statistics in high angular separation and contrast bins, however there are three discernible trends. (i) The error diminishes as the contrast is increased. (ii) The effect increases with angular separation. (iii) Ellipticity components are affected orders of magnitude more than the size. This suggests a binary detector based on the analysis of the size of the profile has a low probability of success. Ideally, we would like to remove all binaries from the sample of stars, but of course this is an utopia. What we can hope for is to detect binaries with a given minimum separation and a give maximum contrast. With this assumption, we can remove all candidates above a separation threshold and below a contrast limit. The resulting errors on the PSF are presented in Fig. 5.8.

This figure is much less noisy, as there are many objects in every bin. The errors are the largest for a low contrast and high angular separation population. Note that the errors for the second

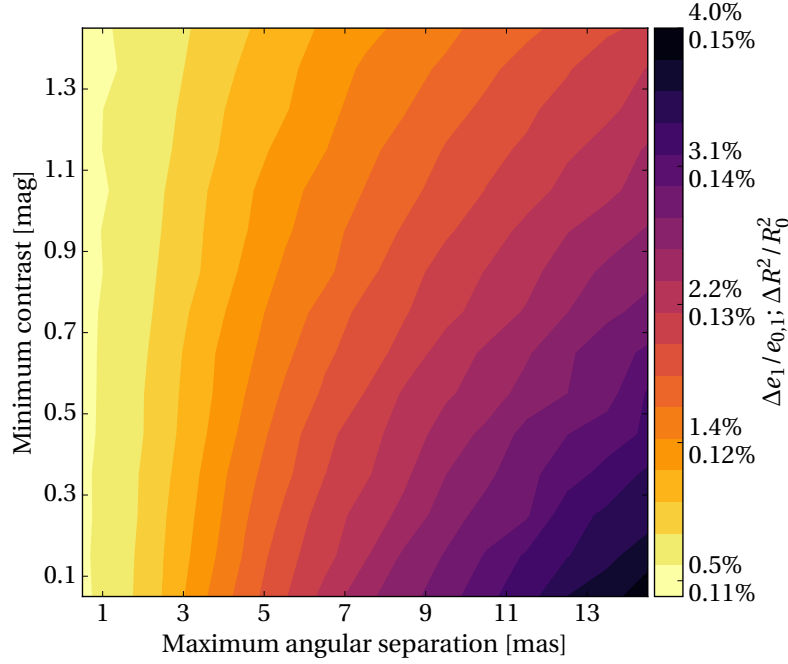


Figure 5.8 – 95-percentile relative errors of the PSF shape parameters e_1 and the size. The relative errors in e_2 are about twice smaller. The values of the relative errors are given in the colour bar for the ellipticity (top number) and for the size (bottom number). The magnitude selection is $18 \leq I(AB) \leq 24.5$. The distributions of relative errors contain all binaries that have at most an angular separation on the x -axis and at minimum a contrast on the y -axis.

ellipticity component are twice as small as the e_1 errors. The reconstruction error for the PSF is constrained by $\sigma(e_i) \leq 2 \times 10^{-4}$ for the ellipticity components and $\sigma(R^2)/\langle R^2 \rangle \leq 1 \times 10^{-3}$ for the size at any point in the image. These requirements are valid when reconstructing the PSF using all data available; these are not requirements on the determination of the profile of each point source in the field. Each VIS image will contain from 2 000 to 3 000 stars (Cropper et al. 2013; Laureijs et al. 2011). If the PSF reconstruction scheme is optimal, the knowledge on each individual star is $\sqrt{2000} \approx 50$ worse than the requirements. The ellipticities will have to be known to $\sim 1\%$ and the size to $\sim 5\%$ for each individual object. Based on Fig. 5.8, we suggest that a detection limit of ~ 3 mas in separation for contrast smaller than $\Delta m < 1.5$ mag. This ensures that the distortions due to binaries are of the same order of magnitude as the measurement error per star for the ellipticity, that is, $\sim 1\%$. We now explore techniques to reach these detection thresholds, or at least to remove a significant fraction of the binaries.

The first idea is simple: we perform a look-up in a binary catalogue and remove all objects from a PSF reconstruction sample known to be in a multiple star system. Multiple stars are poorly known, most of them were found in volume-limited searches in the solar neighbourhood (Sect. 5.1.1). The most complete catalogue when *Euclid* is in operations will be the *Gaia*

catalogue. The *Gaia* mission² (de Bruijne 2012; Gaia Collaboration 2016) is a currently active ESA satellite to map one billion stars in our Galaxy. One of the secondary objective of *Gaia* is to better characterise multiple systems. Looking-up stars observed by VIS in the *Gaia* census will certainly be implemented in the data reduction pipeline to extract, e.g., the SEDs. The *Gaia* catalogue, however, is foreseen to be shallower and cover a smaller footprint than *Euclid* (Eyer et al. 2015). While certainly important and useful to identify several undesirable objects in the PSF sample and providing some of the SEDs, *Gaia* has a limited depth relative to *Euclid* (the *Gaia* binary detection threshold is especially shallow) and smaller extend. This technique can only remove a small fraction of the objects.

The second idea exploits the link between spectral type and multiplicity. Multiplicity fraction, MF , and companion frequency CF are functions of the mass of the star. The mass of the star can be deduced from its spectral type using a mass-luminosity relationship. The spectral type of each star will have to be measured in order to mitigate the colour bias (Sect. 3.2.2). This can be done in several ways: using the *Gaia* survey, SED fitting using multiple bands or inferring directly the spectral type from the VIS image (Sect. 5.2). Identifying individual multiple systems based solely on spectral type is not possible. However, type of stars with high multiplicity fractions can be excluded altogether. Duchêne and Kraus (2013) showed that $\gtrsim 60 - 75\%$ of type A to O main sequence stars are at least in a binary system. Those stars are relatively rare in the *Euclid* stellar population, they tend to have a larger semi-major axis, which, potentially bias the reconstruction more. This technique can only remove a small number of binary objects. It is not sufficient to reach the performance we set in the previous paragraph, even if this spectral approach identifies some of the most harmful objects.

At this stage, the PSF reconstruction objects sample was filtered using stellar parameters only (namely the spectral type and, when available, *Gaia* information on multiplicity). This third idea changes the type of data. In order to simplify the development of the technique, we assume temporal PSF stability. We will drop this assumption later on. The observation plan for the *Euclid* weak-lensing survey is to observe the same region four times. Here again we simplify the picture by assuming very small dither between the frames (also to be abandoned later). Each of the objects will therefore be observed four times. We propose to perform a cross-exposure analysis of the profiles to remove unresolved binaries. PSF spatial variations are smooth in regions observed by the same detector (there can be discontinuities of the PSF field when going from a CCD to another, but we neglect that here). If the PSF field is known (or can be reconstructed) with high enough confidence, we can analyse the bias of each object and identify unresolved binary stars. We know from simple considerations that individual profiles must be known to $\sim 1\%$ for ellipticity and $\sim 5\%$ for the size. Measurements errors for single stars should be random (we will relax this assumption in Paper II, Sect. 5.1.6), thus they should uncorrelated across the exposures. When measuring binary systems however, the errors relative to the fiducial PSF profile are correlated. Binaries systematically and coherently deform the light profile across exposures. The frames are captured in a short timespan (about

²<http://sci.esa.int/gaia/>

5.1. The effect of unresolved binaries on PSF determination

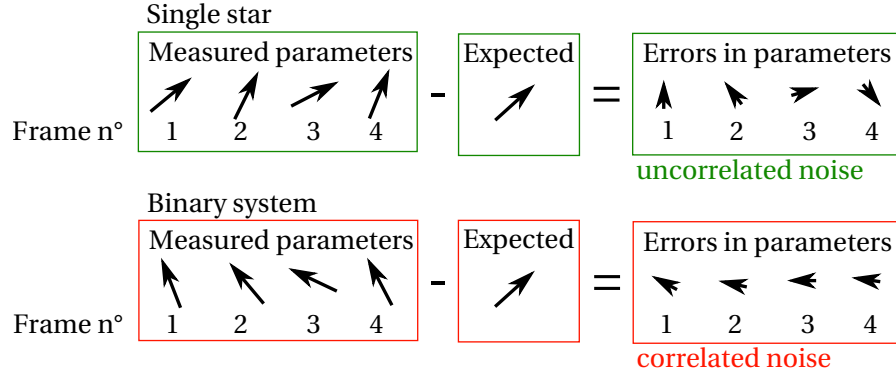


Figure 5.9 – Sketch of the measurement principle. The four observations of an object are compared to the fiducial value of the shape parameters. (*Top.*) Single star shape measurements are noisy, but uncorrelated. (*Bottom.*) Binary stars measurements are noisy too, and correlated as the object profile is systematically biased with respect to the expected PSF shape.

forty-five minutes), meaning that the configuration of the binaries is constant.

We sketch the detection principle in Fig. 5.9, where we illustrate the ellipticity of two objects, a single star and a binary, the expected PSF profile for those spatial coordinates and the errors in the shape measurement. This detection technique relies on precise and accurate shape measurement techniques. Currently, there are no method capable of reaching the *Euclid* VIS targets. We will demonstrate the method at the catalogue level. The catalogue contains at least (i) the image coordinates of the star (two values)³, (ii) the measurements of the size for four exposures (four values), and (iii) the measurements of the shape for four exposures (e_1, e_2 , – eight values). When the PSF field is known, computing the measurement errors is straightforward. The PSF field is not known *a priori*. At this stage of the weak-lensing data reduction we are precisely trying to reconstruct it. Any binary detection method must thus be either robust to a poorly known PSF field or reconstruct it. We propose a scheme that determines the PSF shape parameters field and identifies binaries iteratively (Fig. 5.10).

The iterative scheme proposed in Fig. 5.10 requires an input catalogue, a detector of correlated errors and a stopping criterion. This stopping criterion can be a limit on the number of iterations, or a constant classification. The quality of reconstructed PSF field should increase with the iterations. Every object is classified as meeting the requirements or to be excluded at every iteration to avoid misclassification. This scheme could potentially suffer from misclassifications for high binary fractions or for clusters of binaries at the edges of the frame. The mean density of stars is important for this technique, the higher the better. The VIS instrument will be comprised of 6×6 CCDs. Since 2 000-3 000 objects are foreseen by frame, we can expect about seventy object per detector.

³If we drop the small dither and the PSF temporal stability assumptions, this number grows to twelve (four times two coordinates and one time stamp).

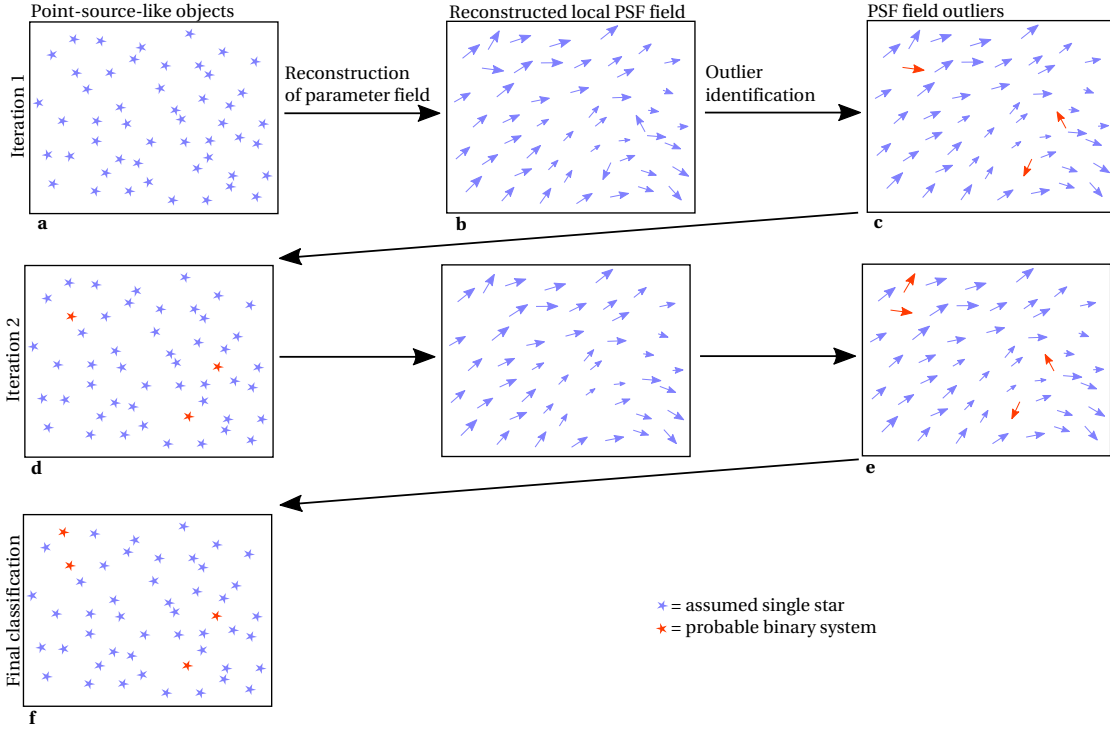


Figure 5.10 – Sketch of the combined PSF field reconstruction and binary identification method. (a.) Point sources in the images are identified. They are assumed to be all single stars for the first iteration. (b.) A shape parameter field is reconstructed from the shapes measured on the point sources designated as single. In the first iteration, that means all objects. (c.) The reconstructed PSF field is smoothed, and its predictions considered as the fiducial PSF parameters. The measurements errors are computed. Objects exhibiting correlated errors are marked as outliers. (d.) Another iteration of the same process begins. Outliers, i.e. presumed binary systems, are not considered in the PSF field reconstruction. (e.) Every object is evaluated again and outliers are flagged. (f.) The algorithm is stopped according to a stopping criterion. The output object catalogue contains an additional multiple star flag.

It remains to find a good correlator for the reconstruction errors. We explore three techniques. In the following, we denote the relative errors for the star i by $\delta_p^{(i)}$, where p is one of the three shape parameter (e_1 , e_2 or R^2).

$$\delta_p^{(i)} = \frac{p^{(i)} - p_0^{(i)}}{p_0^{(i)}} \quad (5.11)$$

where 0 represents the fiducial value of the parameter p . $\delta_p^{(i)}$ has a vectorial form as there are four observations of each of the shape parameters. We denote the frame number by $t = 1, 2, 3, 4$. The vector of the errors can be interpreted as a time-series of the observations.

The first algorithm is a simple auto-correlation of those error vectors $\delta_p^{(i)}$. We compute the

5.1. The effect of unresolved binaries on PSF determination

auto-correlation function (ACF) for star i and parameter p by,

$$\text{ACF}_p(\tau) = \frac{1}{(n-1)\sigma^2} \sum_{t=1}^{n-4} [\delta(t) - \bar{\delta}][\delta(t+\tau) - \bar{\delta}], \quad (5.12)$$

where σ^2 is the sample variance, and $\bar{\delta}$ is its mean value as computed from four observations of δ . For simplicity, we dropped the star index i in the above equation. The quantity τ is the lag. It does not however physical bear units as t represents the identification number of the exposure. We sum the four estimated ACF coefficients to create a single scalar quantity $\Sigma_p = \sum_{\tau=0}^3 \text{ACF}_p(\tau)$. This scalar represents the degree of correlation between the error on a given parameter. A discriminative threshold is chosen such that stars with a high Σ_p are flagged as probable binaries. This ACF method does not incorporate any physics and therefore is completely data driven. ACF analysis is applied to idealised examples in Fig. 5.11. In the above, only one of the shape parameters was exploited. We refer to ACF_{e_1} , ACF_{e_2} and ACF_{R^2} as three different channels. Those channels can be linearly combined to have a single prediction for each star. Weights are empirically determined, and we can already estimate the order of importance of the channels by studying Fig. 5.8. Random measurement errors are $\mathcal{O}(1\%)$ and $\mathcal{O}(5\%)$ for, respectively, the ellipticity components and the size. The 95-percentile errors in the first ellipticity component, $\mathcal{O}(1-10\%)$, are about twice the errors of the second ellipticity component. The error on the size is another order of magnitude lower than those. We can thus deduce that the most important channel will be ACF_{e_1} , ACF_{e_2} and a poor classifier ACF_{R^2} .

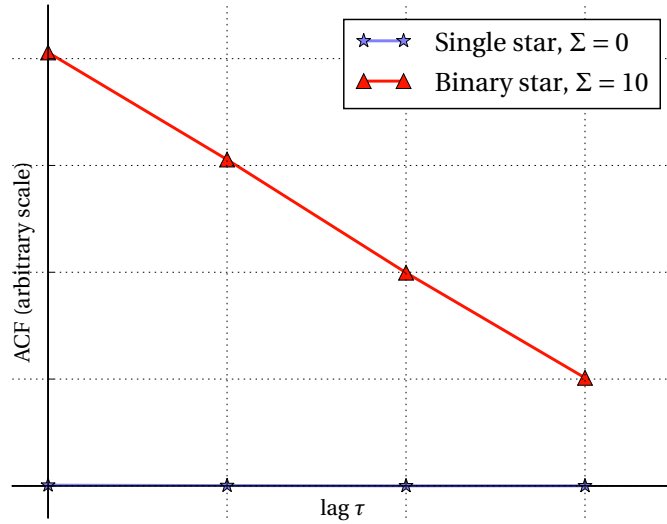


Figure 5.11 – Illustration of the result of the ACF for a relative error δ_p of a given shape parameter. We show the ACF for a single star (blue) and a binary system (red). For simplicity, the noise level in the shape measurement is set to unrealistically low values. The shape parameters of the single stars are uncorrelated between exposures, thus the ACF coefficients are all zero, yielding the scalar quantity $\Sigma = \sum_{\tau=0}^3 \text{ACF}(\tau) = 0$. The shape parameters are correlated for a binary system yielding a high $\Sigma = 10$.

There is more information available than the reconstruction error. The biases for different classes of binary systems can be simulated such that the distribution of biases are known. An algorithm can be trained to detect these patterns in the data. This leads to the second detection technique: random forests (Sect. 4.4). Features for RF are the same than in the ACF technique: the different shape parameter errors. With RFs, however, there are no channel. All the available data is used at once. The algorithm learns on a training catalogue that contains the ground truth, that is the binarity of the object.

When we drop the assumption that there is no dither between the frames, we are confronted to the PSF spatial variability. The relative errors (as in eq. 5.11) are dominated by the spatial variation. So, a more flexible approach is to teach the algorithm the fiducial PSF field as well as the possible bias. To increase the capacity of the detector, we introduce a detector based on ANNs (Sect. 4.3). The number of input features is increased: we use the deviations from the estimated parameters and the estimated parameters themselves. That entails three shape biases times four exposures plus the three fiducial shape parameters. We build networks with three hidden layers, each containing fifteen neurons, and train them with a MSE cost function to output the binarity of a star. For this method as well, a binary/single star criterion must be determined, *via* an additional validation training set. Because networks learn on simulations, more effects can be included, like a significant dither or a temporal evolution of the PSF field.

In the next paragraphs and in Sect. 5.1.6, we present proof-of-concept results on mock *Euclid* VIS-like simulations.

We use six hundred *Euclid* VIS PSFs that were simulated for the four central CCDs. Those PSF images are super-sampled by a factor of twelve. Measurements of these super-sampled images *via* the GalSim adaptive moments are taken as the fiducial value. A fiducial PSF field is generated from these simulations. The binary simulation procedure is the same as in Sect. 5.1.3. To determine the binary bias as a function of relative position host-companion and contrast at any point in the field, we have to interpolate a four dimensional space. The first dimension is the spatial position in the PSF field. We can measure binary biases at six hundred spatial positions. The second and third dimensions are the relative positions between primary and secondary stars. We measure about a thousand possible configurations per spatial positions. As most of the binaries have a small angular separations, we probe the small separations more. The pattern of the configurations is shown in Fig. 5.12. Finally, the fourth dimension is the contrast. Eight different contrast levels are probed. The number of measurements is $\sim 8.5 \cdot 10^2 \times 6 \cdot 10^2 \times 8 \simeq 4 \cdot 10^6$ configurations. From those interpolated measurements, *Euclid*-like fields are drawn. Selection cuts are performed to explore how binaries populating different regions of the contrast-separation space can be detected. Regardless of the value of the cutoff criteria, the fraction of binaries was artificially maintained to 30% to avoid a very low binary fraction, which would likely bias the metrics. As we simulate the four central CCDs only, the number of stars per catalogue is set to 280 ($= 4 \times 70$ stars). Catalogues are drawn with a level of noise such that the shape is known to 1% for the ellipticities and 5% for the size.

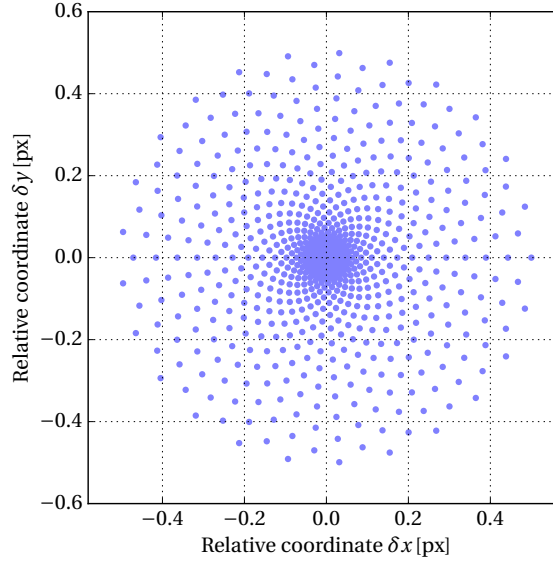


Figure 5.12 – Configurations of the binary systems for which the shape parameters were measured for every of the 600 *Euclid* VIS PSFs. The pixels are super-sampled by a factor twelve, so we explore a disk of radius half a VIS pixel, which virtually engulfs all possible *Euclid* binaries. The units in the plot are in VIS pixel.

The ACF and RF techniques are applied to toy examples for which the PSF fields are known to the *Euclid* requirements. We arbitrarily set the target false positive rate to 10% to have a basis of comparison. RFs in the `scikit-learn` implementation (Pedregosa et al. 2011) directly return a binary classification⁴, but the threshold for the ACF technique still needs to be determined. This is done by evaluating the scalar quantity Σ on training fields and find the threshold such that $\text{FPR} = 10\%$. Weights for the ACF channels are determined by their training F_1 score for the least constraining contrast-separation criteria. Predictions on the importance of channels were correct. The e_1 channel is the most important, than e_2 , and the size channel is completely neglected. The feature importance can also be determined for the RF approach and it presents the same ranking. In RF, the size feature contribute slightly to the results (See Fig. 7 in Paper II, Sect. 5.1.6). The tests show good classification performances on this best-scenario case, and suggest that about 50% of the binaries above the proposed 3 mas threshold are identified. Even if we do not reach our detection target, we do flag the most harmful cases.

The worst case scenario is when we have no knowledge on the PSF field and must apply the scheme proposed in Fig. 5.10. This is of course harder than the best case as the PSF field must be reconstructed from an unknown population of single and multiple stars. When the ACF and RF methods are applied as described in the above, they utterly fail. This is due to the input features. We detail the shortcomings of the relative error features and how to recover the detection capabilities in the next section. The introduction of ANNs with a different set of features can easily overcome the relative error shortcomings and the technique performs well,

⁴Classification should be understood as either class 0 or class 1.

with relatively high F_1 score. Artificial networks are able to detect about 50% of the binaries at a angular separation of 5 mas minimum with no prior knowledge on the PSF field. The PSF field in *Euclid* will vary with time, but not dramatically. This entails that there will be some information on the PSF field available from earlier exposures. If the PSF is stable over more than four exposures, the stellar density can also be increased by stacking catalogues from different fields.

5.1.6 *Paper II: Detecting unresolved binary stars in Euclid VIS images*

In Paper I (Sect. 5.1.4), we introduced the concept and studied the effects of unresolved binaries-induced biases on the PSF reconstruction. In the previous section, we explored strategies to mitigate the binary bias. We found two partial solutions to the problem (*Gaia* catalogue look-up and spectral type determination) and proposed a scheme based on statistical analyses of the point-source candidates light profile. Three methods were tested on mock catalogues to validate the concept of the method.

Paper II ([Kuntzer and Courbin 2017](#)) presents the methodology, the binary detectors and the results. The goal of this paper is to highlight that, while unresolved binaries in *Euclid* are an issue, solutions exist. This study was published in the journal *Astrophysics & Astronomy*, and is reproduced in this section.

(See next page.)

Detecting unresolved binary stars in *Euclid* VIS images[★]

T. Kuntzer and F. Courbin

Institute of Physics, Laboratory of astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny,
 1290 Versoix, Switzerland
 e-mail: thibault.kuntzer@epfl.ch

Received 16 March 2017 / Accepted 22 August 2017

ABSTRACT

Measuring a weak gravitational lensing signal to the level required by the next generation of space-based surveys demands exquisite reconstruction of the point-spread function (PSF). However, unresolved binary stars can significantly distort the PSF shape. In an effort to mitigate this bias, we aim at detecting unresolved binaries in realistic *Euclid* stellar populations. We tested methods in numerical experiments where (i) the PSF shape is known to *Euclid* requirements across the field of view; and (ii) the PSF shape is unknown. We drew simulated catalogues of PSF shapes for this proof-of-concept paper. Following the *Euclid* survey plan, the objects were observed four times. We propose three methods to detect unresolved binary stars. The detection is based on the systematic and correlated biases between exposures of the same object. One method is a simple correlation analysis, while the two others use supervised machine-learning algorithms (random forest and artificial neural network). In both experiments, we demonstrate the ability of our methods to detect unresolved binary stars in simulated catalogues. The performance depends on the level of prior knowledge of the PSF shape and the shape measurement errors. Good detection performances are observed in both experiments. Full complexity, in terms of the images and the survey design, is not included, but key aspects of a more mature pipeline are discussed. Finding unresolved binaries in objects used for PSF reconstruction increases the quality of the PSF determination at arbitrary positions. We show, using different approaches, that we are able to detect at least binary stars that are most damaging for the PSF reconstruction process.

Key words. methods: data analysis – methods: statistical – binaries: close

1. Introduction

Future space-based observatories such as *Euclid*¹ (Laureijs et al. 2011) and WFIRST (Spergel et al. 2015) require exquisite point-spread function (PSF) stability and measurement to achieve their scientific goals, in particular in view of the weak-lensing cosmological probe (Cropper et al. 2013; Massey et al. 2013; Schneider 2006). Several instrumental effects have been identified as nuisances in the PSF determination process, such as the colour of the object and colour gradients across the objects (Voigt et al. 2012; Semboloni et al. 2013), the brighter-fatter effect (Antilogus et al. 2014), or detector distortions that are due to temperature gradients.

Other astrophysical or atmospheric nuisances also contribute to the degradation of the signal, such as satellite trails, cosmic rays, stray light (e.g. Desai et al. 2016), or PSF distortions by unresolved binary stars (Kuntzer et al. 2016a, KC16 hereafter). We showed in KC16 that unresolved binaries can significantly alter the shape of the *Euclid* PSF, even if many (thousands of) stellar images are used together to reduce this effect.

In this paper, we propose a technique to identify unresolved binary stars that may hamper a proper PSF measurement. We base our method on the systematic nature of the deformation induced by binary stars. While the deformation of individual stellar images is of the order of the instrumental noise, correlations between different exposures of the same object can lead to the

detection of a binary star, or at least to flagging an object as non-reliable for subsequent PSF determination.

A similar approach was used by Hoekstra et al. (2005) to detect eclipsing binaries in the Optical Gravitational Lensing Experiment (OGLE). Detecting unresolved binaries from the ground using this approach has already proven successful (Terziev et al. 2013; Deacon et al. 2017). We build on the technique pioneered by Hoekstra et al. (2005) but apply more sophisticated algorithms, including artificial neural networks (ANN). We also study the applicability of our techniques to the specific case of *Euclid*, with the goal of identifying unknown and unresolved binaries that affect the quality of the PSF reconstruction, as described in KC16.

The observing strategy of *Euclid* is to take four dithered exposures (Cropper et al. 2016) of each field for the weak-lensing surveys. The deep fields, located near the galactic poles, will be repeatedly observed over the course of the mission, at different orientations. These multiple images of the same stars can be used, when combined, to reconstruct an estimation of the PSF to account for its effect. They can also be individually used to check for systematic biases in the shape estimates of the light profiles. Flagged objects can be removed from the stellar catalogues used in view of the PSF reconstruction.

In this proof-of-concept paper, we choose to work at the catalogue level, and not to measure *Euclid*-simulated images of single or multiple images. Our choice to work at the catalogue level is deliberate. Our goal is to show that our approach can be an efficient method to flag unresolved binaries when a catalogue of stars with shape measurements is available. We do not aim at building a complete PSF shape measurement and reconstruction pipeline.

[★] The code corresponding to the algorithms used in this work and all scripts to reproduce the results are publicly available from a GitHub repository accessible via <http://lastro.epfl.ch/software>

¹ <http://www.euclid-ec.org/>

We show on simulated catalogues of non-dithered observations that we can robustly detect unresolved binary stars in a sample of point sources. This article is organised as follows: we detail the algorithms and associated performance metrics in Sect. 2. Section 4 presents a summary of the statistical knowledge of the binary systems, and Sect. 5 discusses the mock observations of stars with the VIS instrument. We then illustrate the performance of the algorithm with image simulations in Sect. 6. Finally, Sect. 7 summarises our findings.

2. Definitions, scheme, and training data

We propose different algorithms that learn the difference between systematic and stochastic distortions to predict whether an object is a single or a binary star. The examples used for training (in the machine-learning sense) could be drawn from real data of sources whose binary nature is known or from simulated data.

2.1. Effect of binaries on the PSF shape parameters

A more in-depth discussion of the effect of binaries on the shape of the PSF is provided in KC16. In this paragraph, we summarise the main results of KC16 very briefly. Throughout, we use the term contrast for the difference in magnitude Δm between the host (or main) and companion star $\Delta m = m_{\text{companion}} - m_{\text{host}}$. The presence of a companion can change the measured PSF ellipticity by an order of 1% when the angular separation between host and companion is $r = 0''.01$ (i.e. a tenth of the *Euclid* pixel). Similarly large effects are seen on the size measurement. The distortion of the PSF induced by a companion star depends on the angular separation and on the contrast in magnitude KC16. The expected distortions of the PSF caused by binary objects are systematic in nature. On average, the deformations due to binary stars tend to make the PSF rounder and larger. On a case-by-case basis, however, the deformation of the PSF may not make it rounder, but could increase the ellipticity.

The deformation of the PSF measured on the image of a binary system depends on the relative positions of the host and the companion, and on the magnitude contrast. For the wide-field survey of the *Euclid* mission, the exposures are taken in a sequence, thus the physical properties describing a binary system will not change. In the deep exposures, taken in a much longer time interval, the properties of the binary system may change.

2.2. Definitions

For simplicity, we assume that a given star is imaged in several exposures at similar spatial positions on the detector. The observables used to classify an object are the shape parameters of the PSF. These observables, which are impervious to the measurement method, are the complex ellipticities (e_1 and e_2) and the PSF “size” R^2 . These can be defined using quadratic moments as follows.

Let $I(\theta)$ be the brightness distribution of an image. $\bar{\theta}$ would be the position of the centroid. The tensor of second-order moments q_{jk} computed along the first and second axis of an image is given by

$$q_{jk} = \frac{\int (\theta_j - \bar{\theta}_j)(\theta_k - \bar{\theta}_k) I(\theta) d^2\theta}{\int I(\theta) d^2\theta}, \quad j, k \in \{1, 2\}. \quad (1)$$

A119, page 2 of 10

In a noise-free image, the complex ellipticity components of a light profile are obtained from Kaiser et al. (1995):

$$[e_1, e_2] = \left[\frac{q_{11} - q_{22}}{q_{11} + q_{22}}, \frac{2q_{12}}{q_{11} + q_{22}} \right]. \quad (2)$$

The size of the PSF is defined as

$$R^2 = q_{11} + q_{22}. \quad (3)$$

The standard deviations of the variations in the PSF for the *Euclid* survey must be smaller than $\sigma(e_i) \leq 2 \times 10^{-4}$ for the ellipticity components and $\sigma(R^2)/\langle R^2 \rangle \leq 1 \times 10^{-3}$ for the size according to science requirement 4.2.1.4 (Duvet et al. 2010; Paulin-Henriksson et al. 2008; Laureijs et al. 2011). Each *Euclid* image will contain from 2000 to 3000 stars (Cropper et al. 2013; Laureijs et al. 2011), depending on Galactic latitude. If the PSF reconstruction scheme is optimal, the knowledge on each individual star is $\sqrt{2000} \approx 50$ worse than the requirements, that is, 1% for the ellipticities and 5% for the size on each individual star.

2.3. Simulated data

We stress that we work at the catalogue level and not at the pixel level. We assume that the PSF can be measured to within the *Euclid* expected performances (as stated in the previous paragraph) for each individual star. From a catalogue, we can flag suspicious observations of what has been considered as a single star. The precision and accuracy we take on the shape parameters are given by the *Euclid* requirements. The preparation of a PSF pipeline that meets the *Euclid* requirement is beyond the scope of this paper. For this reason, we chose to work with a catalogue that fulfils the claimed requirements. In this way, our results do not depend on any given shape measurement method either.

For the observations of single stars, we assume that the observed PSF parameters are noisy around their fiducial value (we discuss this assumption in Sect. 6.4). In the case of a binary system, however, the observables e_1, e_2 , and R^2 are systematically modified by a value that depends on the angular separation and the contrast of the system. We assume throughout that the PSF parameters e_1, e_2 , and R^2 are measured to the above precision, that is, 1% (ellipticity) and 5% (size) per star observation.

Catalogues of simulated single and binary system observations are produced by the following procedure and described in more details in Sect. 5:

1. Draw a stellar population from the Besançon Galaxy Model (BGM, Robin et al. 2003) for a *Euclid*-like sky.
2. Using empirical knowledge on the distribution of binary stars (Duchêne & Kraus 2013), compute the confidence that any given star is a binary system and compute its characteristics (semi-major axis of the orbit and mass ratio).
3. Draw a realisation of the binary population with selections in angular separation and contrast.
4. Estimate the observed PSF parameters for each object in the catalogue and for each exposure.

2.4. Principle of the binary detection algorithms

Our detection algorithms of unresolved binaries follows this procedure and is illustrated in Fig. 1.

1. Compute the relative error, $\delta^{(i)}$, based on the PSF parameters $p^{(i)} = e_1^{(i)}, e_2^{(i)}$ or $(R^2)^{(i)}$ for a star i with respect to a fiducial

T. Kuntzer and F. Courbin: Detecting unresolved binary stars in *Euclid* VIS images

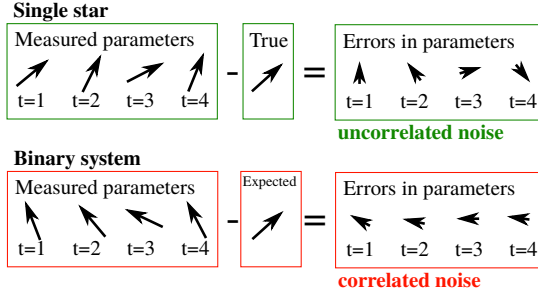


Fig. 1. Sketch of the measurement principles. The four observations are compared to the fiducial value of the shape parameter. A binary star will affect the shape of the PSF systematically across the four exposures (bottom), whereas the measurement errors are uncorrelated (top).

value $p_0^{(i)}$. The relative error is

$$\delta^{(i)} = \frac{p^{(i)} - p_0^{(i)}}{p_0^{(i)}}. \quad (4)$$

For each star i , four such relative errors are measured for each parameter because there are four observations. We denote access to the relative error in exposure t as $\delta^{(i)}(t)$. The fiducial parameter fields are either known a priori or are directly inferred from the measurement PSF parameters using an iterative field estimation and outlier removal algorithm. When the PSF field is unknown, the PSF at a given point is reconstructed using the parameters of neighbouring single stars. The interpolated value can be interpreted as the fiducial value at that point. When the measured relative errors are too different from the fiducial value, the object is considered as an outlier and is removed from the list of single stars. This process of interpolating a fiducial value based on neighbouring single stars and marking it single or outlier is repeated several times. This algorithm is detailed in Sect. 6.2.

2. Use the resulting relative errors $\delta^{(i)}$ as the features for the algorithms to infer the presence of binaries.

3. Algorithms to identify unresolved binary stars

We propose three different classifiers that yield a binary output (binary star or single star) or, similarly, a number between 0 and 1 that translates the confidence of detection. The three classifiers tested here are (i) a simple auto-correlation of the input features, and two supervised-learning methods, namely (ii) random forests (RF) and (iii) ANNs. All three methods are evaluated on two binary system selection criteria: their (i) angular separation, and (ii) contrast.

3.1. Algorithm 1: auto-correlation function

For each object i in the field, we consider a list of four relative errors $\delta^{(i)}$ for each of the shape parameters. These lists can be understood as time series of the relative errors. As described in Sect. 2.4 and illustrated in Fig. 1, when the object is a single star, the errors made on the measured parameters are stochastic. However, when the object is a binary system, the errors are correlated because the relative position of the main and companion star do not change between the first and last exposure. A measure of the correlation of the signal with itself can distinguish

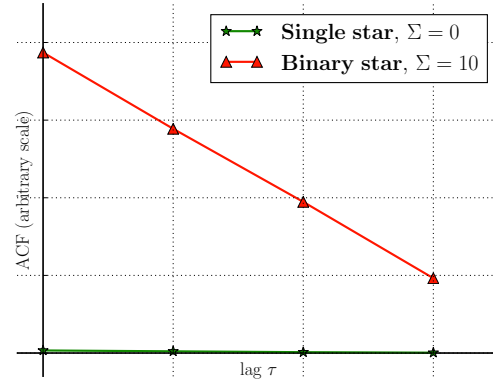


Fig. 2. Illustration of the result of the ACF for a relative error δ of a given shape parameter. We show the ACF for a single star (green) and a binary system (red). For simplicity, the noise level in the shape measurement is set to unrealistically low values. The shape parameters of the single stars are uncorrelated between exposures, thus the ACF coefficients are all zero, yielding the scalar quantity $\Sigma = \sum_{\tau=0}^3 \text{ACF}(\tau) = 0$. The shape parameters are correlated for a binary system, and a strong auto-correlation between exposures is observed, with a high scalar quantity $\Sigma = 10$.

between no correlations (low auto-correlation score) or correlations (high auto-correlation score), which acts as a proxy to the binary or single star nature of the object.

An estimate of the auto-correlation function (ACF) coefficient for a lag of τ for a vector of relative errors δ observed at $t = \{1, 2, 3, 4\}$ is given by

$$\text{ACF}(\tau) = \frac{1}{(n-1)\sigma^2} \sum_{t=1}^{n-1} [\delta(t) - \bar{\delta}][\delta(t+\tau) - \bar{\delta}], \quad (5)$$

where σ^2 is the sample variance, and $\bar{\delta}$ is its mean value as computed from n observations of δ . $n = 4$ exposures in our work. For simplicity, we dropped the star index i in the above equation. This method is data driven: it is based on the assumption that the relative errors are more strongly correlated for binary systems than for single systems. No physics is included in the model. We sketch the result of the ACF method for a binary system and a single star in Fig. 2.

We sum the four estimated ACF coefficients for the four observations to create a scalar quantity $\Sigma = \sum_{\tau=0}^3 \text{ACF}(\tau)$. We choose a value for a separating threshold that classifies highly auto-correlated observables as binary stars and, conversely, links low auto-correlation of the observables to single stars. Each PSF parameter yields an estimate of the auto-correlation. We refer to each of these classifications as e_1, e_2 , and R^2 channels. The outputs of the three detection channels can be combined to improve the classifiers' performance. A three-channel classifier is obtained by taking a weighted average of the output of each individual channel. The weights are empirically determined by evaluating the individual channel's performance. The prediction of the three-channel classifier is again compared to a separating threshold. The ACF classifier is run against a training dataset to determine the separating threshold, which is set such that the false-positive rate is 10%.

3.2. Algorithm 2: random forest

When the expected distortions of the shape parameters are known, we can train an algorithm to recognise patterns in the data. The random forest (RF) algorithm is a large ensemble of simple but weak decision trees. Decision trees are a class of non-parametric supervised-learning classifier or regression methods that hierarchically approach the data. The training of a decision tree consists of learning a set of decision rules based on the features and the ground truth (Breiman 2001). Decision trees taken individually tend to overfit the data. Averaging over a large number of trees – a forest – reduces the classification error.

In the same way as for the ACF classifier, the features adopted for the RF are the relative errors made on the PSF parameters as given by Eq. (4). The ground truth is the nature of the object (binary or single star). The depth of the individual trees is not limited, and a forest of 50 trees is constructed.

We used the RF implementation of the Python package `scikit-learn` with the Gini impurity criterion (Pedregosa et al. 2011).

3.3. Algorithm 3: artificial neural networks

Feed-forward artificial neural networks (ANNs) belong to a class of supervised algorithms that can be used for regression or classification tasks. We applied ANNs to classify binary and single stars. For an introduction to ANNs, we refer to Bishop (1995). The ANNs are made of an input layer, an arbitrary number of hidden layers, and an output layer. The hidden and output layers consist of a number of neurons that take as input a vector \mathbf{x} and return a scalar $h(\mathbf{x}, \mathbf{w}, b)$ through

$$h(\mathbf{x}, \mathbf{w}, b) = h\left(\sum_{i=1}^N w_i x_i + b\right), \quad (6)$$

where \mathbf{w} and b are the weights and the bias, respectively. The monotonic and continuous function h is the so-called activation function. In the following, we use the hyperbolic tangent as activation function. The ability of an ANN to find patterns shared by the input data (the features) and the labels of each image (the ground truth and its prediction) depends on the architecture of the network, in that case, the number of neurons and layers. The architecture must be optimised for specific applications. We found that for our application, three layers of 15 neurons each performs well. We trained the network on a standard mean-square-error cost function. The size of the training and test datasets is given in Sect. 6.3.

The choice of the ANN implementation, like the choice of the RF implementation, is arbitrary from the outset. A comparison of the details of each implementation is beyond the scope of this paper. We used the Python bindings for the Fast Artificial Neural Network Library (FANN) implementation and its optimiser described in Nissen (2003).

The ACF and the RF methods were applied to the simulated data described in Sect. 6. The performance of the ANNs is analysed in Sect. 6.3 on worst-case experiments only, namely on an unknown field in both stellar population and PSF field parameters.

3.4. Metrics for the detection performance

We used three metrics to quantify the effectiveness of detecting binary stars:

- The receiver operating characteristic (ROC) is a graphical representation of the performance when the decision

threshold is varied. The abscissa represents the false-positive rate (FPR, or fall-out),

$$\text{FPR} = \frac{\text{false positives}}{\text{false positives} + \text{true negatives}}, \quad (7)$$

while the ordinate encodes the true-positive rate (TPR, also called sensitivity and recall),

$$\text{TPR} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}. \quad (8)$$

An algorithm that would randomly classify a star as a binary lives on the diagonal $\text{FPR} = \text{TPR}$. An ideal classifier would be represented by a curve passing through the coordinate $[\text{FPR} = 0, \text{TPR} = 1]$. The ROC allows for a quick and reliable comparison between different classifiers, marginalising over the detection threshold. The choice of the separating threshold can be based on the ROC by trading off the TPR for the FPR (Kleinbaum & Klein 2010).

- The *area under the curve* (AUC) is the integral of the ROC over the FPR. It summarises the ROC as a single scalar. A random algorithm would score $\text{AUC} = 0.5$, while an ideal method would reach $\text{AUC} = 1$. For this metric to be high, the true positive rate and the precision (the ratio of the true positives to the sum of true and false positives) must both be high.
- The F_1 score is a scalar metric widely used for binary classification, defined as

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}, \quad (9)$$

where TP, FN, and FP are the numbers of true-positive, false-negative, and false-positive classifications, respectively (e.g. Herlocker et al. 2004). This metric is not only sensitive to the number of correctly detected binaries (TP), but also to the number of binaries classified as single stars (FN) and inversely (FP). A perfect classification would yield $F_1 = 1$, while a completely random decision-making algorithm would have an F_1 score that tends to zero.

4. Binary star population in Euclid

In order to test our binary star finders, we designed *Euclid*-like simulated images, building on a stellar population with realistic binary fractions. The latter as well as the orbital parameter distributions and luminosity contrasts have been studied in volume-limited searches. Early searches were restricted to low numbers of stars, hence leading to incomplete samples (e.g. Abt & Levy 1976; Marcy & Benitz 1989; Duquenoey et al. 1991). More recent searches harvested more data, but the total sample of stars with a detailed characterisation is still limited to a few hundred (Raghavan et al. 2010; Milone et al. 2012; Cummings et al. 2014; Nardiello et al. 2015; Riddle et al. 2015; Rodriguez et al. 2015).

Duchêne & Kraus (2013) compiled the available data and proposed distributions of multiple systems in terms of multiple star fraction, mass ratio to the main star, and angular separation. We used these parameters to create a stellar population based on BGM realisations and with realistic binary fractions. Figure 3 shows the contrast and the angular separation of binary stars as a function of distance to Earth for stars located in the direction of the anti-Galactic centre, with Galactic latitude in the range

5.1. The effect of unresolved binaries on PSF determination

T. Kuntzer and F. Courbin: Detecting unresolved binary stars in *Euclid* VIS images

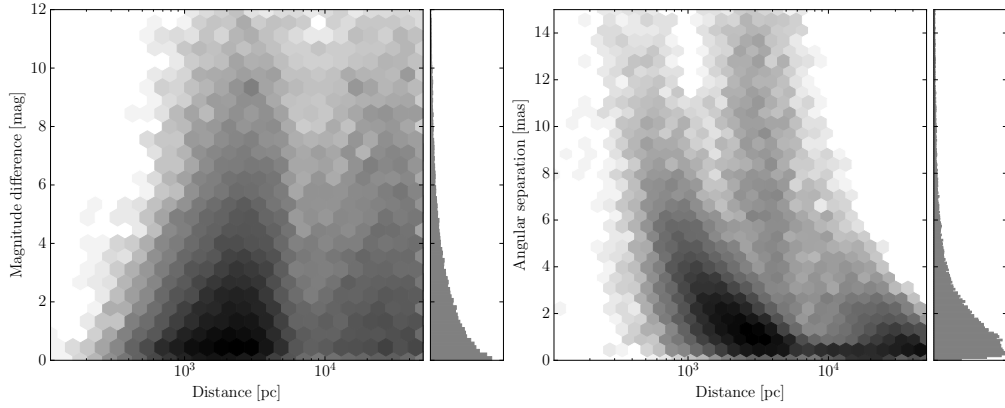


Fig. 3. Numerical model of the binary star density as a function of distance to Earth and contrast (*left*) and angular separation (*right*). The grey scales and the histograms are linear and relative to the maximum bin. We select stars in the magnitude range $18 \leq I(\text{AB}) \leq 24.5$. Both plots exhibit bimodal distributions with low-mass main stars located at shorter distances. More massive stars are not located in our direct neighbourhood, as suggested by the distribution of angular separation with distance. Companions of very low mass stars (down to the brown dwarf limit) are rare with the applied selection cuts.

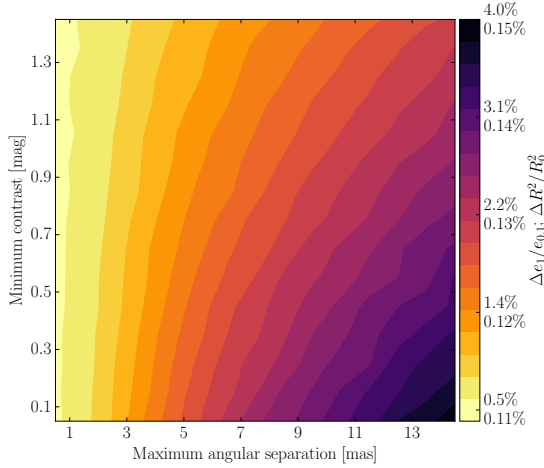


Fig. 4. 95-percentile relative errors of the PSF shape parameters e_1 and the size. The relative errors in e_2 are about twice smaller. The values of the relative errors are given in the colour bar for the ellipticity (top number) and for the size (bottom number). The magnitude selection is $18 \leq I(\text{AB}) \leq 24.5$. The distributions of relative errors contain all binaries that have at most an angular separation on the x -axis and at minimum a contrast on the y -axis.

$+15^\circ < b < +75^\circ$. We selected stars when their apparent magnitude was in the range of the expected magnitude for the *Euclid* wide-field survey (i.e. $18 \leq I(\text{AB}) \leq 24.5$). The bimodal nature of the distributions in Fig. 3 is explained by the mass of the main star. Massive stars in the selected magnitude range are rare in the neighbourhood of the Sun, and inversely, low-mass stars are not selected by our cut in magnitude at large distances. Figure 3 also presents histograms of the stars as a function of contrast and angular separation. They demonstrate that most binary systems that will be seen by *Euclid* have a low contrast, meaning that both stars are similar in terms of magnitude and

have a very small angular separation, with about 75% of the binaries separated by less than 3.5 milliarcseconds (mas) and 50% below 1.7 mas. Most companions have a magnitude similar to the main star of the system, but a wide diversity of contrast is expected. Systems that contain brown dwarf companions are also predicted.

Figure 4 illustrates the magnitude of the relative errors due to binaries at small separations and low contrast. We show the 95-percentile value of the expected relative errors in the shape of PSFs over a given binary population. The intuition that larger relative errors occur at larger angular separation and low contrast is confirmed. Quantitatively, the relative errors follow the same trends, but the effect is strongest on the e_1 component, followed by e_2 , whose relative errors are about half as important. The relative errors induced on the size are an order of magnitude weaker than on e_2 . Since the size is also predicted to be measured five times less accurately than the ellipticities (see Sect. 2.3), the latter are more sensitive indicators of the presence of binaries than the former.

The change in PSF shape due to the presence of binary stars is not yet budgeted in *Euclid*, and there is no requirement available so far on the performance of any binary rejection method. However, Fig. 4 indicates a detection limit of ~ 3 mas in separation almost independently of contrast. This ensures that the distortions due to binaries are of the same order of magnitude as the measurement error per star for the ellipticity, that is, $\sim 1\%$.

5. Mock catalogues of PSF shape parameters

In this section, we explain how we create the catalogues of PSF parameters for any given single and binary system at any spatial position in the four central CCDs of the *Euclid* VIS detector.

For the shape parameters contained in the catalogues to be in a realistic range, we measured super-sampled noise-free *Euclid*-simulated PSFs. We simulated 600 *Euclid* PSFs at randomly selected positions on the detector using the ray-tracing tool Zemax (Ngolè Mboula et al. 2015). The images of these PSFs are super-sampled by a factor of 12 (i.e. the pixel is one-twelfth of $0''.1$). The shapes of the simulated single

and binary stars were measured using the adaptive-moment scheme proposed by Bernstein & Jarvis (2002) and based on re-Gaussianisation (Hirata & Seljak 2003) as implemented in Galsim (Rowe et al. 2015). To obtain a realistic stellar population, we used the BGM, with the *Euclid* cuts on magnitude ($18 \lesssim I(\text{AB}) \lesssim 24.5$) combined with the information on the estimates of the fraction of multiple stars and orbital parameter distributions derived from Duchêne & Kraus (2013). We restricted the stellar population to astrophysical binaries. Coincidental binaries of field stars are negligible given the star density and field of view (of the order of 0.001%). We derived the rate of coincidental binaries from our catalogues of stars, estimating a high density of 5000 stars in the *Euclid* field of view assuming a uniform distribution of the stellar positions. As a simplification, images of binary and single stars were produced using a flat spectral energy distribution (SED). In real data, the stellar SEDs will be provided by the *Gaia* catalogue (de Bruijne 2012; Gaia Collaboration 2016), ground-based multi-band data, or methods such as the VIS single-image spectral classifier as proposed in Kuntzer et al. (2016b). The measurements provide a set of fiducial shape parameters and sample the distortions due to binary stars. Mock catalogues can be drawn from these fiducial values and distortions. These catalogues contain the shape parameters (e_1 , e_2 and the size) for each object in the observed field.

To reflect the *Euclid* observing strategy, four realisations of mock catalogues were prepared of the same field. The noisy shape parameters were assumed to be known, as we described in Sect. 2, to 1% for the ellipticities and 5% for the size. The shape parameters were computed by interpolating their fiducial values from a set of the 600 *Euclid* PSFs. The distortions due to a given binary system were also interpolated. The distortions were interpolated at their spatial position on the image, but also in terms of position of the binary with respect to its host and contrast. The mock catalogues of the stars contained four noisy values for each shape parameter, corresponding to the four dithered observations.

After computing the interpolation, the PSF shape parameters for any system (single or binary in all its variety) can be derived and at any position. Selection cuts were imposed on the contrast and the angular separation of the binary system during the preparation of the catalogues. Regardless of the value of the criteria, the fraction of binaries was artificially maintained to 30% to avoid a very low fraction of positive samples in the data, which would bias the metrics.

In the following numerical experiments, we only work on the catalogues.

6. Numerical experiments

The binary classifier algorithms were applied to two different experiments: (i) the PSF field is known to the *Euclid* requirements, a best-case experiment; and (ii) the PSF field is not known beforehand, the worst-case experiment. For each experiment, a training and test dataset were prepared according to Sect. 5. The training set was necessary in the case of the ACF method to determine the separating threshold. For RF it was used to build decision trees.

In the worst-case experiment, the training set was used to optimise the weights and biases of the ANNs and the threshold was determined on an additional validation set. The results reported in the following were measured on the test set that was similar to the training set, but was not shown to the algorithm during

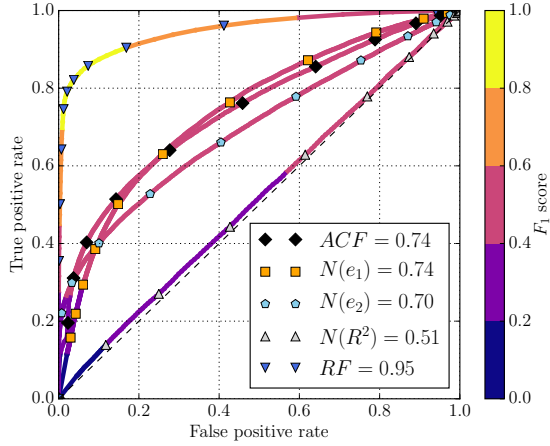


Fig. 5. ROC curve for a 5 mas minimum separation and a contrast of up to one magnitude. The curves for the single channel (e_1 , e_2 , and R^2 , denoted $N(e_1)$, $N(e_2)$, and $N(R^2)$) are shown along with the combined three-channel classifier (denoted ACF) and the RF method. All curves are colour-coded with the F_1 score computed at each threshold. The dashed line shows the performance of a random-guess algorithm. The numerical value in the legend is the AUC of the method.

the training. The selection criteria (contrast and angular separation of the binaries) in both experiments evolved on a grid. The angular separations ranged between 1 and 15 mas. The angular separation is a lower bound criterion, while the contrast is an upper bound criterion.

6.1. Known PSF field experiment

In this experiment, we assumed that the PSF shape parameters are known to the *Euclid* requirements at any spatial position, namely $\sigma(e_i) \leq 2 \times 10^{-4}$ for the ellipticity components and $\sigma(R^2)/\langle R^2 \rangle \leq 1 \times 10^{-3}$ for the size (Duvet et al. 2010; Paulin-Henriksson et al. 2008; Cropper et al. 2013; Massey et al. 2013). The number of objects in the datasets depends slightly on selection criteria, but amounts on average to a few thousand objects.

For every pair of criteria, the threshold selected on the training set produced the targeted 10% FPR in the testing phase. An ROC curve was calculated for each of the individual channels (i.e. classifying on each of the shape parameters individually) and the three-channel classifier (i.e. taking into account all shape parameters to classify). The ROC curves are displayed in Fig. 5. The resulting AUC, an indicator of the performance of each of the channel, was used as weight for the three-channel ACF. The weights were left fixed for all criteria pairs.

The most useful information to predict the nature of the object is inferred from the e_1 relative error feature (Eq. (4)). The second most useful information is the second ellipticity component. While its overall performance is poorer than the e_1 channel, the e_2 channel has a better TPR at low FPR. This better performance at low FPR implies that a combination of the two ellipticity channels leads to a better overall classification. The ACF size channel classifier is a very poor predictor of binary systems. This remains true for most choices of selection criteria, short of very large angular separations.

5.1. The effect of unresolved binaries on PSF determination

T. Kuntzer and F. Courbin: Detecting unresolved binary stars in *Euclid* VIS images

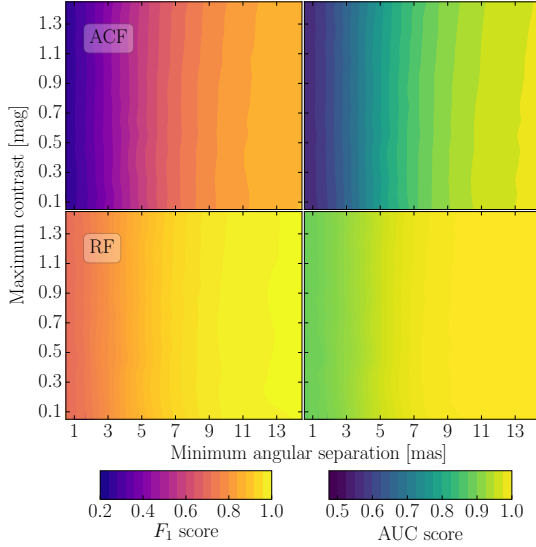


Fig. 6. Performance metrics F_1 (left) and AUC (right) for different combinations of selection criteria in the known PSF experiment, using the ACF method (top panel) and RF method (lower panel) to detect the binaries. The parameters of the PSF shapes are known to 1% for the ellipticity and 5% for the size.

As expected, better performances in terms of both AUC and F_1 were obtained when the angular separation was increased. This is illustrated in Fig. 6. At small angular separations, the ACF classifier is barely better than a random classifier, with $F_1 \sim 0.2$ – 0.3 . At large angular separations however, the performance of the classifier is greatly improved. In general, the performance of the classifier mostly depends on the angular separation and very little on the maximum luminosity contrast. The separating threshold chosen was chosen on the training dataset such that the FPR was 10%. The value of this separating threshold was the same for any choice of selection criteria. The separating threshold was thus not influenced by the selection criteria, which simplified the application of the ACF technique.

We then tested the RF method using the same training and test datasets as for the ACF. The RF method consistently obtained better scores than the ACF, as is shown in Fig. 6. The ROC curves for the RF method (Fig. 5) also indicate a better AUC and a higher F_1 score for the same FPR. The separating threshold is determined during the RF training to reach the goal FPR of 10%. The relative importances of the features in the RF algorithm show, according to the ACF method, that the e_1 relative errors (see Fig. 7) encompass the most decisive information. The e_2 features come in second, while the size information is third and negligible in the decision-making process. The RF method is only weakly dependent on the contrast, like the ACF. When the PSF field is known (to the *Euclid* requirement), the RF method in particular can be used to find binary systems even at small angular separations.

6.2. Unknown PSF parameter experiment

We now turn to an experiment where we assumed no prior knowledge of the PSF field. To follow the procedure laid out in Sect. 2, the PSF shape parameters must first be determined at the position

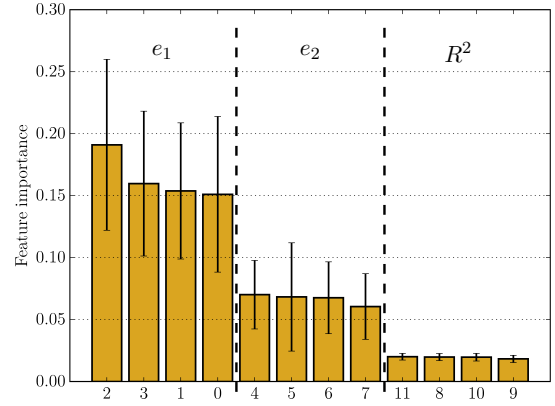


Fig. 7. Feature importance for an RF classifier with the selection criteria set at an 8 mas minimum separation and a contrast of up to half a magnitude. Features 0 to 3 encode e_1 relative errors for each exposure, features 4 to 7 are the same for the e_2 component, and features 8 to 11 are the same for the size. The error bar shows the 1σ deviations of the importance of the feature on the learned dataset.

of the objects of interest and the field of PSF shape parameters estimated. To this end, we adopted a “leave-one-out” scheme, described in algorithm 1. The PSF shape parameters at the position of the objects are interpolated from the nearest ten and assumed single neighbours and taken as fiducial parameters. The shape parameters of the objects are then compared to the fiducial parameters. The resulting relative errors are then used as the vector of features for the classifiers. The object is then classified into as a single or binary star. At the first iteration of algorithm 1, we assumed that all stars in the field are single. The above procedure was repeated to iteratively construct a good estimate of the PSF field, and to determine where the binaries are in the field of view.

The VIS instrument will be comprised of 6×6 CCDs, each with $4k \times 4k$ pixels, with a field of view of $0.787 \times 0.709 \text{ deg}^2$ (Racca et al. 2016). About 2000–3000 objects will be used to reconstruct the PSF field in any pointing (Cropper et al. 2013; Laureijs et al. 2011). In the present work, we only have PSF estimates for the four central detectors. The number of stars seen in this reduced field is 280, and we still use a binary fraction of 30%.

We determined the threshold for the ACF and inferred the decision rules for the RF on the training set in a similar way as in the known PSF field experiment. To test the performance of the methods, we created 20 different fields, thus different spatial positions and binary populations, each containing 280 stars. The reported metrics were averaged over the 20 fields. The relative error estimator during the testing phase was modified from Eq. (4) to

$$\delta^{(i)} = \frac{p^{(i)} - p_0^{(i)}}{p_0^{(i)} + \epsilon}, \quad (10)$$

where ϵ is a calibration coefficient accounting for the PSF field reconstruction errors. For RF, ϵ_{RF} varies from 0.03 to 0.07 as a function of the selection criteria, while ϵ_{ACF} is in the range [0.01, 0.035]. If this modification is not enforced, the FPR increases several times.

Algorithm 1 “Leave-one-out” reconstruction scheme used in the case of an unknown PSF field for the ACF and RF methods in the training phase. p denotes the PSF parameters for all stars and the four exposures, and δ is the relative errors.

```

1: procedure UKNPSF( $\text{star\_list}, p$ )
2:    $\text{single\_stars} \leftarrow \text{star\_list}$ 
3:   for iteration < max_iterations do
4:      $\text{iter\_single\_list} \leftarrow []$ 
5:     for each star in  $\text{star\_list}$  do
6:        $\text{usable\_stars} \leftarrow \text{single\_stars} - \text{star}$ 
7:        $K \leftarrow \text{Get10ClosestStars}(\text{usable\_stars}, \text{star})$ 
8:       #  $K$  is the list of the 10 nearest stars
9:        $w \leftarrow \text{InverseDistanceSquare}(K, \text{star})$ 
10:       $p^{(K)} \leftarrow \text{MedianOverExposures}(p^{(K)})$ 
11:      #  $p^{(K)}$  is an array of  $K$  stars  $\times$  3 shape parameters,
      averaged over exposures
12:       $p_0^{(\text{star})} \leftarrow \text{WeightedAverage}(p^{(K)}, w)$ 
13:      #  $p_0^{(\text{star})}$  is a vector of the 3 shape parameters
14:       $\delta^{(\text{star})} \leftarrow (p^{(\text{star})} - p_0^{(\text{star})}) / p_0^{(\text{star})}$ 
15:      #  $\delta^{(\text{star})}$  is an array of 4 exposures  $\times$  3 parameters
16:      if Classification( $\delta^{(\text{star})}$ ) == Single then
17:        append(star, iter_single_list)
18:      end if
19:    end for
20:     $\text{single\_stars} \leftarrow \text{iter\_single\_list}$ 
21:  end for
22:   $\text{binary\_list} \leftarrow \text{star\_list} - \text{single\_stars}$ 
23:  return binary_list
24: end procedure

```

The ACF method performs well in the case of an unknown PSF field. Even if its overall performance is reduced, it remains close to the known PSF field experiment. Figure 8 shows the ROC curve in the unknown PSF case. The ACF principle is simpler in the sense that it does not need to train on the features themselves. It only has to find a threshold between single and binary stars based on the degree of correlation of the features. The RF method relies on the individual relative error estimates to infer the binary nature. The interpolation of PSF parameters and determination of a binary star loop is not encoded in the training of the RF, which reduces the performance. For both methods, the FPR is significantly reduced between the first iteration (in which no binaries are assumed in the field) and the second iteration to settle to about 10% as required. After this, the FPR stays roughly constant, while the F_1 score tends to increase slightly. The ACF seems to outperform the RF technique, both in terms of AUC (reaching a maximum of about 0.9 – Fig. 9) and F_1 score, especially at high angular separations. The performance of the classification again mostly depends on the minimum angular separation and not on maximum contrast. In Figs. 6 and 9, the selection is made based on maximum contrast. If the selection were made on minimum contrast, there would be a stronger effect. However, we are interested in removing the most damaging objects from the PSF samples, and systems with a high contrast will not be the most damaging during the PSF reconstruction (see Fig. 4). The inclusion of the hyperparameter ϵ does succeed in producing good detection, but it is determined manually, which can hamper the image reduction pipeline. For this reason, we present in the next paragraph the ANN approach, which does not require the inclusion of the hyperparameter ϵ .

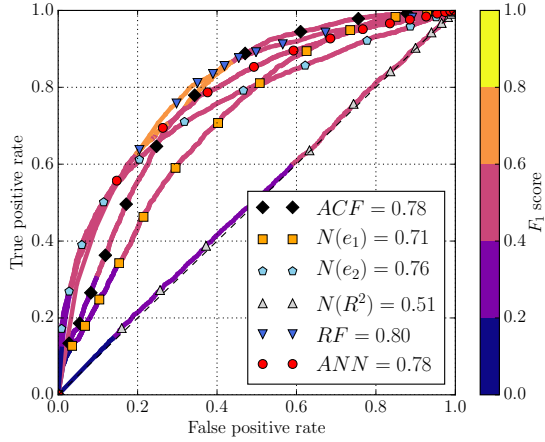


Fig. 8. ROC curves for the unknown PSF experiment. The angular selection criterion is 10 mas and the contrast is of up to one magnitude. The curves for the single channel (e_1 , e_2 , and R^2 , denoted $N(e_1)$, $N(e_2)$, and $N(R^2)$) are shown along with the combined three-channel classifier (denoted ACF), the RF, and the ANN methods. All curves are colour-coded with the F_1 score computed at each threshold. The dashed line shows the performance of a random-guess algorithm. The numerical value in the legend is the AUC of the method.

6.3. Applying ANNs to unknown PSF field experiments

The reconstruction scheme, based on the calculation of the relative errors and correlations between exposures, is subject to poor performance if there is a significant dither between exposures. The PSF can indeed spatially vary to such a degree that there is only little correlation in the PSF parameters. A more flexible approach to capture this effect is to teach the method about the deviations from the fiducial value and how it evolves with the position of the object on the chip. This latter scheme by construction handles any dithering and has the additional advantage of not having to tune an additional parameter such as ϵ in the relative error as in Eq. (10) to calibrate out the errors that are due to the PSF field reconstruction.

The method was trained on parameters observed on 25 unknown fields, each containing 280 stars, with the fiducial parameters computed by the “leave-one-out” reconstruction scheme. We give as features the deviations from the estimated parameters and the estimated parameters. The procedure of identifying the binary star is similar to the ACF or RF procedures: first, a run on all objects, with no prior on which are binaries. Then a second pass is made, with an interpolation ignoring the objects that were blacklisted as binaries in the first run. This algorithm is almost identical to algorithm 1, with the difference that the features used to classify are the interpolated fiducial parameters and the measured parameters. The second pass reduces the FPR by a few percentage points and slightly increases the true-positive rate. Two passes are sufficient to reach the best performance, while ACF and RF typically require three to five passes. In the current implementation, there are 15 input features (three deviations times four exposures plus the three interpolated fiducial values of the parameters) with a hidden three 15-neuron layers and 1 output neuron. The architecture used in this work is thus: 15 inputs, 15 hidden neurons, 15 hidden neurons, 15 hidden neurons, and 1 output. We validated the training on a set of five fields to select the separating threshold value and tested 15 fields of 280 stars.

5.1. The effect of unresolved binaries on PSF determination

T. Kuntzer and F. Courbin: Detecting unresolved binary stars in *Euclid* VIS images

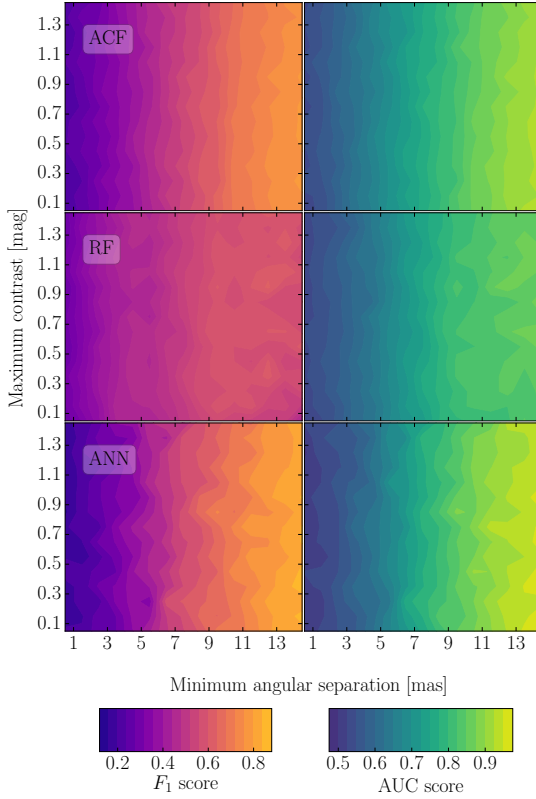


Fig. 9. Performance metrics F_1 (left) and AUC (right) for different combinations of selection criteria in the unknown PSF experiment, using the ACF method (top panel), RF (mid panel), and ANN (bottom panel).

Overall, the ANN method performs better than either ACF and RF, especially at high minimum angular separation, reaching an AUC and F_1 score of almost one, as shown in Fig. 8. The F_1 score, as shown in Fig. 9, is higher for most of the selection criteria. The main drawback of this method is the complexity of the training scheme.

6.4. Performance dependence on the assumptions

In the following, we discuss different effects or assumptions that can significantly affect the performances of the algorithms.

- The measurements of the shape parameters are assumed to be accurate. The features should reflect a systematic bias between the exposures. If there is a systematic additive bias in the shape measurement, the algorithms could be misled into classifying single stars as binaries. For the ACF classifier, the performance depends on the value of the additive bias. A slow decrease in metrics value is observed until an additive bias levels of the order of 10^{-3} . Past this threshold, the ACF classifier is essentially random. For the machine-learning classifiers, if the additive bias is included in the training, no significant degradation of the performance is noted. If the bias is not included, the same threshold in bias level is observed with the same consequences.

- The errors on the shape measurements are assumed to be Gaussian. We tested the shape measurement algorithm based on adaptive moments on noisy PSF images at the *Euclid* resolution. Although the errors are larger than the *Euclid* requirements, their distribution is Gaussian.
- If the precision on the measurement of the size is improved from 5% to 0.5–1%, the size channel carries much more weight. If the precision of measurement on the ellipticities is improved by a factor of two from the baseline of 1%, the performance increases by a few percent. A degradation of the same magnitude is observed if the precision is worsened by a factor of two.
- The number of stars per field in the *Euclid* survey will vary depending on Galactic latitude. The estimates range from 1800 objects for high Galactic latitude to twice this number at low Galactic latitudes (Cropper et al. 2013; Laureijs et al. 2011). For the ACF, both performance metrics are stable for a number of stars per field higher than 1000. For the RF, the AUC metrics increase by 5–10% between fields where 1000 stars are present and 3000 stars per field. The AUC for the ANN similarly improves in the same conditions.
- We finally discuss the completeness of the detection when the algorithms are trained with a full binary population, that is, with no selection of the binaries for training. As expected from Fig. 9, the AUC value is 0.5 and the overall F_1 score is poor. The completeness of the detection of binaries at large angular separation is also significantly reduced compared to methods trained on binaries selected for their large angular separation. The value of the completeness is roughly divided by two between a classifier trained to find unresolved binaries in an unknown PSF field at 10 mas minimum and a classifier trained to detect all binaries.

7. Conclusion

Unresolved binary stars can create significant biases on the PSF determination in space-based weak-lensing surveys like *Euclid*. As binary stars are ubiquitous in the sky, their observation cannot be avoided. The catalogues of single and binary stars provided by *Gaia* will be useful to flag a number of undesired objects, but the binary stars identified by *Gaia* are not expected to match the depth of *Euclid* or match its footprint on the sky (Eyer et al. 2015).

We here presented an approach to detect unresolved binaries using catalogues of shape parameters of PSFs observed multiple times. Repeated measurements of the same objects are provided by the dithering plan and the deep field observations.

We used relative errors of the complex ellipticity and size to their fiducial value as input features for our classification algorithms. We proposed three methods. The first (ACF) is based on the auto-correlation of the relative errors of the parameters. The two others are supervised machine-learning algorithms: random forest (RF), and artificial neural networks (ANN). The methods were tested using two numerical experiments: a best case, in which the PSF parameters are known to the *Euclid* requirements, and a worst case, in which the fiducial PSF parameters are unknown. Based on an analysis of the distortions caused by binary stars in a realistic *Euclid* setting, we suggested a detection limit for binaries separated by at least ~ 3 mas with a contrast lower than 1.5 mag to remove binaries whose distortions on the PSF ellipticity is of the order of the measurement error.

Supervised machine-learning approaches perform well in the best-case scenario, and the detection limit of 3 mas is reached. The tests on the ACF method indicate that about 50% of the

binary stars above the detection limit are correctly flagged. The influence of the contrast is weak in the detection performance, which depends mostly on the angular separation. In the worst-case experiment, the performances are degraded because the fiducial PSF shape parameters must first be determined. The RF method is limited in the worst-case scenario. ANNs performs better than the other two methods because of the increased ability of learning the relation between the deviations from the fiducial values and the binary nature of the object. The ANN approach is able to detect about 50% of the binaries at an angular separation of 5 mas minimum. We stress that in this worst-case scenario, no prior knowledge of the PSF was used at all.

The methods proposed here are given as a proof-of-concept. Dithering between exposures is left for future implementation. The treatment of dithering can be taught to a machine-learning approach without any major change in the method, as shown by our tests with the ANN. Priors based on the apparent spectral class of an object can be added to overcome the difficulty of the dependence of the PSF shape on the spectra of the observed point source. The incorporation of PSF parameters knowledge, even if not at the *Euclid* requirements, can significantly increase the effectiveness of the proposed methods.

Acknowledgements. The authors would like to thank Jérôme Amiaux, Patrick Hudelot, Koryo Okumura, and Samuel Ronayette for providing the simulated *Euclid* PSFs. We would like to thank Henk Hoekstra and Joana Frontera-Pons for reading a draft of this paper and Vivien Bonvin for useful suggestions. We thank the anonymous referee for their time and valuable comments. This research has made use of NASA's Astrophysics Data System. We are grateful to the authors of the following Python packages: Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), Matplotlib (Hunter 2007), Scipy (Jones et al. 2001) and Numpy (van der Walt et al. 2011). This work is supported by the Swiss National Science Foundation (SNSF).

References

- Abt, H. A., & Levy, S. G. 1976, *ApJS*, **30**, 273
- Antilogus, P., Astier, P., Doherty, P., Guyonnet, A., & Regnault, N. 2014, *Journal of Instrumentation*, **9**, C03048
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bernstein, G. M., & Jarvis, M. 2002, *AJ*, **123**, 583
- Bishop, C. M. 1995, *Neural Networks for Pattern Recognition* (New York, NY, USA: Oxford University Press, Inc.)
- Breiman, L. 2001, *Machine Learning*, **45**, 5
- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **431**, 3103
- Cropper, M., Pottinger, S., Niemi, S., et al. 2016, *VIS: the visible imager for Euclid*, 9904, 16
- Cummings, J. D., Geisler, D., Villanova, S., & Carraro, G. 2014, *AJ*, **148**, 27
- de Bruijne, J. 2012, *Astrophys. Space Sci.*, **341**, 31
- Deacon, N. R., Magnier, E. A., Best, W. M. J., et al. 2017, *MNRAS*
- Desai, S., Mohr, J. J., Bertin, E., Kümmel, M., & Wetzstein, M. 2016, *Astron. Comput.*, **16**, 67
- Duchêne, G., & Kraus, A. 2013, *ARA&A*, **51**, 269
- Duquenois, A., Mayor, M., & Halbwachs, J.-L. 1991, *A&AS*, **88**, 281
- Duvet, L., Gondoin, P., & Saavedra Criado, G. 2010, *Euclid Science Requirements Document*, Tech. Rep. issue 4, revision 0, ESA, dEM-SA-Dc-00001
- Eyer, L., Rimoldini, L., Holl, B., et al. 2015, in *Living Together: Planets, Host Stars and Binaries*, eds. S. M. Rucinski, G. Torres, & M. Zejda, ASP Conf. Ser., **496**, 121
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. 2004, *ACM Trans. Inf. Syst.*, **22**, 5
- Hirata, C., & Seljak, U. 2003, *MNRAS*, **343**, 459
- Hoekstra, H., Wu, Y., & Udalski, A. 2005, *ApJ*, **626**, 1070
- Hunter, J. 2007, *Comp. Sci. Eng.*, **9**, 90
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open source scientific tools for Python*, <http://www.scipy.org/>
- Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, **449**, 460
- Kleinbaum, D., & Klein, M. 2010, *Logistic Regression: A Self-Learning Text* (New York: Springer-Verlag)
- Kuntzer, T., Courbin, F., & Meylan, G. 2016a, *A&A*, **586**, A74
- Kuntzer, T., Tewes, M., & Courbin, F. 2016b, *A&A*, **591**, A54
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *Euclid Study Report* [[arXiv:1110.3193](https://arxiv.org/abs/1110.3193)]
- Marcy, G. W., & Benitz, K. J. 1989, *ApJ*, **344**, 441
- Massey, R., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **429**, 661
- Milone, A. P., Piotto, G., Bedin, L. R., et al. 2012, *ApJ*, **744**, 58
- Nardiello, D., Milone, A. P., Piotto, G., et al. 2015, *A&A*, **573**, A70
- Ngolè Mboula, F., Starck, J.-L., Ronayette, S., Okumura, K., & Amiaux, J. 2015, *A&A*, **575**, A86
- Nissen, S. 2003, Report, Department of Computer Science University of Copenhagen (DIKU), 31
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, **484**, 67
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Racca, G. D., Laureijs, R., Stagnaro, L., et al. 2016, *The Euclid mission design*, *Proc. SPIE*, **9904**, 23
- Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJSS*, **190**, 1
- Riddle, R. L., Tokovinin, A., Mason, B. D., et al. 2015, *ApJ*, **799**, 4
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, **409**, 523
- Rodriguez, D. R., Duchêne, G., Tom, H., et al. 2015, *MNRAS*, **449**, 3160
- Rowe, B., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, **10**, 121
- Schneider, P. 2006, *Gravitational Lensing: Strong, Weak and Micro*, Saas-Fee Advanced Course (Springer), **33**, 269
- Semoloni, E., Hoekstra, H., Huang, Z., et al. 2013, *MNRAS*, **432**, 2385
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, *ArXiv e-prints* [[arXiv:1503.03757](https://arxiv.org/abs/1503.03757)]
- Terziev, E., Law, N. M., Arcavi, I., et al. 2013, *ApJSS*, **206**, 18
- van der Walt, S., Colbert, S., & Varoquaux, G. 2011, *Comp. Sci. Eng.*, **13**, 22
- Voigt, L. M., Bridle, S. L., Amara, A., et al. 2012, *MNRAS*, **421**, 1385

5.2 Stellar spectral classification based on VIS images

Another block in the PSF module is the determination of the spectral classification of point-source objects. Stellar SEDs must be reliably estimated to obtain a sufficiently good wavelength-dependent reconstructed PSF (see Sect. 3.2.2 and [Cypriano et al. 2010](#)). The knowledge of the spectral type of stars is scientifically interesting beyond the demanding *Euclid* PSF reconstruction, in particular for the study of stellar populations and their formation history.

In this section, we study a novel fast and reliable approach to determine spectral type. The approach would serve to establish a first approximate catalogue to empower other algorithms that would require initial guesses or allow to screen through the data quickly. After examining the motivations for the technique, we highlight the simple physical mechanism we conceive and present proof-of-concept tests. We then showcase an application on real HST data.

5.2.1 Spectral type determination using single-band imaging

This determination could serve other purposes than spectral type determination. Small galaxies could trick a point-source finder and be listed as stars. Galaxies are extended objects, even if their profile appear to be point-like at first glance. Their profile is not the representation of the PSF. Galaxies and stars do not share the class of spectra. Trying to infer a stellar spectra for a galaxy should fail and unveil the true nature of the object.

Methods to determine the spectral type are typically based on the analysis of spectra, or on fitting SEDs to observations of multiple photometric bands, which is resource intensive. In the context of *Euclid*, the *Gaia* catalogue will be available and provide the spectral type for the brightest stars in the sample. The *Gaia* limiting magnitude for stellar spectra determination was foreseen to be $G \approx 17$ mag and was confirmed in the first data release ([de Bruijne 2012](#); [Gaia Collaboration 2016](#)). This limiting magnitude is in the VIS very bright range, so the coverage of the *Gaia* catalogue method is poor. The *Euclid* weak-lensing survey is complemented by an important effort on the ground to image and acquire low resolution spectra down to *Euclid* faint magnitude limit. Surveys like DES or LSST will provide spectral determination using classical methods ([Rhodes et al. 2017](#)). The integration and treatment of multiple-source data (*Euclid* plus ground-based data) needs to be designed to be fast and seamless. It is likely, however, that the pipeline will suffer from delays between the capture of the VIS image and the publication of the stellar spectra catalogue. We propose a method that overcomes this delay and any red tape by requiring only VIS images. A high galaxy number density is required to drive down the statistical noise in reduced shear, and in turn reach the precision and accuracy aimed by *Euclid* ([Laureijs et al. 2011](#)). There are basically two ways of achieving higher galaxy number density: (i) increase the exposure time, which implies a longer lifetime for the same survey area or (ii) a broader filter. VIS will feature a single broad filter. While colour biases are hindering cosmological parameter determination, then can also be exploited to infer the SED of point-like sources.

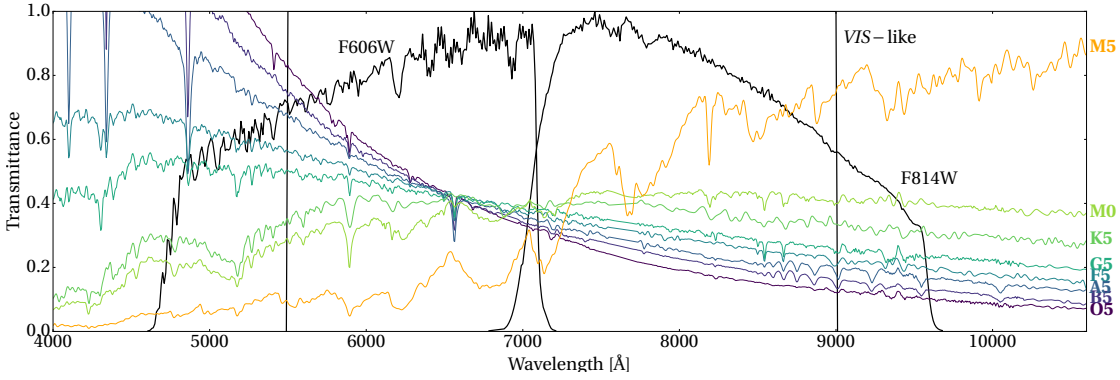


Figure 5.13 – Stellar templates for a few of the spectral types that will be used in this section along with the transmittance B of filters used in this work (*Euclid* VIS, and HST F606W; F814W). The spectra have been normalised by their total flux in the VIS band and plotted in arbitrary units of flux.

Stars are grouped by their spectral features. Those are directly linked to their effective temperature, that is the temperature of their photosphere. The temperature of a star is also linked to its mass. Stars exhibiting a low effective temperature are red, and of low mass, while massive and hot stars are blue. The Morgan-Keenan (MK) scheme classifies stars using the letters O, B, A, F, G, K, and M from hottest to coolest and subdivides each classes into ten subclasses (0 being the hottest). An additional roman number distinguishes the different stellar phases (giants, main-sequences, WDs, ...). In this scheme, the Sun, a $T_{\text{eff}} \simeq 5500$ K main-sequence star, is classified as G5V. Some spectral templates from the Pickles library (Pickles 1998) are shown for reference in Fig. 5.13. Spectral types depend on the effective temperature. In the text, we refer to half a spectral class as the gap between O5 and B0, or between G0 and G5. Note that the MK scale is based on historical consideration about colour such that the difference in temperature between two classes is not constant.

The method provides quick and reliable estimates of the stellar spectral type in VIS images. While it was developed with *Euclid* applications in mind, it is generalisable (and will be generalised in Sect. 5.2.3) to any broad filter space-borne observations. Our algorithm is based on the study of the observed stellar profile. The observed PSF representation in an astrophysical image depends on the spectrum of the object, S_p , as

$$\text{PSF} = \int_0^\infty \text{PSF}(\lambda) S_p(\lambda) B(\lambda) d\lambda, \quad (5.13)$$

with $\text{PSF}(\lambda)$, describing the instrument PSF as it evolves with wavelength and B the transmittance. This latter quantity B essentially describes the filter band. Broad filters and a space-based telescope are the two requirements for the approach. The results are better when the former criterion, the filters, are very broad. A broader band is able to capture more spectral features than a narrow band. The latter criterion implies that the images are diffraction-limited. The PSF per wavelength $\text{PSF}(\lambda)$ and the band are fixed and, we assume, not evolving

in time. For *Euclid*, these assumptions are true for period of times of $\gtrsim 45$ minutes (at the level of the requirements) and reasonable for a much longer period beyond that. If two of the three quantities in equation (5.13) are fixed, the subtle differences in the PSF profile are therefore related to the spectrum of the object. We base our approach on a semi-supervised machine-learning analysis of the PSF profile. After some pre-processing, the second step is to find an alternative representation of the PSF which is sensitive to the subtle differences in profiles between spectral types. The third is to the classification itself.

We illustrate the underlying physical process, namely a wavelength-dependent angular resolution. The angular resolution θ depends on the diameter of the telescope, D , and on the wavelength, λ ,

$$\theta = 1.22 \frac{\lambda}{D}. \quad (5.14)$$

This equation is valid for a circular aperture and for space-based observations. Light bundles that pass through the atmosphere experience turbulence and wavelength-dependent perturbations that affect the measured angular resolution. This effect disqualifies ground-based observations for our technique. As blue light has a shorter wavelength than red light, blue unresolved objects will be smaller. This is illustrated in the left panel of Fig. 5.14.

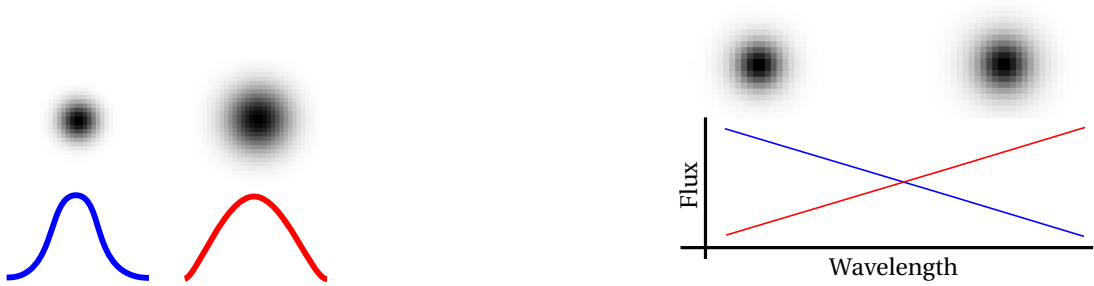


Figure 5.14 – (*Left panel.*) PSF profiles for (*left.*) a purely blue object and (*right.*) a purely red object. Due to its longer wavelength, the red PSF is broader. (*Right panel.*) PSF profiles for (*left.*) the object with the bluest spectrum and (*right.*) the object with the reddest spectrum.

Stars however are not purely blue or red, their energy distribution spans the electromagnetic spectrum. We create two mock spectra: one with a preponderant blue component, fading towards the redder part of the spectrum and an opposite one. The two PSF profiles have different SEDs. The resulting images are shown in the right panel of Fig. 5.14. The difference between the two images become more subtle, but each profile can still be clearly identified by a human eye. The stars observed by *Euclid* will not have such simple spectrum. We show three very different stars in Fig. 5.15: a very blue O5, a solar and a red dwarf.

At first glance it is hard to identify any difference between the O5 and G0 stars; only the red dwarf can be identified due to its somewhat more uniform wings. When the pairwise differences are inspected, it becomes clear that the central part of the profile and an annulus

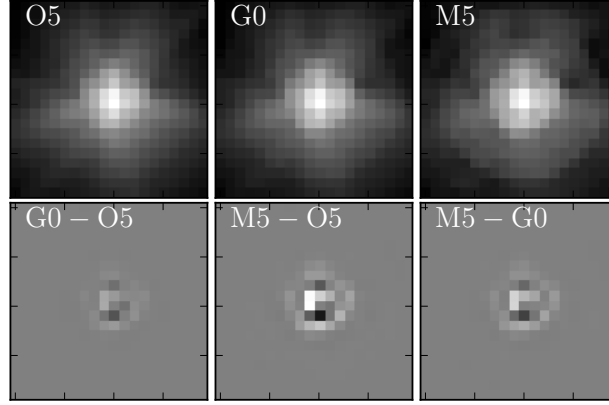


Figure 5.15 – (*Top.*) Simulated stellar images of different spectral types, as seen by the *Euclid* VIS imager, shown with a logarithmic scale. (*Bottom.*) Differences between pairs of these images, shown with a linear scale. White is positive and black is negative. The images are idealised: they do not contain any noise, and the profiles are centred at exactly the same position with respect to the pixel grid.

around concentrate the information.

Our single-band classifier exploits these subtle differences between the profiles to predict the spectral type of a star. There are three main steps to the algorithm:

1. A pre-processing of the data to extract postage stamp images, centered on the star centroid. In the first part of this work, the images are simulated observations by *Euclid* and HST. In Sect. 5.2.3, we turn to applications to real data.
2. The representation of observed pixelised PSF is changed by projecting on the principle components. This effectively reduces the dimension of the data by concentrating the information in only a few features.
3. A classification of the features by spectral class is performed by a committee of ANNs.

This technique is semi-supervised: the PCA algorithm, what we refer in the following as the encoder, does not need a ground truth, but the classifier part, the ANNs do. The training set can be either simulated or taken directly from previous observations. Simulating a training set is best as it would allow to generate a dataset with no misclassification and, most importantly, free the user from having to tediously harvest archive data and catalogues. The complexity and realism of simulated data can also be controlled. Training and testing with perturbed⁵ data can be useful to assess the robustness of the approach to e.g. the temporal variation of the PSF. Using archive data ensures that the training images are similar to the data, contrary to a simulation. Dataset size when sampling from real data could be an issue, even if data augmentation techniques are applied. Increasing the size of mock dataset requires computing

⁵E.g. less physical datasets, or with a different distribution of spectral class, ...

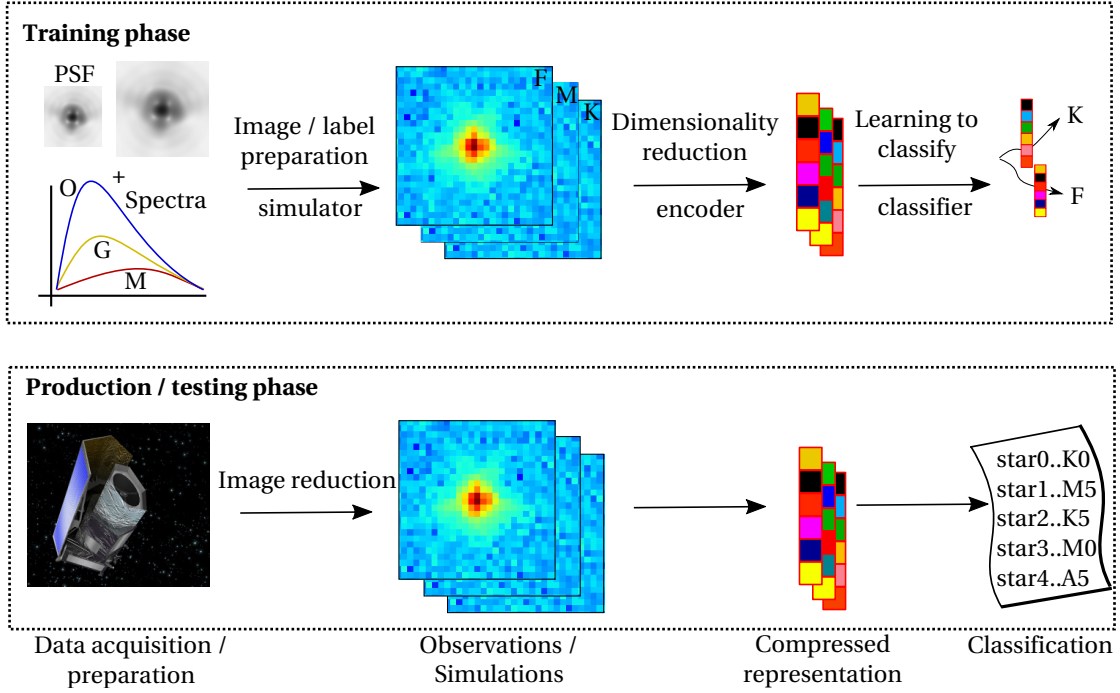


Figure 5.16 – Concept of the stellar single-band classifier. (*Top.*) Training phase of the algorithm with the simulation of PSF images with different spectra. Encoding this data with PCA yields a more compact representation that serves as a set of features. (*Bottom.*) Production/testing phase of the algorithm in which the data follows the same process as before, but the final product is not the classifier and encoding parameters, but a catalogue of spectral types. Credits for the artist view of *Euclid*: ESA and *Euclid* consortium.

resources, but can be achieved usually quickly if the simulation process is reasonable in its complexity. Figure 5.16 summarises the approach visually with a mock dataset example and *Euclid* observations.

At the second step, we exploit the PCA algorithm to find a better representation of the data. As shown in Fig. 5.15, the discriminative features concentrate in an annulus closely surrounding the peak of the PSF and the peak itself. This means that, and unsurprisingly, the slope of the profile is varying as function of the stellar type. We find that the predominant eigenstars⁶ are those describing the same annulus and peak. For *Euclid* and HST simulated data, the highest scoring architectures need about twenty-five components.

We designed the classifier part as a continuous regression task. Each of the spectral classes are assigned a numerical value from hottest to coolest star. The networks predict this numerical value. This design choice has the advantages that the output layer is only one-dimensional and there is a fast post-treatment to round to the closest class number. The disadvantage of

⁶The eigenvectors returned by PCA can be mapped back to an image, to eigenimages. Since the images represent stars, we refer to those eigenvectors as eigenstars.

this technique is that there is no second or third choice, like with a classical softmax last layer classifier. That being said, the underlying physical effect is continuous and depends on the effective temperature. We train a committee of networks. We can exclude the badly trained networks and assess the confidence of the prediction. Having a regression rather than a fixed number of class has an added advantage: it allows out-of-range predictions. This entails that a classifier could very well predict 0, while the minimum is 1.5 (which is mapping to O5). When the members predict very different classes (i.e. have a large standard deviation), the objects are likely at very low S/N or could not be stars altogether. Networks will not be reacting the same way to an out-of-range signal, since their (stochastic) training did not prepare them for such cases. This gives our method the capacity of detecting non-stellar objects in the samples. Star-(small point-like) galaxy disentanglement is important both for PSF determination and the selection of the weak-lensing analysis objects. This application remains to be explored in the case of VIS data. Optimisation of ANN in the proof-of-concept tests showed a great variability of the architecture. For example, a VIS classifier requires only two layers of 26 neurons to reach peak performance, while HST F606W demands three 29-neuron layers. The capacity of the networks must be adapted in function of the difficulty of the task.

The second and last steps require algorithms capable of feature extraction and classification. PCA and ANNs could be replaced in a future implementation by a CNN that might pick-up the necessary discriminative features in the profiles more efficiently.

The number of retained eigenstars and the architecture of the networks depend on the difficulty of finding the mapping between stellar images to their spectral types. The main drivers of the difficulty are (i) S/N and (ii) the transmittance function of the filter. There of course a lower limit in S/N, below which it becomes impossible to reliably measure the discriminative features of the different light profiles. From our proof-of-concept tests (described in the next paragraphs and in Sect. 5.2.2), we estimate that this limit is not reached for the depth of both *Euclid* and HST. Narrower bands, like the HST F606W and F814W, are sensitive to a smaller fraction of the spectrum than a VIS-like filter.

The technique must be robust to the presence of extinction. Extinction causes scattering of the incoming light bundles by dust on the line-of-sight. Due to the characteristics of interstellar dust, the shorter wavelengths are more scattered than their redder counterparts. This implies a reddening of the sources. The consequence for our classifier is direct: the object change their spectrum, thus the profile is modified too and a bias arises. Importance of extinction depends on the dust density along the line-of-sight, the distance to the emitting object and the direction of the line-of-sight. Extinction-induced biases can be mitigated if reddening is included during the training phase. The training images must simply be reddened. Failing to include the reddening in the training set will bias the classification towards redder types. In the proof-of-concept tests, reddening causes a bias of half a spectral class redder.

Another issue could arise from the spatial variability of the PSF. As the PSF changes, the discriminative features in the PSF profile can be significantly modified, leading to biased

results or failures. In the tests we performed, this seems not to be a significant effect, however if the spatial variations of the PSF were to be large, there would be a need to train several instances of the methods on different, PSF coherent, regions of the image. In essence, this marginalises over the different PSF in the image or region. Alternatively, the networks could account for the PSF by providing PSF shape parameters fields to the network. This can be achieved by, e.g. giving the image coordinates as additional inputs.

We also considered the effect of close contaminants in the stamps. Close-by object degrade the representation of the PSF, thus affecting the performance. We found that objects fainter by a flux ratio of two did not impact the performance. More work is to be done to characterise the effects of star-star and star-galaxy blends. In light of that statement, extended objects, which are usually fainter in terms of surface brightness should not impact much the classification.

5.2.2 *Paper III: Stellar classification from single-band imaging using machine learning*

We designed simple experiments to assess the performance of the method. To that end, we simulate mock images of stars by *Euclid* and HST F606W. We prepare thirteen different spectra of MS stars and at ten spatial positions with a large range of S/N, without reddening. The spectra are drawn from the Pickles library (Pickles 1998). Two sets are prepared in this way: a training set of 32 000 images and a validation set containing 20 000 samples. The validation set is used to optimise the hyperparameters of the methods. The test set is more complicated than the other sets: there are more spectra, a lower range of S/Ns, extinction and more spatial PSF variability. Real stellar spectra are more complicated and variable than the template library. Including more spectra (whose ground truth is rounded to the nearest training spectrum), increases the realism of the tests. If drawn from real data, it is likely that the samples would be at relatively high S/N (spectroscopic determination needs more flux than photometry). This motivates including lower S/N images. The advantage of training on higher S/N data, is that the sample size is smaller and thus, the training faster. We note here that the templates and PSF positions used for simulating the training set are also present in the validation and test set. This means that we are using the same models for training and testing. We defend this choice as the number of spectra and PSF is limited (23 and 600 respectively). Due to other noise realisations and the combination of spectra and PSF, the images are still different in training and test, even when the underlying models are the same. In all datasets, we prepare an equal number of stars per class. This does not reflect the Milky Way stellar population: most of the stars are classes G, K, M. This equal fraction is there to ensure that we evaluate the performance of the method in an unbiased way. Having a flat distribution of spectral type actually degrades the performances. Spectral types between G5 and M5 show the best classification metric. A test performed on realistic populations of stars would therefore yield better results. As for spectral type, we chose to generate a flat distribution in S/N. This, however, is an optimistic assumption as the number density of objects is higher at lower S/N. At this proof-of-concept stage, a high realism of the simulation is not necessary, as we aim at

demonstrating the feasibility of the technique.

More details on the image simulations, the choice of machine-learning algorithm and the test results can be found in the paper III, published in the European journal *Astronomy & Astrophysics* and reproduced here ([Kuntzer et al. 2016b](#)). The goal of the paper was to show that simple alternative approaches can contribute valuable information quickly to the weak-lensing pipeline. The results of this approach were presented at different *Euclid* conferences. A poster presented at the eighth astronomical data analysis conference⁷ in Greece (in 2016) won the best poster award.

Malte Tewes, co-author of this research paper, contributed by helping in the implementation of the software and proposed ideas that increased the realism of the simulations. In the spirit of open research, the accompanying software can be downloaded from the Lastro webpage⁸.

(See next page.)

⁷<http://cosmo21.cosmostat.org/>

⁸<http://lastro.epfl.ch/software>

Stellar classification from single-band imaging using machine learning

T. Kuntzer¹, M. Tewes², and F. Courbin¹

¹ Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
 e-mail: thibault.kuntzer@epfl.ch

² Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany

Received 7 April 2016 / Accepted 29 April 2016

ABSTRACT

Information on the spectral types of stars is of great interest in view of the exploitation of space-based imaging surveys. In this article, we investigate the classification of stars into spectral types using only the shape of their diffraction pattern in a single broad-band image. We propose a supervised machine learning approach to this endeavour, based on principal component analysis (PCA) for dimensionality reduction, followed by artificial neural networks (ANNs) estimating the spectral type. Our analysis is performed with image simulations mimicking the *Hubble* Space Telescope (HST) Advanced Camera for Surveys (ACS) in the *F606W* and *F814W* bands, as well as the *Euclid* VIS imager. We first demonstrate this classification in a simple context, assuming perfect knowledge of the point spread function (PSF) model and the possibility of accurately generating mock training data for the machine learning. We then analyse its performance in a fully data-driven situation, in which the training would be performed with a limited subset of bright stars from a survey, and an unknown PSF with spatial variations across the detector. We use simulations of main-sequence stars with flat distributions in spectral type and in signal-to-noise ratio, and classify these stars into 13 spectral subclasses, from O5 to M5. Under these conditions, the algorithm achieves a high success rate both for *Euclid* and HST images, with typical errors of half a spectral class. Although more detailed simulations would be needed to assess the performance of the algorithm on a specific survey, this shows that stellar classification from single-band images is well possible.

Key words. methods: data analysis – methods: statistical – techniques: photometric – stars: fundamental parameters

1. Introduction

Traditional methods to infer the spectral type of stars rely, as the name suggests, on the analysis of expensive spectra or multi-band photometry. Knowledge of spectral types and stellar parameters such as mass and age for large numbers of stars is of course of direct interest for stellar population studies and to study the formation history of our Galaxy (e.g. [Smiljanic et al. 2014](#); [Yang & Li 2015](#); [Ness et al. 2015](#)).

More indirectly, stellar classification is also relevant for the future space telescopes *Euclid*¹ ([Laureijs et al. 2011](#)) and WFIRST ([Spergel et al. 2015](#)), as a reliable classification improves the quality of the reconstruction of the wavelength-dependent point spread function (PSF; e.g., [Cypriano et al. 2010](#)) and as accurate knowledge of the PSF is mandatory to reach the scientific requirements for the weak gravitational lensing surveys (for *Euclid* see e.g. [Cropper et al. 2013](#); [Massey et al. 2013](#)). The VIS imaging instrument of *Euclid* will feature a single broad filter. While this is needed to reach the required number density of galaxies ([Laureijs et al. 2011](#)) to measure cosmic shear with sufficient precision, broad-band imaging also implies a number of complications in measuring galaxy shapes ([Voigt et al. 2012](#); [Semboloni et al. 2013](#)). In addition, aside from the chromatic dependence of the PSF, a notable indirect effect arises from the spatially variable abundance of stars with companions ([Kuntzer et al. 2016](#)). Stellar data from *Euclid* can provide a wealth of information and contribute to a possible

extension of the ESA *Gaia* catalogue as *Gaia* will provide stellar spectra for stars down to magnitude 17 ([de Bruijne et al. 2015](#)).

In this paper, we present a novel technique to estimate the stellar spectral type of spatially unresolved sources, based solely on their image shape in a single wide band. This is important to carry out a first classification on the optical data of *Euclid* quickly and even for faint stars, beyond the reach of *Gaia* or with no multi-band photometry available. Our technique will also be useful to classify stars in archival images of the *Hubble* Space Telescope (HST). These images were taken in only one filter and therefore function as a general-purpose tool for stellar work.

The method exploits the subtle differences in diffraction limited images of point sources with contrasting spectra. A broad filter is generally advantageous for this approach, as it accentuates these differences between sources with varying spectral slopes. For this first approach, we perform the classification of sources into spectral types through a regression of a continuous scalar parameter, C_s , that roughly represents an effective temperature and covers adjacent bins of different spectral types. For each source, estimates for C_s are predicted by artificial neural networks (ANN, see, e.g., [Bishop 1995](#)), using coefficients from a Principal Component Analysis (PCA, [Pearson 1901](#)) of the source image as input. These neural networks perform a supervised machine learning, via training on stars with known spectral types.

All the images used in our exploratory work are simulations of stars along the main sequence, as observed either with *Euclid* or the HST. This allows for a controlled proof of concept. But

¹ <http://www.euclid-ec.org/>

importantly, using these simulations, we also demonstrate the proposed technique in a purely data-driven application. For this, we mimic a situation in which a training set, with known true spectral types, is obtained by high resolution spectroscopy. To emulate an incomplete sampling of the training stars, we set aside some of the stellar spectra and spatial locations within the focal plane of the instrument during the training phase. We then analyse the performance of the method on stars with a lower signal-to-noise (S/N) cut, a greater variety of spectral types, and suffering from reddening by extinction. This complex test probes the interpolation behaviour of the classifier, and gives a first assessment of the reliability of results that could be expected on real data.

This article is organised as follows: we detail the algorithm and associated performance metrics in Sect. 2. We then describe the preparation of the different simulated data sets for training and testing in Sect. 3. In Sect. 4, we discuss the optimisation of the hyper-parameters. A proof-of-concept classifier and the performances of the classifiers for both *Euclid* and HST are detailed in Sect. 5. Finally, Sect. 6 summarises the work.

2. Scheme and algorithms

The proposed method, which we refer to as single-band classification, takes advantage of the fact that the diffraction-limited PSF of a telescope varies with wavelength. The precise shape of a stellar image, integrated over an observing filter, is therefore dependent on the transmission profile of the filter, as well as the stellar spectrum within this profile (for an illustration, see Figs. 1 and 2). Our single-band classifier exploits these shape differences to predict the spectral class of a star. In the following, we succinctly lay down the different steps of the classifier, before describing them in more detail.

1. Pre-processing of the data
To analyse images from a space-based survey, a catalogue is first created. This is performed through the detection of all stars, or, more generally, unresolved objects. Square stamps centered on the objects are prepared and normalised. Note that in this work, we simulate all the data, and directly produce stamps of pure stellar nature.
2. Dimensionality reduction
Instead of using the normalised pixel values of a stamp as input to the machine learning, the image information is compressed, in order to reduce the dimensionality of the problem. To do so, the stamp images are projected onto a common basis, and only the most significant components are retained. In the vocabulary of machine learning, this reduces each stellar image to a chosen number of “features”.
3. Classification
The goal of this step is to create a robust mapping from the features to the spectral class of each object, using supervised machine learning. As commonly done in machine learning, we use an ensemble (“committee”) of classifiers and compare their outputs to (1) increase the confidence in the results; (2) estimate the uncertainty of the classification; and (3) detect unclassifiable objects.

2.1. Dimensionality reduction

As the images of stars with different spectra do undeniably share common structures, they can be reconstructed, up to their noise, using a combination of components that are defined on a basis highlighting the differences between these images. Finding

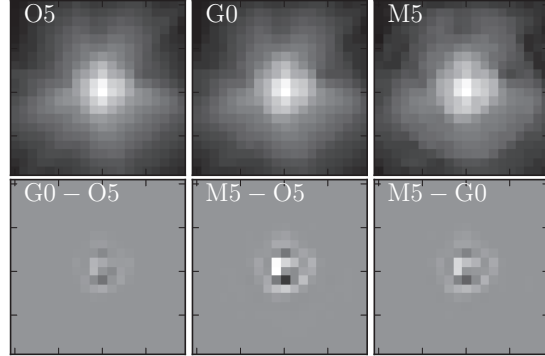


Fig. 1. *Top:* simulated stellar images of different spectral types, as seen by the *Euclid* VIS imager, shown with a logarithmic flux scale. *Bottom:* differences between pairs of these images, shown with a linear flux scale. White is positive and black is negative. Note that for demonstration purposes, this illustration is highly idealised: the above stellar images do not contain any noise, and the profiles are centred at exactly the same position with respect to the pixel grid.

this basis and retaining only a number of elements that represent the data well enough is the aim of dimensionality reduction. To this effect, we use the principal component analysis (PCA) technique. This algorithm projects the data onto the most meaningful basis (see, e.g., Shlens 2014) that represents the input data. A useful feature of the PCA decomposition is that it naturally provides a mean to compare the importance of each dimension. Since the projection is made along axes of decreasing importance for the reconstruction of the original data, all dimensions of order greater than a n_{PCA} cut-off threshold can be dismissed. PCA has the advantage of being non-parametric, so that no hyper-parameters must be fine-tuned other than the number n_{PCA} of components to be retained. PCA is widely used in astronomy, for example in PSF reconstruction (e.g. Jarvis & Jain 2004; Gentile et al. 2013) and in weak lensing catalogue post-processing (e.g. Niemi et al. 2015), to study properties of objects.

In practice, we simply use all available stellar images to construct the PCA basis onto which each star can be projected. In our analysis we compare results obtained by retaining from 12 to 27 PCA-coefficients for each star. We use the implementation of PCA provided by *scikit-learn* (Pedregosa et al. 2011).

Note that as an alternative to PCA, we have tried to feed a moments-based width-measurement of the light profile as well as fluxes in different apertures as input features to the classification step. However these simple attempts turned out to be less successful than the PCA reduction. Other dimensionality reduction techniques, such as independent analysis component (ICA) or manifold mapping (e.g. NMF, Ivezić et al. 2014), can also be applied to this problem, but they are not retained here as early attempts hinted at their similar or worse performance for the problem of classifying single-band stellar images.

2.2. Classification: the machine learning

At this stage, through the dimensionality reduction, each stellar image can be seen as a point in an n_{PCA} -dimensional feature space. Classification methods such as k-nearest neighbour (k-NN) or support vector machine (SVM) rely on the clustering of the data into groups with the same labels, that is, the

T. Kuntzer et al.: Stellar classification from single-band imaging using machine learning

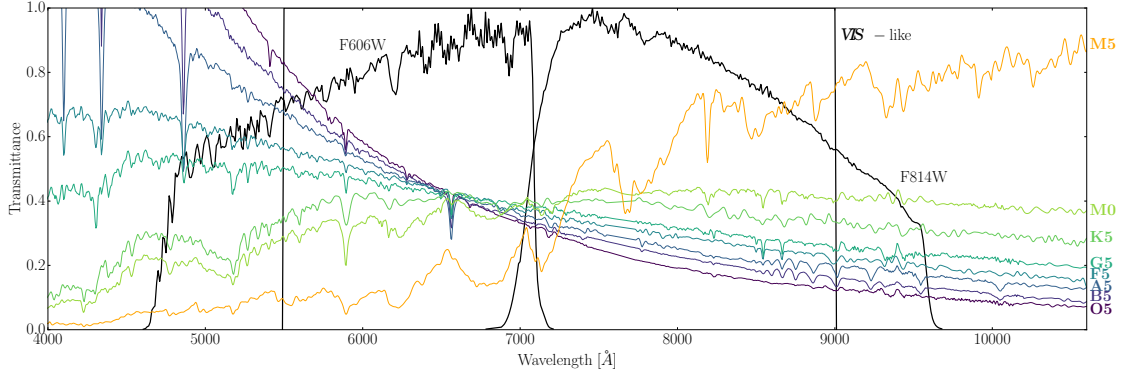


Fig. 2. Filter profiles of the three bands used in this work (VIS from 550 to 900 nm, *F606W* and *F814W*) along with some stellar spectra of different types from Pickles (1998). For display purposes, the spectra have been normalised by their total flux in the VIS band and plotted in arbitrary units of flux. Note that to simulate stellar images, we must integrate over the wavelength-dependent PSF models, using spectra in units of photon number counts.

same spectral type (see Ivezić et al. 2014, for an overview). Due to the image noise and the imperfect centering of stars with respect to the pixel grid, the different spectral types do not form clear disjoint clusters in PCA space, but exhibit a noisy but continuous evolution of the features. This will be illustrated, in the projection of two PCA components, in Fig. 6. The distribution of labelled data suggests a regression of a continuous scalar parameter, C_s , whose value evolves along the spectral classes. Eventually, the predicted class of a star is determined via a binning of C_s .

2.2.1. Artificial neural networks

We propose the use of simple artificial neural networks (ANN) to perform this regression from feature space to C_s . Feed-forward ANNs of perceptrons (Bishop 1995) consist of several nodes, each taking an input vector \mathbf{x} and returning a scalar output $h(\mathbf{x}, \mathbf{w}, b)$ via the equation

$$h(\mathbf{x}, \mathbf{w}, b) = h\left(\sum_{i=1}^N w_i x_i + b\right), \quad (1)$$

where \mathbf{w} and b are the weights and the bias, respectively. The monotonic and continuous function $h(x)$ is the so-called activation function. For our application, we use the sigmoid activation function $h(x) = 1/(1 + e^{-x})$, except for the last node of the network, which uses the identity $h(x) = x$. The nodes in the ANN are arranged into one or more layers. In each layer, nodes treat the input data through Eq. (1) with different values for the weights and the bias. In general, this input \mathbf{x} of each node consists of the outputs of the nodes in the previous layer. Nodes of the first layer take the vector of features as input, and the single node of the last layer returns the estimate for C_s . The capacity of a neural network to represent intricate dependencies depends on the number of nodes, and how these nodes are distributed into different layers. Choosing the number of nodes per layer and the number of layers is not straightforward, and we explore different combinations of number of layers and number of nodes per layer. Layers that are not the input layer nor the output are called hidden layers.

For a given and fixed network structure, training of the ANN aims at finding optimal values of the weights and biases of each

node, in order to minimise a cost function between the estimated and known true C_s values of a training set where C_s encodes the true spectral type (see Sect. 2.2.3). We use the typical least-square cost function to evaluate the goodness of fit.

Various implementations of the multilayer perceptron could be used for the purpose of this study. We use the Fast Artificial Neural Network Library (FANN) by Nissen (2003). We have also tried the SkyNet implementation (Graff et al. 2014), yielding very similar results. As we do not aim to compare implementations of ANNs in the scope of this paper, we only report results obtained with FANN in the following sections. Other algorithms such as random forests (RF) can be applied here. Simple tests carried out with RF instead of ANNs yielded similar performance.

2.2.2. Committees for better robustness and anomaly detection

Due to the complexity of an ANN training, and random initialisation of weights and biases, the final values of the parameters obtained through the minimisation of the cost function are not deterministic. A training attempt can also remain trapped in a poor local minimum of the cost function.

To address these difficulties, and increase the prediction accuracy, several independent ANNs, forming a so-called committee, can be trained individually (Bishop 1995). This allows us to reject the worst training failures, based on the cost function performance achieved on the training set, and retain only the n_c best committee members. When analysing unknown data, the different predictions from these retained committee members can be averaged, to yield a robust combined estimate for each object. A large variance of predictions is an indication that the unknown object was not represented in the training data. Another possible response to such an anomalous object would be an ensemble of predictions that fall far from the range of known values of C_s . The committee approach increases the confidence in detecting anomalies (Nguyen et al. 2015). In the present context, such anomalies could range from slightly resolved objects such as small galaxies, to unresolved objects with unusual spectra (binary stars, quasars) or too-noisy data. In the following section, we define how exactly these outliers are identified.

2.2.3. Classification into spectral types and anomalies

In this paper we consider the classification into a set of 13 separate classes of stellar spectra, with a discretisation of “half” a spectral type: {O5, B0, B5, A0, A5, F0, F5, G0, G5, K0, K5, M0, M5}. We define the continuous parameter C_s by attributing a sequence of numerical values to these classes, in steps of 0.5. Training stars of type O5 get a true C_s of 1.5, and $C_s(B0) = 2.0$, $C_s(B5) = 2.5$, ..., $C_s(M5) = 7.5$.

For each unknown object to be analysed, the combined average C_s -estimates from the retained well-trained committee members determines the classification: O5 if $1.25 < \langle C_s \rangle \leq 1.75$, B0 if $1.75 < \langle C_s \rangle \leq 2.25$, and so on until M5 with $7.25 < \langle C_s \rangle \leq 7.75$. We refer to an estimation error of 0.5 on the C_s -scale as an error of half a spectral type.

In addition, if the variance of the individual C_s -estimates is larger than 1.0 or if $\langle C_s \rangle$ is out of range, we classify the object as an anomaly.

2.3. Metrics to quantify the classification performance

To analyse the performance of the single-band classifier applied to a large sample of objects, we introduce a set of simple metrics. We describe them below.

- The *confusion matrix*, whose elements M_{ij} correspond to the relative abundance of the estimated spectral type i given the true spectral type j . Correctly classified objects contribute to the diagonal terms of the matrix, while classification errors are represented by the off-diagonal elements. The distribution of the objects in the confusion matrix can reveal systematic biases and give a detailed overview of the classification errors.
- The F_1 -score is a metric which summarises further the performance to one scalar value. For a binary classification, the F_1 -score is defined by

$$F_1 = \frac{2TP}{2TP + FN + FP}, \quad (2)$$

where TP, FN, and FP are the numbers of true positive, false negative, and false positive classifications, respectively. We compute F_1 individually for each of the spectral types, and average these results to get a single F_1 -score describing the overall classification performance. An error-free classification corresponds to $F_1 = 1$, and imperfect classifications reach lower scores. Note that this is a very strict measure of performance, as it will consider an object to be wrongly classified if the estimate falls into a class immediately adjacent to the true spectral type. In other words, given the spectral classes used in this work, it even penalises errors corresponding to only half a spectral type (e.g., G5 instead of G0).

- The *success rate* S is the classification accuracy including a tolerance of one class (i.e., half a spectral type). In practice, S is the trace of the confusion matrix plus the sum of the elements directly above and below the main diagonal, divided by the overall number of classified objects. In this paper, we optimise the configuration of the single-band classifier according to this success rate S .

3. Simulated data

In this section, we describe the preparation of synthetic data sets mimicking stellar images obtained by the HST and *Euclid*. We

first present the structure and methodology that we use for creating the mock images, and then discuss the telescope-specific tools to produce realistic images.

3.1. Training, validation, and testing

In line with machine learning practices (e.g. [Hastie et al. 2009](#)), for each observational setup to be simulated, we generate a group of three disjoint data sets, all with known true spectral type. A similar structure could be adopted to split the subset of data with known spectral classification when working with real observations.

- First, a training set is needed, on which the neural networks learn by adjusting their weights and biases. Potentially, over-fitting of the neural network parameters could lead to exceedingly high apparent performances on this training set. Over-fitting arises when the dimensionality reduction or/and the neural networks become too specific to the data, for example, by fitting the noise contained in the training set.
- The validation set is not seen by the neural networks during the optimisation of their parameters. By comparing the classification performance on the training set and the validation set, over-fitting of the neural networks can be detected. If no over-fitting is detected, and if this validation set is large enough, it can in turn be used to optimise the hyperparameters of the machine learning algorithm, such as, in the case of this work, the number n_{pca} of PCA coefficients and the size of neural networks.
- Finally, a test set is prepared, to independently test the performance of the optimised algorithm.

In the context of this paper, for some analyses we add additional astrophysical and observational complexity to the test set. Compared to the training and validation sets, we include fainter stars, more variants of the PSF corresponding to different spatial positions on the detector, additional stellar spectra, and wavelength-dependent extinction by dust. Thereby, our test sets can also be used to explore the performance of the classifier on significantly more complex data, mimicking a purely data-driven approach in which the training could not be performed on fully representative samples.

3.2. Mock stellar images: generalities

We restrict the range of stellar spectra to main sequence spectra using the templates prepared by [Pickles \(1998\)](#). A few of these are shown in Fig. 2. For all our data sets, we adopt flat uniform distributions of these spectral types and of the S/N. Inevitably, the global performance of the single-band classifier depends on the stellar distribution, as the different stellar type yield different performances. For real data, the stellar distribution would depend on the galactic coordinates ([Chabrier 2003](#); [Robin et al. 2003](#)). Our choice of a flat distribution has the advantage that a sufficient number of stars can be drawn in each stellar type bin while maintaining a tractable total size of the data sets. Tests on flat distributions could be later weighted to predict results for arbitrary stellar distributions. The same arguments motivate our choice of working with flat S/N distributions.

Table 1 summarises the characteristics of the three data sets, which we generate for each considered band and telescope. For the training and the validation sets, we restrict the diversity of PSFs to 10 different spatial locations on the detector, and use only the 13 different spectra (two per spectral class with an

5.2. Stellar spectral classification based on VIS images

T. Kuntzer et al.: Stellar classification from single-band imaging using machine learning

Table 1. Summary of the characteristics of the three data set families.

Data set	# Spectra	# PSF	A_v	S/N_{Euclid}	$S/N_{HST F606W}$	$S/N_{HST F814W}$	# Stars
Training	13	10	0	50–400	120–1000	200–1000	~32 000
Validation	13	10	0	50–400	120–1000	200–1000	~20 000
Test	27	600	0.3	20–400	80–400	150–400	~20 000

Notes. For each data set we give the number of different spectral templates, the number of different spatial positions on the detector, the maximum extinction A_v (in magnitude), the considered S/N ranges, and the number of simulated stars. The extinction in our simulated data is randomly drawn between $A_{v,min} = 0$ and A_v (see text).

exception as we start from O5) defining the classification. For the more complex test sets, we finely sample all detector positions, and use all spectral templates from the library provided by Pickles (1998) (roughly four per spectral class). For the purpose of evaluating the performance metrics, the true spectral types of these templates are rounded to the nearest classification bin (e.g., M4 becomes M5). Finally, we also add the effect of reddening by dust to the test sets only, using a Milky Way extinction curve with $R_V = 3.1$ and the extinction A_v randomly chosen between 0 and 0.3, to reflect the typical visual extinctions for the sky of the *Euclid* weak lensing survey (Cardelli et al. 1989; Schlegel et al. 1998; Schlafly & Finkbeiner 2011).

Instrument-specific codes then produce the image of the objects, according to the different bands, spectra, and fluxes. In all our simulations, objects are randomly mis-centered by up to half a pixel in each direction both on the x and y axes, to obtain a uniform coverage of the sub-pixel positions and simulate a non-interpolating stamp extraction from survey data.

3.3. Simulated *Euclid* images

The PSFs we use for the *Euclid* telescope (Laureijs et al. 2011) are simulated using the pipeline for the VIS instrument (P. Hudelot, private comm.) and consists of 600 PSFs at random spatial positions within the four central CCD chips of the VIS camera. Depending on the position on the detector, the measured axis ratio evolves from 1 to 1.15. Each PSF is a FITS datacube containing 100 wavelength slices, hence allowing us to accurately describe realistic SEDs. To produce stellar images for VIS we consider a top-hat window function between 550 and 900 nm. The pixel size is that of the VIS detector (no sub-sampling), that is $\Delta x = 0.1''$.

The S/N range for the *Euclid* training and validation sets spans $50 < S/N < 400$, while the test set images have a lower S/N cut of $S/N = 20$. The limiting AB magnitude for *Euclid* is $V \approx 24.5$, which corresponds to $S/N \approx 10$ (Laureijs et al. 2011).

The number of training stars is of the order of 32 000. The validation and test sets contain about 20 000 images. We observe that each of these samples is large enough to exclude any over-fitting when using machine learning methods.

3.4. Simulating Hubble space telescope images

For the HST simulations, we simulate stellar images in the *F606W* and *F814W* bands of the Advanced Camera for Surveys (ACS, Ford et al. 1996; Sirianni et al. 2005). Both bands have similar widths, but are centred on different wavelengths (see Fig. 2). The HST bandwidths are both about 1.5 times smaller than the *Euclid* VIS band. In addition, using the actual throughput curves, instead of an idealised top-hat function, will also reduce the potential performance of the single-band classification. The images are produced via the TinyTim

software (Krist et al. 2011) in its ACS configuration (both CCDs are used), using the same template spectra from Pickles (1998) as for the *Euclid* simulations. The lower bound for the S/N range, $S/N = 80$, corresponds to a limiting AB magnitude of Johnson $V \approx 23.5$ for O5V stars and $V \approx 24.1$ for M5V stars with an exposure time of one hour (Avila et al. 2016). For $S/N = 1000$, (the higher bound of the training set), the corresponding limiting magnitudes are $V \approx 19.5$ and $V \approx 20$ for O5V and M5V stars respectively.

Our aim in simulating these two *F606W* and *F814W* bands is not to compare their performance as input to a single-band classifier. Any such comparison would only be possible given a particular scientific question, and for a particular stellar population. Instead, we adjust here the arbitrary S/N ranges so that our classifiers yield results of roughly similar quality from both bands. This demonstrates that the single-band classification is possible both with *F606W* and *F814W* images.

4. Optimisation of the hyper-parameters

The performances of machine learning techniques such as neural networks depend on a number of hyper-parameters, for which successful values can be difficult to guess a priori. We now describe how we evaluate a grid of possible settings for the hyper-parameters of the classifier, in order to determine optimal configurations. We perform these optimisations only for the *Euclid* and HST *F606W* cases. For the *F814W* filter, we use the same optimised configuration as for the *F606W* filter. The hyper-parameters considered here are: the number of retained PCA components n_{PCA} , the number of hidden layers of the ANN, n_l , and the number of nodes per hidden layer n_{nn} . The capacity of a neural network to learn a task is determined by the values of n_{pca} , n_l and n_{nn} . Large values of the parameters are difficult to train and are prone to over-fitting (Bengio 2009). Small values of the parameters usually result in a somewhat faster training than for large value, but poorer performance, because of under-fitting.

We study the following possible values, whose ranges are determined empirically from preliminary trials:

$$n_{pca} \in \{12, 15, 18, 21, 24, 27, 30, 33\}, \quad (3)$$

$$n_l \in \{2, 3\}, \quad (4)$$

$$n_{nn} \in \{5, 8, 11, 14, 17, 20, 23, 26, 29\}. \quad (5)$$

For each resulting combination of hyper-parameters, we train 96 ANNs and retain only the n_c best ANNs. The number $n_c \in \{24, 48, 72, 96\}$ is selected to yield the highest F_1 -score on the validation set. The use of the separate validation set instead of the training set penalises potential over-fitting, although no over-fitting is detected in the present application. In the context of this paper, we do not systematically explore further hyper-parameters for each setup. In particular, the size of the image stamps on which the PCA is performed is kept constant (40 pixels on-a-side).

Table 2. Optimal configurations of the single-band classifiers, yielding best results on the validation sets.

Observational setup	n_{PCA}	n_{hn}	n_l	n_c	$F_{1,\text{va}}$	$F_{1,\text{tt}}$	S_{va}	S_{tt}
<i>Euclid</i>	24	26	2	48	0.75	0.42	0.98	0.90
HST <i>F606W</i>	27	29	3	24	0.57	0.30	0.94	0.68

Notes. The hyper-parameter n_{PCA} is the number of retained PCA components, n_{hn} is the number nodes in the hidden layers of the ANN, n_l is the number of hidden layers, and n_c is the number of ANNs retained in the committee (out of the 96 trained). The F_1 score and the success rate S are given for the validation (va) and test (tt) sets. If the output catalogues were randomly drawn, the metrics would be $F_1 \approx 0.07$ and $S \approx 0.21$.

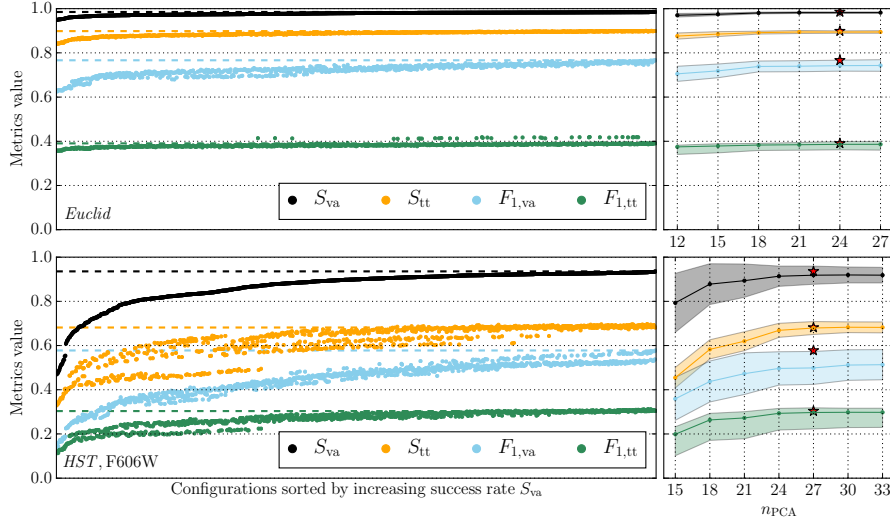


Fig. 3. Classification performances achieved by the different hyper-parameter combinations. The *top panels* are for the *Euclid* data sets, while the *bottom panels* are for the HST *F606W* filter. The *left-hand plots* show the performance of the all tested configurations in terms of the F_1 -score and of the success rate S , for the validation (va) and test (tt) sets. The dashed lines show the performance of the best configuration in both cases, selected by the highest S_{va} score. The *right-hand plots* depict the median metrics values for configurations with a given number of PCA components. The shaded regions depict the 1σ envelope on the median. The red stars correspond to the optimal classifiers.

Table 2 presents optimal settings, meaning with the best success rate, S , on the validation sets. We stress that the given metrics reflect the performance given the artificial flat distributions of spectral type and S/N, as described in Sect. 3 and Table 1.

In Fig. 3, we show the range of performance metrics achieved by the different combinations of hyper-parameters, that is the configuration. The plateaux in the left panels of the figure suggest that the choice of the configuration does not influence much the results and that poor performance of some configurations can easily be identified using the validation set. The same holds true for the number of PCA coefficients used to describe the stellar images. The relatively broad plateaux in the right-hand panels of Fig. 3 indicate that this parameter, n_{PCA} , has only a minor impact on the metrics values. Thanks to this behaviour, a crude optimisation of the hyper-parameters is sufficient.

4.1. On the significance of the PCA components

The PCA as described in Sect. 2.1 is performed on a large ensemble of stars, mixing widely different spatial locations on the detector, different sub-pixel stellar positions, and different spectral classes. We illustrate the first 20 eigen-stars from this PCA, for the *Euclid* case, in Fig. 4.

Instead of selecting the *first* n_{PCA} components as features for the machine learning, one could pick those components that are

the most “significant” for the purpose of spectral classification. For each PCA component, we quantify this specific significance by evaluating how sensitive the coefficient is to the spectral class when the nuisance parameters (spatial PSF variability, sub-pixel position, noise) are averaged over. To do so, using the same sample of stars on which the PCA was performed, we first compute the median of the eigenvalues for each component and for each true spectral class. For each component, we then compute the standard deviation across these median coefficients from the different spectral classes. The larger this standard deviation, the stronger a PCA component reacts to the morphological differences resulting from the different spectra. This is illustrated in Fig. 5 for *Euclid*, highlighting the high value of the PCA component number eight in this particular case. We find that selecting the nine most significant coefficients as input features for the network allows us to achieve 90% of the performance obtained when we use the full 24 coefficients. Even when we use only the single most significant PCA component, the classifier does not lead to catastrophic failures. Adding information from other significant coefficients of course improves the performance.

Considering Figs. 4 and 5, one can observe that the most significant PCA components usually represent the outer part of the profile, while the first eight coefficients account mainly for the central parts and the centering. Using components eight and six is, for example, an efficient way to measure the slope of the

T. Kuntzer et al.: Stellar classification from single-band imaging using machine learning

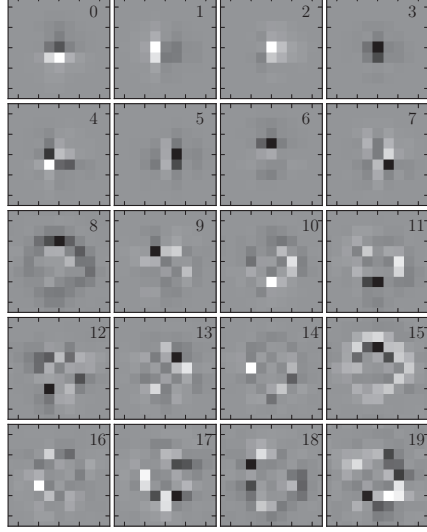


Fig. 4. Eigen-stars for the *Euclid* PCA decomposition for the first 20 components (10×10 central pixels). White is positive and black is negative. The first eight components deal with the center of the image while the others describe the wing of the profile.

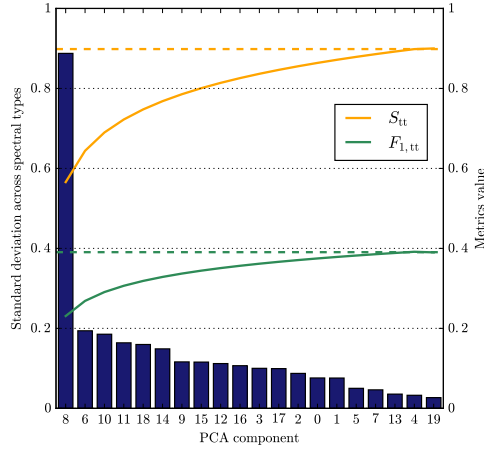


Fig. 5. Performance of the classifier with the most significant PCA components and significance of the PCA components. The bar chart shows the standard deviation in the distribution of the first 20 PCA components across the spectral classes in decreasing importance in the context of *Euclid* simulations. The orange and green lines represent the performance metrics S and F_1 for classifiers that use only the most “significant” PCA components, as defined in the text, and given leftwards in the bar chart. The dashed lines depict the results of the optimisation.

profile. Values of these two components are shown in Fig. 6, which illustrates a strong correlation between these coefficients (position of the points in the plot) and the spectral type (color of the datapoints).

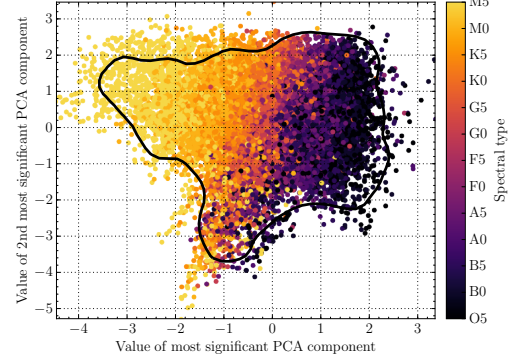


Fig. 6. PCA decomposition of the test set data. The black line depicts the envelope of the PCA decomposition of the training data set. The x -axis corresponds to the component with the largest standard deviation (see text) in coefficients across the spectral types. The component with the second largest standard deviation is shown on the y -axis.

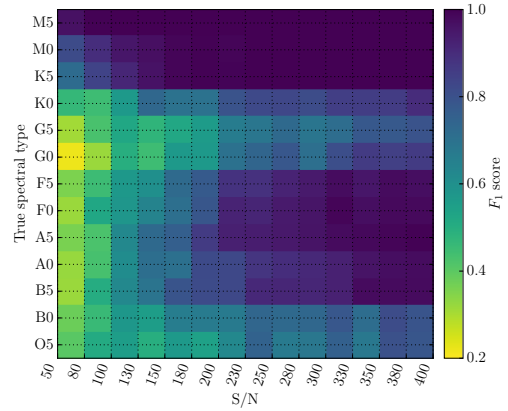


Fig. 7. Classification performance as measured by the F_1 score, as a function of the true spectral class and the signal-to-noise ratio, for *Euclid*.

We note, however, that the results presented in the next section, use the optimised value n_{PCA} for the number of PCA coefficients to ensure the maximum performance.

5. Results

This section presents the performance of the single-band classifier in different conditions, using the optimal configuration as summarised in Table 2.

5.1. Classification results: simple proof-of-concept situation

We first present results obtained from a simple and well-controlled toy model. We use the wavelength-dependent PSF at a single spatial position of the detector to simulate all stellar images, corresponding to a spatially invariant PSF. Furthermore, we use the same S/N ranges and the same stellar spectra for

A&A 591, A54 (2016)

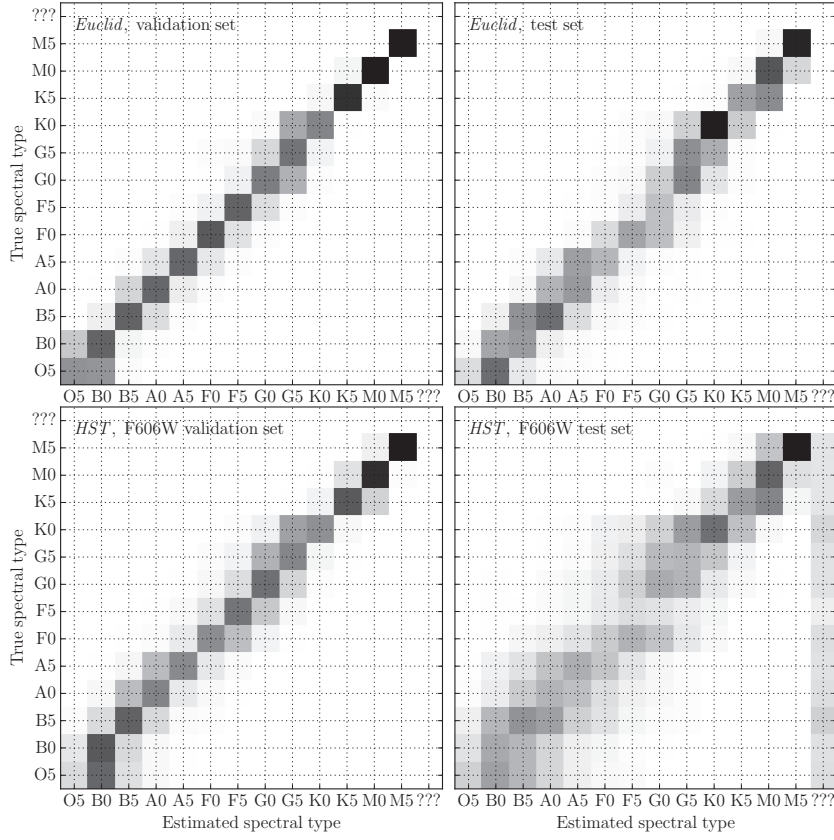


Fig. 8. Confusion matrices (see Sect. 2.3) for *Euclid* (top) and for the HST *F606W* filter (bottom). The left-hand panels correspond to the training set, while the right-hand panels show results from the more complex test sets. The label “???” denotes the “anomaly” class (see Sect. 2.2.3).

training and testing, and we do not include any extinction effects in the test set.

This simplified situation results in the best possible performance for the problem at hand. For *Euclid*, and using a uniform distribution of S/N between 50 and 400, we obtain $F_1 = 0.78$, and a success rate $S = 0.99$. The value of S is significantly closer to one than the F_1 score as S includes a tolerance of half a spectral class, as compared to the F_1 score. This indicates that the vast majority of the classification failures correspond to errors of only half a spectral type.

Figure 7 shows the F_1 score as a function of the true stellar class and of the S/N. The spectral types G0, G5 and K0 present poorer results than their neighbouring spectral classes, reflecting similarity in their spectra. The S/N barely impacts the performance for the reddest objects, but appears more important for bluer objects.

5.2. Classification results: realistic PSF field

We now move to the situation of a spatially variable PSF, and we analyse the classification performance on the validation and test sets as described in Sect. 3. The analysis of the test sets mimics a data-driven approach, in which the training would be performed

on a set of spectroscopically-classified stars with imaging data of higher S/N than for the stars to be classified (the test set).

Figure 8 shows the confusion matrices for the *Euclid* and the HST *F606W* validation and test sets. We could also consider the HST *F814W* filter but since the spectral slopes of the stars in this filter are similar we expect lower performances. In the following we will only explore the behaviour of our classifier using the bluer *F606W* filter. For both observational setups, most of the stars are distributed along the diagonal, with a noticeable excess of prediction errors concentrated in the G0 to K0 region, as previously observed for the simpler test described in Sect. 5.1. Also, we observe again that the classification of blue stars (e.g., O5 and B0) is less successful than for the reddest stars.

Figure 8 shows the degradation of performance between the training and the test phases. The difference can be explained by the inclusion of low S/N images in the test set, as described in Table 1. This degradation, while severe in the *F606W* filter, still allows for a useful classification of the spectral types, with a typical error of one spectral class. For these *F606W* simulations, a significant number of stars are classified as anomalies, denoted by the class “???” in the figures. In the present case where the test set contains only stars, anomalies are objects that are actually stars but that are classified as not being in the range of objects

5.2. Stellar spectral classification based on VIS images

T. Kuntzer et al.: Stellar classification from single-band imaging using machine learning

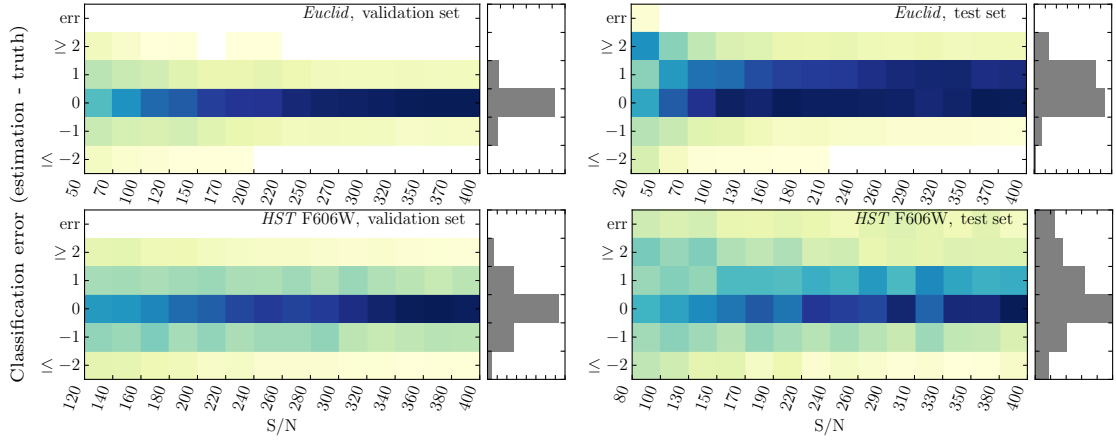


Fig. 9. Classification error as a function of the S/N for *Euclid* (top) and HST F606W (bottom). Results from the validation set are shown in the left-hand panels, and results for the test sets are displayed in the right-hand panels. These panels show the same data as Fig. 8.

known by the classifier. The stellar images wrongly classified as anomalies are low S/N objects.

5.3. Effect of reddening and extinction

Interstellar dust reddens stellar spectra. In our simulations, we have deliberately included such reddening to the test stars, but not to the training and validation stars. This mimics a situation where the training set is simulated from templates but where the test set has unknown reddening.

Figure 9 presents the same results as shown in Fig. 8, but projected on different axes, namely classification error and S/N. A classification error of +1 corresponds to classifying an object as redder than it really is. It becomes apparent that the reddening of the test set results in a bias in the predicted classes, for all S/Ns. However, this can be overcome by including a randomly distributed reddening in the training set. We carried out such an experiment and noticed that the bias disappeared. The performances increased almost at the same level as when the classifiers were run on data sets without any extinction.

5.4. The effect of contamination by companion objects

Objects angularly close to the stars degrade the quality of the PCA decomposition and consequently affect the performance of the classifier.

In order to test this, we created an additional test set for the *Euclid* case, containing only double stars (here we do not care if the stars are physically related or not). The contaminating stars are randomly placed in the considered image stamps, with a minimum distance of 1.5 pixels from the main star and they have a random spectral type. The separation of 1.5 pixels corresponds roughly to the FWHM of the *Euclid* PSF. We only simulate contaminants that are fainter than their host stars. Using the new test set but the original training set with single stars, we observe that:

- The metric S increases with the distance between the main star and its contaminant.
- Faint contaminants, that is stellar pairs with a flux ratio of larger than two, have little impact on the classification performance.

- The presence of contaminants increases the fraction of low S/N stars being classified as anomalies.

We conclude from this simple study that the general functionality of the single-band classifier is not critically endangered by the astrophysical reality of close companion objects. The presence of companion objects in the training set may, however, severely degrade the performance of the classifier. Depending on the training set selection strategy, the importance of the purity of the training set should be investigated.

6. Summary and conclusion

In this paper, we demonstrate the feasibility of inferring the spectral classes of stars from images taken with a space telescope with a single broad-band filter. This single-band classification relies on the wavelength-dependence of the PSF, which leads to small yet significant changes between images of stars with different spectra. We use supervised machine learning to interpret these changes and predict spectral classes. Such a single-band classification can quickly deliver information about stellar types and colours, even in the absence of multi-band photometry or spectroscopic follow-up. Such information may be useful for selecting stars to be used for modelling the wavelength-dependent PSF of, e.g. *Euclid*. The inner workings of the single-band classifier that we developed for this study can be summarised as follows.

First, we project the stellar images onto a basis obtained from principal component analysis. This reduces the information content of each stellar image to a set of coefficients. Through experimentation, we find that good results are obtained when considering about 25 PCA coefficients from 40×40 pixels stamps centered on the target stars. Second, we train committees of feed-forward artificial neural networks to predict the stellar types based on these PCA coefficients. We obtain best results for networks with 2 to 3 hidden layers of 25 to 30 nodes each.

We perform all our analyses with simulated stellar images from several optical setups: HST ACS using the F606W or F814W filter, and the *Euclid* VIS filter. While we use simple uniform distributions of spectral types and S/Ns, we include the complications of spatially variable PSFs, reddening, and contamination by companion objects. We stress that the purpose of

testing these different instrumental and observational conditions is not to compare them, but to demonstrate the general feasibility of the suggested approach. Performing a comparison would require focusing on a particular scientific interest, involving a specific stellar population.

Our technique is most efficient with broad pass-bands such as the *Euclid* VIS band. However, we show that even the commonly used filters of the ACS (*F606W* and *F814W*) are broad enough to obtain a reliable stellar classification. This may allow one to use archival HST data taken in one single band to infer information, for example on the stellar populations of resolved stellar clusters. Still, the goal of the present work is to describe a proof-of-concept classifier. Improvements leading to a full classifier, possibly used for *Euclid*, may include the following items:

- The efficiency of the dimensionality reduction could benefit from a better prior centering of the sources, potentially on a finer pixel grid. In the present paper we simulate centering errors as large as half a pixel.
- The PCA decomposition could be replaced with a different one, specifically suited to catch the wavelength-dependent features in the PSF, for example wavelets, starlets, shapelets, etc.
- Any spatial variation of the PSF across the detector could be properly accounted for, and not just marginalised over. This could be achieved by training different classifiers for different locations of the detector, or by using the detector location as input feature to the machine learning.
- Instead of performing a regression of a continuous parameter whose value encodes the classification, the requested output could be better adapted to the desired use. For example, it might be more meaningful to predict colours instead of spectral types, or to use a softmax regression to obtain probabilities for distinct classes of interest (Nielsen 2015).

The results of this method do not depend much on the exact value of the hyper-parameters, which facilitates the optimisation. However, the training strategy is still survey-dependent. For a space telescope, we are fortunate that the PSF can be modelled fairly easily, hence leading to clean and arbitrarily large training sets. Another strategy is to train the ANNs on actual stellar images with known spectral types. This might be a viable strategy for *Euclid*, given its exceptional PSF stability, the depth of the survey beyond that of *Gaia*, and the broad-band VIS filter.

Acknowledgements. The authors would like to thank Rémy Joseph for useful discussions as well as Patrick Hudelot, Koryo Okumura and Samuel Ronayette for providing the *Euclid* simulated PSFs. We are grateful to the anonymous referee for the valuable comments that improved the quality of this work. This work is supported by the Swiss National Science Foundation. M.T. acknowledges

support from a fellowship of the Alexander von Humboldt Foundation. This research made use of *Astropy* (Astropy Collaboration et al. 2013), *Matplotlib* (Hunter 2007) and *Numpy* (van der Walt et al. 2011).

References

- Robitaille, T. P., Tollerud, E. J., Astropy Collaboration, et al. 2013, *A&A*, **558**, A33
- Avila, R., et al. 2016, ACS Instrument Handbook, version 15.0 (Baltimore: STScI)
- Bengio, Y. 2009, *Foundations and trends® in Machine Learning*, **2**, 1
- Bishop, C. M. 1995, *Neural Networks for Pattern Recognition* (New York, NY, USA: Oxford University Press, Inc.)
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, in *Interstellar Dust*, eds. L. J. Allamandola, & A. G. G. M. Tielens, IAU Symp., 135, 5
- Chabrier, G. 2003, *PASP*, **115**, 763
- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **431**, 3103
- Cypriano, E. S., Amara, A., Voigt, L. M., et al. 2010, *MNRAS*, **405**, 494
- de Bruijne, Allen, M., Azaz, S., et al. 2015, *A&A*, **576**, A74
- Ford, H. C., Feldman, P. D., Golimowski, D. A., et al. 1996, in *Space Telescopes and Instruments IV*, *Proc. SPIE*, **2807**, 184
- Gentile, M., Courbin, F., & Meylan, G. 2013, *A&A*, **549**, A1
- Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2014, *MNRAS*, **441**, 1741
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The elements of statistical learning: data mining, inference and prediction*, 2nd edn. (Springer)
- Hunter, J. 2007, *Comput. Sci. Eng.*, **9**, 90
- Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, *Statistics, Data Mining and Machine Learning in Astronomy* (Princeton University Press)
- Jarvis, M., & Jain, B. 2004, ArXiv e-prints [[arXiv:astro-ph/0412234](https://arxiv.org/abs/astro-ph/0412234)]
- Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in *SPIE Conf. Ser.*, **8127**, 0
- Kuntzer, T., Courbin, F., & Meylan, G. 2016, *A&A*, **586**, A74
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *Euclid Study Report*, ArXiv e-prints [[arXiv:1110.3193](https://arxiv.org/abs/1110.3193)]
- Massey, R., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, **429**, 661
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *Apl*, **808**, 16
- Nguyen, A., Yosinski, J., & Clune, J. 2015, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- Nielsen, M. 2015, *Neural Networks and Deep Learning* (Determination Press)
- Niemi, S.-M., Kitching, T. D., & Cropper, M. 2015, *MNRAS*, **454**, 1221
- Nissen, S. 2003, Report, Department of Computer Science University of Copenhagen (DIKU), 31
- Pearson, K. 1901, *Philosophical Magazine Series 6*, **559**
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Pickles, A. J. 1998, *PASP*, **110**, 863
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, **409**, 523
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, **737**, 103
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Semoloni, E., Hoekstra, H., Huang, Z., et al. 2013, *MNRAS*, **432**, 2385
- Shlens, J. 2014, ArXiv e-prints [[arXiv:1404.1100](https://arxiv.org/abs/1404.1100)]
- Sirianni, M., Jee, M. J., Benítez, N., et al. 2005, *PASP*, **117**, 1049
- Smiljanic, Korn, A. J., Bergemann, M., et al. 2014, *A&A*, **570**, A122
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints [[arXiv:1503.03757](https://arxiv.org/abs/1503.03757)]
- van der Walt, S., Colbert, S., & Varoquaux, G. 2011, *Comp. Sci. Eng.*, **13**, 22
- Voigt, L. M., Bridle, S. L., Amara, A., et al. 2012, *MNRAS*, **421**, 1385
- Yang, T., & Li, X. 2015, *MNRAS*, **452**, 158

5.2.3 Predicting a colour-magnitude diagram from HST single-band data

The stellar classifier presented in the above is foreseen to be applied to *Euclid*-like data. It can also be applied to HST data taken in broad filters. In this section, we predict a colour-magnitude diagram (CMD) from single-band data. CMDs are an important tool to study stellar evolution, and stellar clusters (for an introduction, see [Maeder 2009](#)). The colour is computed from two magnitudes. In this case, the colour is $V - I$ from the V (filter F606W) and I (F814W) bands, which are narrower than the VIS filter, and have a lower efficiency (see Fig. 5.13). We will use data taken by the advanced camera for survey/wide field channel (ACS/WFC) instrument. We tweak the stellar classifier into a regression technique that predicts directly the colour of an object from single-band images of stars.

Description of the data

We will use data from six different clusters: (i) NGC 6397, the second closest globular cluster, (ii) NGC 5053, (iii) NGC 6779, (iv) NGC 5466, (v) NGC NGC6366, and (vi) Lyngå-7, all observed by the HST ACS/WFC. Each of these clusters were imaged in F606W and F814W. Further characteristics are presented in Tab. 5.1. A catalogue containing at least the position, magnitude in one of the filter and colour is available for each of the clusters.

Table 5.1 – List of clusters used in the stellar classifier application to real data. More physical characteristics can be found in the references.

Cluster	R.A.	Dec.	Total time (s) ^a	Dataset	Obs.	E(B-V)	Refs
NGC 5053	13h16m27.0s	+17°41'53"	1730 (V), 1780 (I)	J9L902	03/06	0.01	1,2
NGC 5466	14h05m27.3s	+28°32'04"	1730 (V), 1780 (I)	J9L903	04/06	0.02	1,2
NGC 6366	17h27m44.3s	-05°04'36"	570 (V), 570 (I)	J9L907	03/06	0.75	1,2
NGC 6397	17h40m42.1s	-53°40'28"	2170 (V), 4200 (I)	J97101	04/05	0.186	2,3,4
NGC 6779	19h16m25.5s	+30°11'05"	1730 (V), 1780 (I)	J9L905	05/06	0.26	1,2
Lyngå-7	16h11m03.0s	-55°18'52"	1835 (V), 1835 (I)	J9L904	04/06	0.78	1,2

References. (1) [Sarajedini et al. \(2007\)](#); (2) [Anderson et al. \(2008\)](#); (3) [Hansen et al. \(2007\)](#); (4) [Richer et al. \(2008\)](#);

Notes. ^(a) This is the exposure time per frame used in this study. This does not reflect individual frame exposure time.

NGC 6397 will be the main cluster for this study. [Richer et al. \(2008\)](#) built a CMD based on 2324 stars following reduction techniques detailed in [Anderson et al. \(2008\)](#). The cluster was observed under proposal GO-10424 (PI: H. Richer). Calibration of the HST ACS data was done following [Sirianni et al. \(2005\)](#). Validation of the calibration was carried out by comparing the photometric results to ACS Galactic Globular Clusters Survey (ACS GGC Survey, [Sarajedini et al. 2007](#)). Objects were identified *ab initio*; the detection catalogue was cleaned by cutting too large or small objects and artefacts. The final catalogue was produced by selecting stars based on their proper motion. This catalogue only lists stellar positions relative to an image that is no longer available. Drizzled images of NGC 6397 are however available, but a linear transformation of the catalogue was necessary to match the catalogue to the stars. Seventy-two

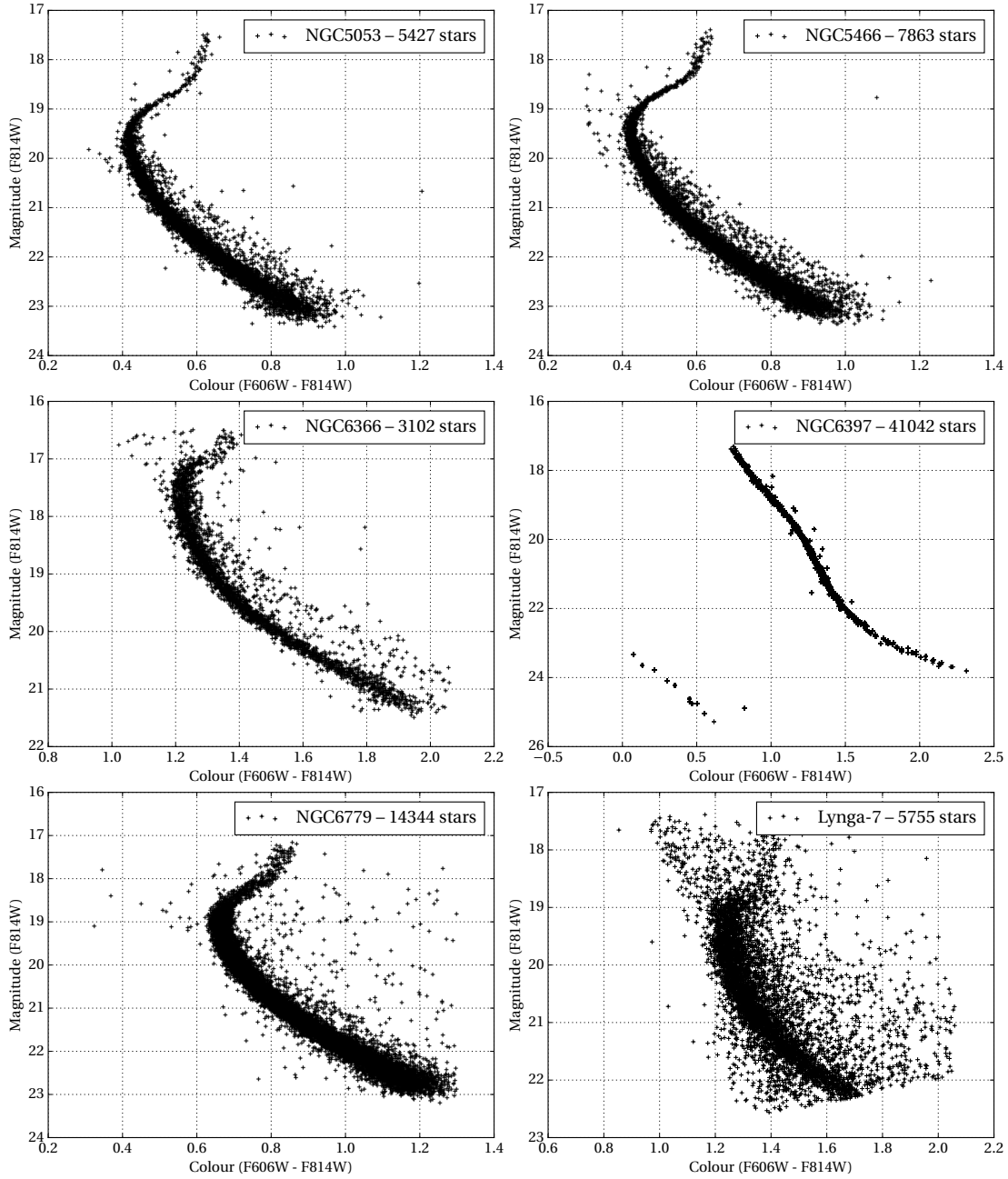


Figure 5.17 – Cleaned CMDs of the cluster used in this study. The subtle double main sequences (clearly visible in the two top panels) indicate the binary population in the cluster. This could also be a technique used to solve the issue raised in Sect. 5.1. For illustration purposes, the WD branch is included in the NGC 6397 diagram. We note that this WD branch does not show all members found in [Richer et al. \(2008\)](#). This cluster is particularly clean as stellar members were proper-motion selected.

images (thirty-six images in each filter) were downloaded and analysed. Source identification was made without input catalogue with the SExtractor package (Bertin and Arnouts 1996). Detected objects and catalogues are cross-matched. To avoid close stellar blending, objects detected with the profile blend flag are rejected. Objects with a too large full width at half maximum (FWHM), which could be badly matched to small galaxies or artefacts are also removed. Finally, glsplwd are removed too, in order to simplify this first real data application. On average, about 1400 stars could be identified in each of the thirty-six pairs of F606W and F814W images. A total of $\sim 40\,000$ stars are extracted from the images, in units of electrons per second. We note that this means that there are about twelve realisations of each of 2324 stars in the catalogue.

Observations of the five other clusters, NGC 5053, NGC 5466, NGC 6366, NGC 6779 and Lyngå-7, were part of the ACS GGC Survey campaign (GO-10775, PI: A. Sarajedini, Sarajedini et al. 2007). This program aimed at providing photometry for stars with $S/N \gtrsim 10$ using ACS/WFC. Clusters in the ACS GGC Survey were not observed before 2006. A likely reason is that these clusters have a low stellar density. The reduction of the data is carried out following Anderson et al. (2008). This entails that observations of these clusters similar to NGC 6397 are reduced in the same way. Low stellar density decreases blending, hence reduces errors due to poorly identified nearby stars and blends in training sets. For each cluster, a drizzled image is available per filter. The catalogue contains the position in equatorial coordinate system, such that extracting stellar stamps is somewhat easier⁹. CMDs for the six clusters are shown on Fig. 5.17. We apply FWHM, magnitude error, and blending flag cuts to the matched catalogues. We extract about 25% of the objects in the catalogues in each cluster. This low efficiency is due to the fact that a lot of the objects in the catalogues are not stars (some small galaxies appear in multiple entries), or suffer from artefacts. The stamps, as for NGC 6793 are in units of electrons per second and are 24 pixels on a side.

Mock HST data is also generated. We use the TinyTim software (Krist et al. 2011) to prepare the simulation of the images. The procedure used to prepare HST-like observations for the stellar classifier, as described in Sect. 3.2 of Paper III (Sect. 5.2.2), is slightly altered. Mock stars are not simulated with a template SED, but as a black body, with temperatures uniformly distributed between 1 600 and 3 600 K. An extinction $E(V - B)$ of 0.186, the same as NGC 6397, is applied by the TinyTim library. Fluxes are uniformly sampled between F814W magnitudes 17 and 24, to reflect the apparent magnitude of members of NGC 6397. The PySynPhot code (STScI Development Team 2013) is used to estimate the flux in the F606W band given the flux in F814W taking into account the extinction. The library is designed to simulate HST photometric data. 50 000 stellar images are simulated.

F814W stamps are prepared for processing by dividing the sample into two dataset: training and testing. Stars that have multiple realisations, which is the case of some of the simulations and NGC 6397 observations, cannot be included in both sets. In line with other studies conducted in this thesis, we dedicate 60 to 75% of the data to training.

⁹Somewhat, as there is a $\mathcal{O}(10)$ pixels shift between the coordinate reference of the catalogue and the image...

Modifications of the stellar classifier algorithm

Predicting a colour is a regression task while predicting spectral type was a classification task. Due to the design of the classifier, the necessary adaptations to the algorithm are minor. The output layer in the network predicts the colour, not the class representation. The technique is still a two-step process: finding a new representation of the data and ANNs to predict the colour. The ground truth is no longer a numerical representation of the spectral class but directly the unscaled colour. When working on real observations, the architecture of the network can be simplified from the optimised architecture described in Tab. 2 of Paper III. This new architecture was selected empirically and is not justified by a thorough exploration of the hyperparameter space.

While an human inspection of the simulated PSFs is satisfactory, it is known that TinyTim does not produce accurate PSF profiles. Using empirical reconstruction techniques is the recommended path when the required fidelity is high¹⁰. In our case, a high PSF fidelity seems to be necessary to correctly infer the colour from the profile. In any case, we reduce the size of the committee to seven members and four voting members in the first numerical experiment, and to 18 voters out of 21 for the others. The predictions are averaged only at the post-processing stage, when creating the CMD from the predictions. As for the classification task, ANNs are trained using the MSE cost function. Results presented in the following were obtained for networks trained within thirty minutes.

Numerical experiments

We design four experiments. First, we train a committee on NGC 6397 stars, and predict the colour of test stars in NGC 6397. This experiment would mimic a case where sufficient amount of data is available in two filters, and there is missing data. We refer to this experiment as same cluster experiment.

Second, we train on NGC 5466 stars, and predict colours for NGC 5053. These two clusters have very similar extinction. This is thus the different cluster, but same extinction experiment.

A third committee learn to predict colours from the five ACS GG clusters, and is tested on NGC 6397. This is akin to a case in which a large observational database of stars was available and used to predict colours of stars observed in a different region of sky. In this experiment, the extinctions are not the same in all clusters.

The fourth and last experiment optimises the networks from TinyTim simulations and tests the performance on similar simulations and NGC 6397. In this application, a modelled PSF is used to predict the colour of observed stars.

¹⁰See <http://www.stsci.edu/hst/observatory/focus/TinyTim> for a short discussion.

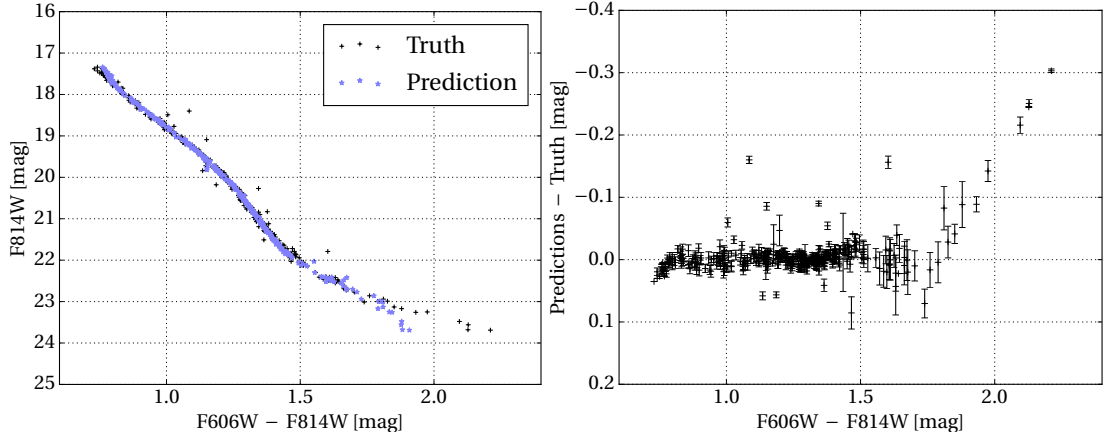


Figure 5.18 – Same cluster experiment. (*Left.*) Predicted CMD from the F814W image of the stars (blue stars) compared to their true value (black crosses). (*Right.*) Errors as a function of the true colour. Error bars show the 1σ uncertainty on the predictions from the prediction of the different realisations.

Results

The results of the same cluster experiment shows that the committee can predict unbiased colours, up to $V - I \approx 1.7$ mag, as shown in Fig. 5.18. At these colours, the S/N of the stars is low, which could explain the divergence of the predictions. The 1σ error below the diverging colour limit is 0.02 mag. The network architectures are set to ten PCA components, and three layers of twenty neurons. There are some objects that do not lie in the main sequence. Such objects were not purged out of the training set. They represent outliers of the data, but are still predicted to be in the MS. The committee learned to predict the MS from a clean training set. While the networks are extremely successful at predicting MS colours, the limitation of this same-cluster approach is directly apparent. Outliers are ignored and a large error is made. This same-cluster approach can only be used for cases in which the data are extremely similar.

In the second experiment, the different cluster but same extinction, we train with a much noisier MS. The red giant branch at the bright end of the sequence shows the higher complexity of the data. The capacity of the networks is increased to twenty principle components and three layers of twenty-five neurons. The results, displayed in Fig. 5.19, show unbiased predictions, but the 1σ error doubles to ~ 0.04 mag. This is not due to different clusters: metrics for the training (NGC 5466) and test (NGC 5053) datasets are very similar. The mapping between magnitude and colour is no longer bijective. In both clusters, a binary sequence is present. The binary sequence is displaced by about 0.1 mag towards a redder colour in NGC 5053. A fraction of the binary stars are not predicted correctly. This suggests a partial success in reproducing the binary sequence. Partial because there is a hint of the binary sequence in the residues of the prediction around a colour index of 0.8 (Fig. 5.19). The turn-off between I magnitude ~ 19.5 and 18 is not well modelled by the networks. Incorrect modelling could be

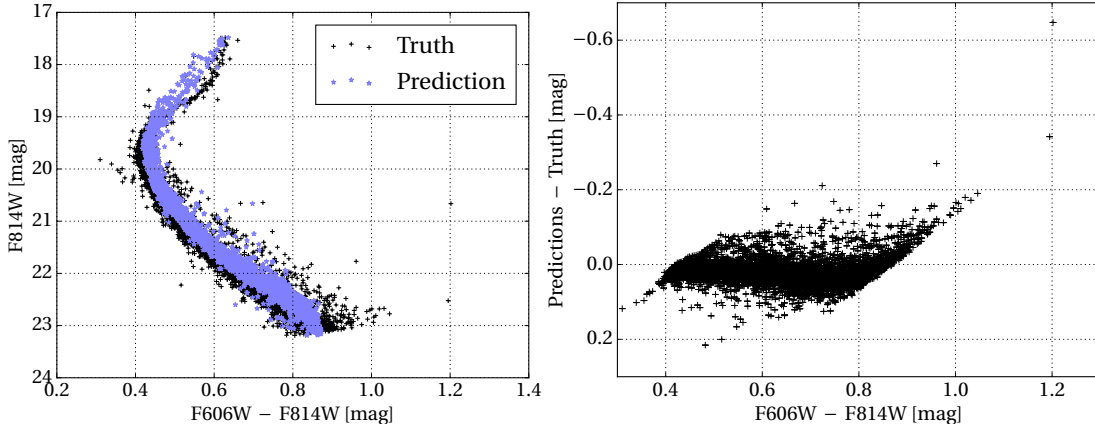


Figure 5.19 – Committee trained with NGC 5466 and tested on NGC 5053. Same panels as in Fig. 5.19. The right panel, showing the residues, does not have error bars, as there is only one realisation of the stars.

picked up by the large dispersion of the predictions. While the turn-off is relatively brisk, a variation of 0.2 colour index in less than one magnitude, this happens at high S/N data. We note however, that it is a relatively low density region of the CMD, which can explain the poor performance. That being said, our colour estimator is accurate and, relatively to the sequence noisiness, precise.

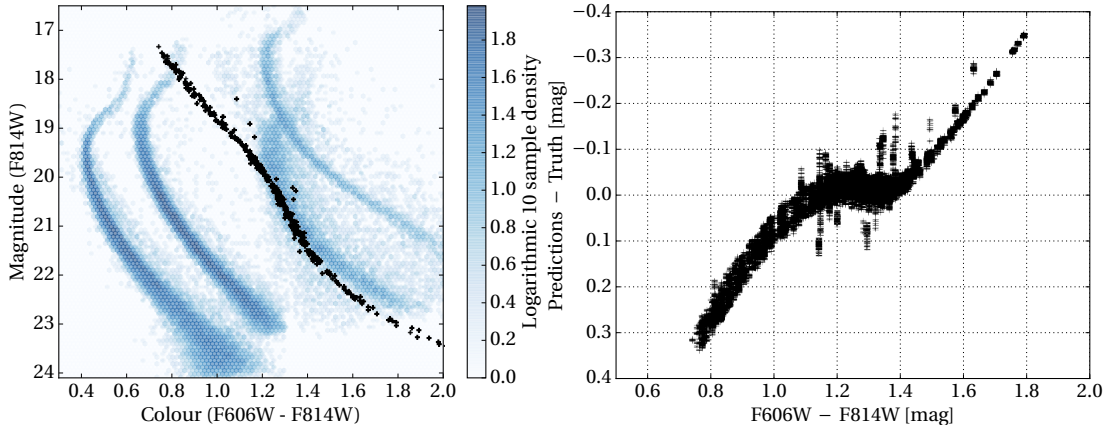


Figure 5.20 – Committee trained with ACS GGC images, and tested against NGC 6397. (Left.) Comparison between the NGC 6397 data (in black) and the training dataset. (Right.) Residues of the predictions of NGC 6397 data. The error bars show the 1σ variability of the committee's members predictions.

In the third experiment, networks taught with the ACS GG clusters and tested on NGC 6397, the features of the stellar sequences are very different (see left panel of Fig. 5.20). The turn-off in the main sequence is present in some clusters of ACS GGC, but absent in NGC 6397. Colours are also different: the typical index range of ACS GGC is 0.4 – 2.2, while NGC 6397's index

spans 0.6 – 2.5. There are also differences in extinction. We loosened some of the selection criteria to extend the training set to lower S/N stars. The training sample however does not span the whole NGC 6397 main sequence, especially at faint magnitude and high colours. Poor predictions in this region can therefore be expected. As noted in Sect. 5.3 of Paper III, mismatch in the training and test sets extinctions can bias the results. The test predictions on NGC 6397 were indeed biased by 0.2 magnitude in colour, that is the value of the cluster’s extinction. Predictions must therefore be calibrated by a straightforward correction: adding the value of the extinction. Residues are shown in the right panel of Fig. 5.20. In the region well predicted, between 0.9 and 1.5 magnitude, the 1σ error is less than 0.05 magnitude, and gives a non-biases estimates. As anticipated above, there is a clear deviation for stars whose colour index is $\gtrsim 1.5$ magnitude. This is due to the absence of samples in this region of the CMD training dataset. At low colour indices, the errors are also important. Reasons for this mismatch probably reside in the data included in the training samples. At these low colour indices, the training stars have left the main sequence: they are in the red giant phase. Those stars are not representative of the MS stars at the same colours in NGC 6397.

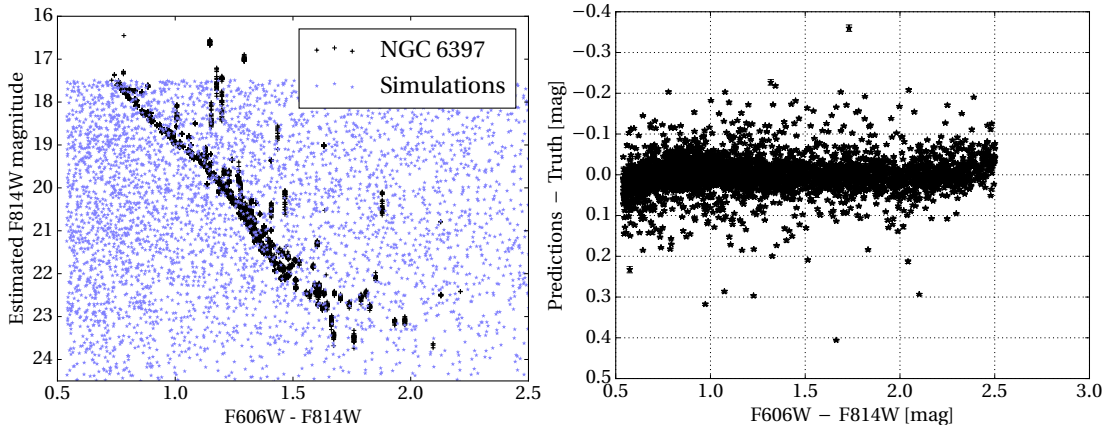


Figure 5.21 – Committee trained with simulations, and tested against simulations too. (*Left.*) Comparison of the simulated training set in blue and NGC 6793 stars on a CMD plot. About 6 000 simulated data points are shown. Real stars brighter than the training sample are outliers and does not influence the results. (*Right.*) Residues of the predictions on a validation set made up of simulations similar to the training data. The error bars show the 1σ variability of the committee’s members predictions.

We turn to the fourth experiment: training with simulations and testing on NGC 6397. An important difference between the three above experiments and this one is the distribution of the objects in the CMD. There is no longer a sequence of stars along which most objects live, the stars are at random positions in the diagram. With this approach, removing the WD branch, or specially boosting the binary population becomes unnecessary, as we properly sample the colour-magnitude space, as demonstrated by the left panel of Fig. 5.21. This is more ambitious than the other experiments. We first present the result on a simulated test dataset. Flux and temperature distributions are the same in the training and testing samples.

Training is more difficult than with real data which exhibit a sequence in the CMD. It takes a longer time, and models likely requiring deeper architectures. A deeper architecture is, by itself, harder to train than a shallower network, because of the vanishing gradient problem (see Sect. 4.3.3). The model is, as for the other experiments, unbiased, but less precise: the 1σ distribution of the error on the colour is $\mathcal{O}(0.04)$ mag. The result is promising as this means that this method does not necessarily need a colour-magnitude sequence to produce good results. When the networks are applied to NGC 6397, they break down: the predictions are not correlated at all with their ground truth. The reason for this breakdown is not known at the time of writing. Distributions of flux, colour and count rate in the images are similar. This leaves open two main options: (i) an undiagnosed improper processing of the data, or (ii) TinyTim-generated PSFs are unsuited for such an application.

Summary and outlooks

We demonstrated that a modified stellar classifier can be used as an accurate and relatively precise estimator of the colour. Four numerical experiments were carried out, from the simple training and testing on stars extracted from the same clusters to training on simulated data. All experiments, with the notable exception of the simulated dataset training, exhibit acceptable to good performance. We note that increasing the size of the training set, by including more clusters, or even field stars, would most likely translate into a better performance. When training and predicting simulated datasets, the performance remains good. The architecture of these networks is more complex than if model parameters are inferred from real data. This increase in the complexity of networks either suggest that mock data do not reflect well the data, or that networks exploit the properties of the sequence of stars to be fitted. As there is no stellar sequence buried in the simulated stars, the flux in the stamp does not yield information. Ergo, it may be harder for networks to distinguish stars, as the flux is easily computed from the input features. The extremely poor performance simulation-trained networks are applied to NGC 6793 data is currently not yet understood.

It seems that individual committee members are weak learners. Combining eighteen of them reduces predictions errors by a factor of two. This forest of ANN is however costly to train. A single member is a two-step algorithm: first finding a good representation, and second predicting some parameter from this new representation. This process could be advantageously replaced by convolutional neural networks; a route to be explored in further work. Throughout this study, we extracted F814W data to predict the colour. Changing to F606W stamps does not significantly alter the performance, at least on real data. Instead of returning the $V - I$ colour, the effective temperature of the star could be predicted, or an other colour index. While the first three experiments were successful, further work is needed to demonstrate the training on simulated data. This could very entail modelling the HST PSF with another tool than TinyTim. We highlight that the performances obtained are promising. They suggest that such a tool could be applied to *Euclid* data.

5.3 PSF measurement and reconstruction

This section describe the last step in the PSF pipeline module: PSF reconstruction. The output of the previous modules were which point-source-like objects to use and their SEDs. Outputs of the PSF reconstruction are parameters that completely describe the PSF model. PSF models can be completely data-driven, without constraints from the optical design. Opposite models are also possible: strong design constraints and a few free parameters to optimise.

In this section, we introduce a data-driven PSF reconstruction based on auto-encoders (Sect. 4.5). We demonstrate that AEs are capable of modelling narrow-band and wavelength-dependent PSFs. In addition we show that we can interpolate the PSF in code space, which allows to determine the PSF at any point on the image, and with any SED. Requirements on the PSF reconstruction for stage III and IV weak-lensing surveys are tight. Wavelength PSF interpolation had not been required before *Euclid* to account for colour biases. Those demands prompted a recent interest in new PSF methods.

This study was conducted in collaboration between the author of this thesis and Romain Meyer, master student at LaSro, currently pursuing a PhD at University College London. He implemented the final version of the approach, and ran the PSF reconstruction pipeline to measure the performance of the method.

5.3.1 *Euclid* typical PSF and stellar population

Features of a PSF are determined by the optical design of the telescope, the temperature of the hardware and various impurities. The design of the telescope does not change in the months preceding the launch. The operating temperature of the telescope is a controlled quantity. The ideal PSF can thus be simulated based on the physical characteristics of the system. PSFs are simulated by propagating wave fronts in a model of the telescope. A well-known example is the TinyTim software for HST PSFs. As the PSF depends on wavelength, monochrome slices of PSFs must be simulated. The PSF image, as seen on the detector, is a sum of the monochrome slices weighted by the SED. PSFs build in this way remain simulations, and do not encapsulate full complexity. Effects of vibrations of the launch and impurities are not accounted for. Impurities could take the form of, e.g., damage sustained during assembly, which cannot be modelled.

Jérôme Amiaux, Patrick Hudelot, Koryo Okumura, and Samuel Ronayette worked towards proposing six hundred *Euclid* VIS PSF images in the four central CCDs. The resolution of the images is twelve times better than the real VIS pixel scale, and the images are sampled in wavelength every 5 nm. The image of one of these PSFs is shown in Fig. 5.22, left panel. Those PSFs have been extensively used over the course of this thesis (See Sect. 5.1.6, 5.2.2, and this one). Six hundreds is not a sufficiently large number to train a network, let alone assess its predictions with confidence. We used the GalSim library to create a wavelength-depend mock PSF that look reasonably similar to *Euclid* VIS (Fig. 5.22, right panel). In this section, we



Figure 5.22 – (Left.) *Euclid* VIS-like PSF as simulated with the software Zeemax. (Right.) VIS-like PSF used to test AE PSF reconstruction methods, generated with GalSim. Two PSF images are sampled at one twelfth the VIS pixel scale at a monochrome wavelength of 550 nm.

explore novel concepts for PSF reconstruction. The mock PSF that is used to demonstrate the capabilities does not need to match perfectly the expected PSF, but does need to capture the main features and parameters typical variations across the field.

Cropper et al. (2013) suggest that there will be between 2 000 and 3 000 stars per field (that is per field-of-view of 0.54 deg^2) with 30% of corrupted stellar images. There would be ~ 1750 stars per frame, and following *Euclid* requirements, the PSF should be stable for four exposures, yielding a sample size of ~ 7000 stars per epoch. We use BMG simulations of typical *Euclid* weak-lensing pointings to infer realistic distributions of spectral types and magnitude. Magnitudes can be converted to fluxes and S/N estimated (see details of the conversion in

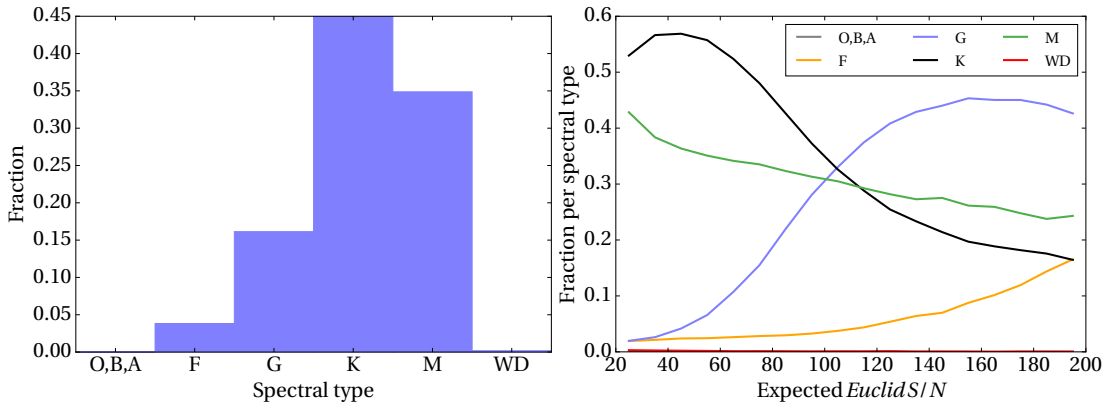


Figure 5.23 – Typical distributions of *Euclid* stars observed by VIS, selected to have $20 \leq S/N \leq 200$ from BMG realisations. (Left.) Overall distribution of spectral class. (Right.) Evolution of the fraction of stars per spectral class as a function of S/N. We note that these distributions of S/N are only order of magnitude estimates. A detailed computation of the S/N could lead to different distributions.

Sect. 5.4.6). Figure 5.23 shows distributions of *Euclid* VIS stars, selected to have a minimum S/N of 20 and be fainter than magnitude 18. This represents a faint magnitude limit at ~ 22.5 mag. The distribution of stars per S/N reveal that, with these cuts, $\sim 50\%$ of the stars are at S/N lower than 50. An overwhelming fraction of stars will be in types F through M in the MK classification. The fraction of stars per spectral type evolves with magnitude, or, equivalently, with S/N. At low S/N, the sample is dominated by K and M type stars. In a region centred around $S/N = 160$, G and M stars dominate. This is important, as the sampling of SEDs changes, we are forced to use different levels of S/N to reconstruction different wavelength-dependent PSFs. Red PSFs will be easier to reconstruct than bluer PSFs, because the flux will be higher for the former. That being said, a *Euclid* PSF reconstruction method should meet the requirements with the typical distributions of stars presented previously, whatever the SED.

Table 5.2 – List of the different datasets used in the PSF reconstruction algorithm study.

Dataset	Dither?	S/N distribution	SED	SED distribution
A	✗	Flat	1 (G5V)	N/A
B	✓	Flat	1 (G5V)	N/A
C	✓	Flat	27 (Pickles)	Flat
D	✓	Typical	27 (Pickles)	Typical

We test our proof-of-concept approach on increasingly complex data, without reaching full complexity. The datasets are based on the GalSim PSF presented in Fig. 5.22, right panel with *Euclid*-like PSF parameter fields. Table 5.2 lists the different characteristics of the datasets. The code used to simulate the data is publicly available¹¹. The size of training sets is $5 \cdot 10^4$ images, 48 pixels on a side, irrespective of the complexity of the data. This size represents roughly five to six times more data that would be obtained from a single PSF epoch. For the non-realistic datasets, we prepare three levels of Gaussian noise: (i) no noise, (ii) objects with $S/N = 100$, and (iii) a low signal, $S/N = 20$. For the non-realistic datasets, we measure the S/N with SExtractor¹². In summary, we test AE-based methods on three simple datasets (A: unrealistic no dithering, no SED; B: dithering, no SED; C: dithering and SEDs), each at three noise levels, and (D) a VIS-like dataset with typical SED and S/N distribution.

5.3.2 Description of the approach

We propose to separate the problem of PSF reconstruction into two sub-problems: (i) finding a new representation of the PSF, and (ii) interpolating the new representation on a PSF field. We set out by simulating the data according to the procedure laid out in the previous section. This requires simulated PSFs, SEDs, an observed stellar population and image characteristics (e.g., pixel scale). The two sub-problems are treated as two independent modules, with the interpolation module relying on the outputs of the representation module. A visualisation of the concept is shown in Fig. 5.24.

¹¹<https://github.com/kuntzer/pibe>

¹²Defined as $FLUX_AUTO/FLUX_AUTOERR$.

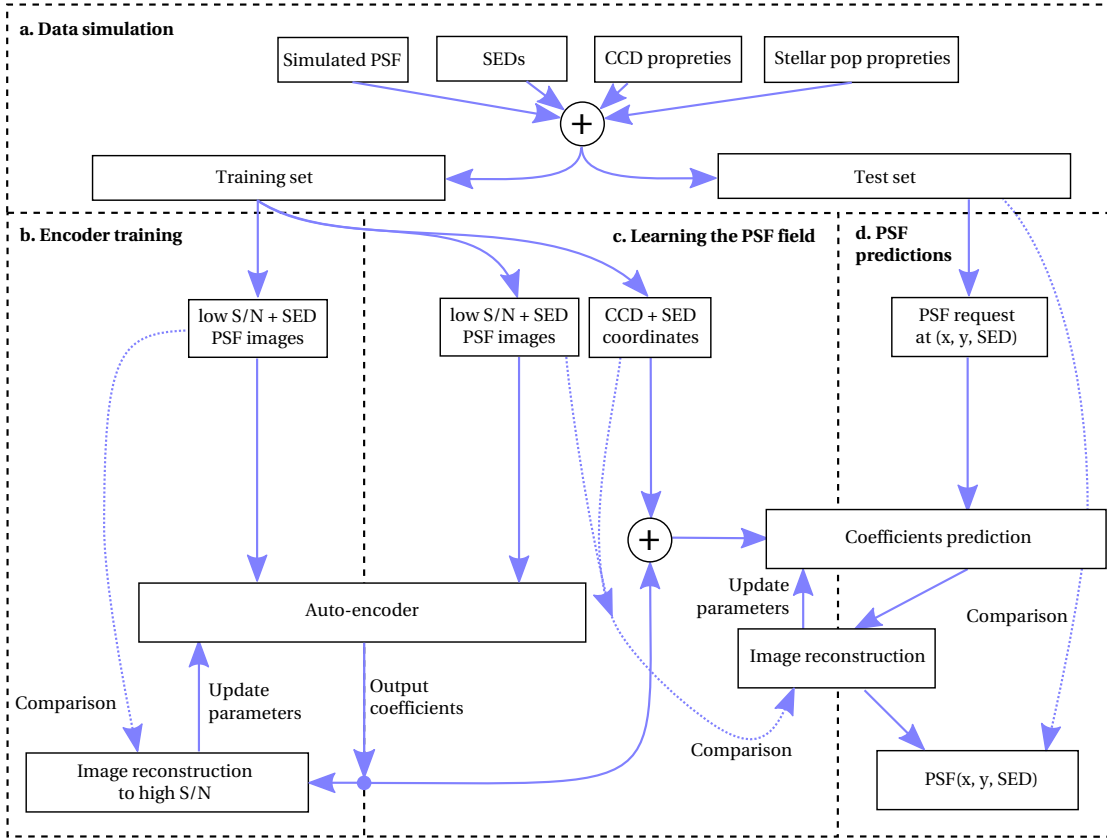


Figure 5.24 – Concept of the AE PSF reconstruction algorithm. (a.) Data simulation and processing in two separate and independent sets. (b.) Training of the auto-encoders. Training samples are PSF images. (c.) Preparation of the interpolations, one coefficient per interpolant. (d.) Test/production phase. A PSF with given SED is request at given coordinates. Coefficients are interpolated, and decoded with the auto-encoder to yield the PSF image at high S/N.

Auto-encoders find (i) a better representation of the PSF than noisy pixels, and (ii) an encoding/decoding mapping from the image to the code and conversely. An image undergoing encoding and decoding, the two stages of AEs, is reconstructed at a higher S/N than the input. AEs are similar to PCA in the sense that the amount of information that remains after encoding/decoding is limited, and that noise is not relayed unless the dimension is not reduced. The output of the encoding module is the AE model parameters.

The second module is the interpolation. Point-sources images are encoded with trained AE yielding coefficients to be interpolated. If the coefficients trace the variations of the PSF in the field, they can be interpolated to any given coordinates to provide estimations of the PSF. There is no reason to get variations that are similar across the different coefficients. Interpolation rules should be derived for every coefficients. Coefficients are interpolated in three-dimensional space: two positions and one PSF SED. Any further variables, like ageing of the sensors, increase the dimension of the interpolation space. In the initial concept,

interpolants would likely have been ANNs. The motivation was that they can learn well in high dimensions despite a low density of points (see Sect. 4.2.2). Coefficients maps revealed to be smoother than anticipated, which suggested using simpler and faster methods, like polynomial fitting.

5.3.3 Quality metric evaluation

The next step is the validation of the method with a test set of arbitrary complexity, and the evaluation of a performance metric. We propose a metric Q which was derived from the GREAT3 metrics (see eq. (8) of Mandelbaum et al. 2015) and is calculated for N examples as follows:

$$Q = \frac{\eta}{\sqrt{\sigma_{min}^2 + \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{e_{1,i}^{int} - e_{1,i}^{true}}{\alpha} \right)^2 + \left(\frac{e_{2,i}^{int} - e_{2,i}^{true}}{\alpha} \right)^2 + \left(\frac{(R_i^{int} - R_i^{true})}{\langle R^2 \rangle^{int}} \beta \right)^2 \right]}}, \quad (5.15)$$

with calibration terms, $\eta = 2000$ and $\sigma_{min}^2 = 1$, such that $Q = 1000$ corresponds to meeting the Euclid PSF reconstruction requirements. Calibration terms additionally ensures that Q cannot diverge. $\alpha = \sigma(e) = 2 \cdot 10^{-4}$ is the target ellipticity stability and $\beta = \sigma(R^2) / \langle R^2 \rangle = 1 \cdot 10^{-3}$ is target size stability.

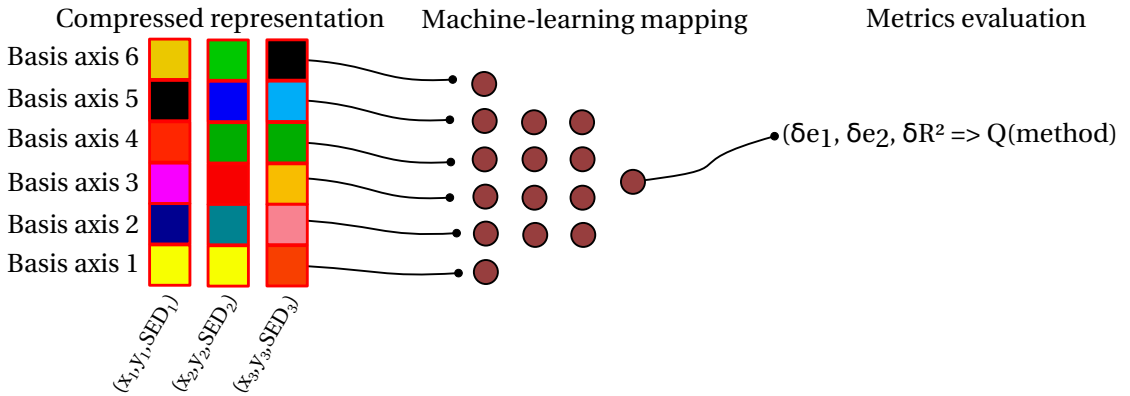


Figure 5.25 – Concept of the learned-performance metric, which is a similar to a look-up performance table, but with ANNs. PSFs representations are the input for ANNs which map them to ellipticities and size. The last step is to evaluate the Q metric.

Measuring shape parameters of PSF images in under-sampled and noisy data is biased, even with state-of-the-art methods (see Sect. 5.4). To circumvent this unreliable shape measurement step, we propose to use a learned-performance metric. This metric directly maps the compressed representation of the PSF to the shape parameter space. The extracted shape parameters are compared to the fiducial shape parameters at the right spatial position as per equation (5.15). This process requires building a metric training set for each of the PSF reconstruction method. Compressed representations of the PSFs have $\mathcal{O}(10^1 - 10^2)$ coefficients,

that is the mapping of compressed representation to shape parameters is many dimensional. We populate a training set at the coordinates for which we evaluate AE PSF reconstruction algorithms. To train the method, we encode profiles whose shape parameters are perfectly known, feed the coefficients as the input features to the network and use shape parameters as ground truth. This implies encoding images of the PSF at the evaluation coordinates for each of the implementations. This learned performance metric uses ANNs to learn the mappings (Fig. 5.25). We found that simple two-layered five neurons networks showed were sufficient. No bias in the recovered shape parameters was detected. These bias tests included exploring sensitivity of the predictions as a function of the ellipticities and the size, and no such conditional bias was discovered. Such a learned-performance metric can directly map coefficients to shape parameters. There could be a PSF reconstruction method whose internal representation of the PSF is complicated, and could not be learned by ANNs. Further limitations of the method reside in the training phase of the metric. Simulations must be generated such that the mapping coefficients to shape parameters can be learned with confidence, which demands a lot of resources. To evaluate the performance of the PSF reconstruction, we trained the learned metric with 50 000 samples.

Testing the performance of a PSF reconstruction technique is hard for two main reasons. First, there should be no bias when measuring the shape parameters. When measuring resolved, but small and faint objects, like WL-galaxies, shape estimates are biased, notably because of the noise and at small sizes. Second, the sample size must be large to push the statistical errors down below the *Euclid* requirements. To reach an uncertainty comparable with the ellipticity stability, we need $N \sim (2 \cdot 10^{-4})^{-2} \simeq 25 \cdot 10^6$ examples. This is a number comparable with the amount of galaxies stamps used to test the galaxy shape measurement algorithms. There is a definite need to find an evaluation procedure to assess the performance of the PSF reconstruction technique. Unbiased and accurate performance evaluation is not straightforward for PSF reconstruction.

5.3.4 Auto-encoder architectures and PSF reconstruction

We present the different regularisations and AE architectures that were tested. We train all networks with the MSE cost function, that is,

$$J_{\text{MSE}} = \frac{1}{2N} \sum^N \|\tilde{\mathbf{x}} - \mathbf{x}\|_2^2, \quad (5.16)$$

where $\tilde{\mathbf{x}}$ is the reconstructed data and \mathbf{x} the original data. We introduce the stacked residuals penalisation, which could reduce the importance of artefacts that are below the noise level of individual stellar images,

$$J_{\Sigma} = \frac{\lambda_{\Sigma}}{2} \left\| \sum^N \tilde{\mathbf{x}} - \mathbf{x} \right\|^2. \quad (5.17)$$

Another penalisation term would enforce a smooth representation (SR) in the coordinate frame. The aim of this term is to ensure that we can interpolate smooth coefficients in the image coordinates,

$$J_{\text{SR}} = \lambda_{\text{SR}} \sum_i \sum_j w(\mathbf{x}_i, \mathbf{x}_j) \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2, \quad (5.18)$$

where $f(\mathbf{x})$ is the encoded representation. The weights w are defined using a Gaussian kernel,

$$w(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{Z} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}\right), \quad (5.19)$$

with a normalisation constant, Z , and the width of the Gaussian, σ . Another regularisation term is the positivity constraint on the reconstruction image: all pixels must be positive. In the description of the AEs, we will declare what regularisations terms are used. The classical L^1 and L^2 regularisations can also be added to the cost function. The activations functions however, are always the same. We use exponential linear units (ELU) activation function, that speeds up computations, alleviates the vanishing gradient problem and improves learning (Clevert et al. 2015),

$$\sigma(x) = \begin{cases} x, & \text{if } x > 0 \\ \exp(x) - 1, & \text{otherwise} \end{cases}. \quad (5.20)$$

All networks were implemented in TensorFlow¹³ (Abadi et al. 2015), and trained in \mathcal{O} (minutes).

The simplest AE, AE-8, has an architecture $2304 \rightarrow 64 \rightarrow 8 \rightarrow 64 \rightarrow 2304$ trained with a stacked residuals term J_{Σ} , $\lambda_{\Sigma} = 0.1$. This AE is tested on dataset A (no SED, no dither; Tab. 5.2), and compared to PCA, with sixteen components, PCA₁₆. For all S/N levels (no noise, S/N= 100, 20), we test the pixel-to-pixel reconstruction error that we denote L_{pp}^2 , which is MSE, the AE cost function, without regularisation terms. The quantities L_{pp}^2 evaluated on both methods are similar, and the reconstruction errors are below the noise level. The coefficient maps, i.e. the variation of coefficient with detector coordinates, are different between PCA and AE. When there is no noise in the input images, AE maps show simple, almost linear, gradients, see Fig. 5.26. This suggests that the interpolation of the coefficients is easier for AE than PCA. When the noise is present, the coefficient maps become extremely noisy, and AE interpolation yields absurd results.

To address this issue of noisy maps, we propose to use asymmetric contractive AEs (aCAEs). The asymmetry is introduced by adding two input features: the coordinates of the images. Reducing the number of free parameters of the representations, the number of coefficients, decrease the dilution of the spatial information. The bottleneck layer in aCAE, therefore only contains four components. The architecture is now $2304 + 2 \rightarrow 64 \rightarrow 4 \rightarrow 64 \rightarrow 2304$. The contractive loss is computed with respect to the PSF image pixels, not the spatial coordinates.

¹³<https://www.tensorflow.org>

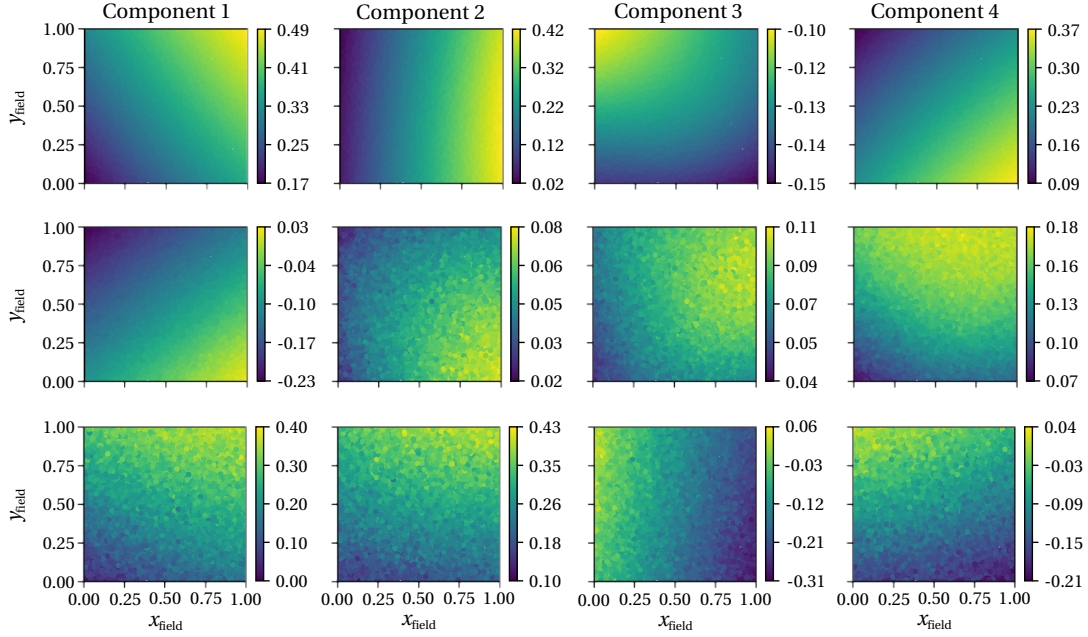


Figure 5.26 – aCAE component maps on the PSF field for dataset A for different PSFs S/N. The colour bars show the value of the aCAE representation. *From top to bottom*. Noiseless images, S/N = 100, S/N = 20. Components maps get noisier as the S/N decreases. Contrary to principal component analysis, components are not ordered. Features of the maps are different as the training is stochastic. Each instance of the network learns a different mapping.

Table 5.3 – Value of the Q metric for PSF reconstruction with aCAEs and order four polynomial interpolation on dataset A. Best values are shown in bold. No denotes no noise in the images.

Q_A		Test on levels		
		No	S/N = 100	S/N = 20
Train on levels	No	1967	1117	126
	S/N = 100	1895	1809	615
	S/N = 20	1929	1835	1294

Contractive regularisation terms reduces the sensitivity of the representation to small variations to the input (see Sect. 4.5.3). We also remove the stacked residuals penalisation in the cost function. The cost function can be minimised by concentrating the errors on a single pixel, while the rest of the image of the image is well reconstructed. The errors did concentrate on the central pixel. This might be mitigated by an additional flux conservation constraint. Adding terms increase the computation time and the complexity of training. In an effort to stay simple, we sided for removing the stacked residual term and explore aCAEs. The coefficient maps of the aCAE are smooth, and the gradients are at high signal-to-noise, allowing a simple order four polynomial to reach high Q values, see Tab. 5.3. Training the AE at low S/N does not impair its performance when reconstructing higher S/N images. AEs trained on images at S/N = 20 score high Q values at all S/N levels. It is not the case for networks trained at high S/N and tested on low S/N: the inclusion of noise in the training set seems to be beneficial, and to

increase the robustness of the network. For comparison, a PCA encoding paired with any kind of interpolation scheme scores values larger than five hundreds only when trained and tested at the same level of noise, i.e. only for the diagonal of Tab. 5.3. This would suggest that the quality of the PSF reconstruction with PCA depends on the distribution of S/N in the training set. This is not the case for auto-encoders. The positivity constraint partially controls the level of denoising in the reconstructed images. This is linked to the vanishing gradient problem, to the choice of activation function and to back-propagation. Output pixels set whose values are below zero are set at zero. The ELU (and the ReLU) returns $\sigma(0) = 0$, these output pixels are effectively stuck at zero indefinitely. Instead of enforcing positivity, we introduce a clipping threshold hyperparameter, $\delta < 0$, such that we clip all values $f(x) < \delta \mapsto \delta$. The choice of the threshold impacts the denoising.

Table 5.4 – Value of the Q metric for PSF reconstruction with Conv+aCAEs and order four polynomial interpolation on dataset B. Best values are shown in bold. No denotes no noise in the images.

Q_B		Test on levels		
		No	S/N = 100	S/N = 20
Train on levels	No	1959	1892	978
	S/N = 100	1826	1778	755
	S/N = 20	1963	1953	1934

In dataset B, the sub-pixel dither is added. AEs learn the mapping dithered PSF image to the dithered PSF representation. Dithering is not a property of the PSF, but a random instrumental effect, with no spatial correlation. Dithered PSF thus generate noisy coefficient maps. We aim at learning a mapping from dithered PSF images to a centred PSF representation. To that end, two convolutional layers precede the aCAE. The convolution kernels are learned during training, and are of dimension 3×3 . The architecture is: stamp \rightarrow conv \rightarrow conv \rightarrow 2304 + 2 \rightarrow 64 \rightarrow 4 \rightarrow 64 \rightarrow 2304 Reconstructing individual, undithered, PSF images with Conv+aCAE is at the same level as with dataset A. The residuals are at the level of the noise. PCA stacked residuals show a strong signal at the peak of the PSF, because it cannot handle dithering. Resulting coefficients maps are smooth and as simple as for dataset A, which allow to reach similarly good Q values (Tab. 5.4). The best PSF reconstruction network is trained at the lowest S/N level, consolidating the conclusions drawn from the experiments on dataset A.

Table 5.5 – Value of the Q metric for PSF reconstruction with Conv+aCAEs and order four polynomial interpolation on dataset C. Best values are shown in bold. No denotes no noise in the images.

Q_C		Test on levels		
		No	S/N = 100	S/N = 20
Train on levels	No	81	82	78
	S/N = 100	1745	1789	1756
	S/N = 20	1777	1784	1765

We turn to wavelength-dependent PSF reconstruction with dataset C. The encoding part of the PSF reconstruction algorithm does not need the SED information to reconstruct the images: stamp \rightarrow conv \rightarrow conv \rightarrow 2304 + $(x, y) \rightarrow$ 64 \rightarrow 4 \rightarrow 64 \rightarrow 2304.

We use quantity representing the SED of the star for the interpolation. This quantity is the effective temperature of the star. The Q results are shown in Tab. 5.5. Performances for the noisy datasets are high. When no noise is present, the performance is extremely poor. Conv+aCAEs seem to overfit the noiseless data. In-depth investigations into this issue are yet to be carried out at the time of writing. The cost function of the aCAEs contains a contractive term that is design to handle noise. This term is superfluous for noiseless data, or could even prove to be counter-productive as the results in Tab. 5.5 suggest. This is a secondary issue however, since the scheme is performing well on noisy datasets. The limitations encountered here could hint at a need for more sophisticated interpolation techniques, like ANNs.

Finally, we test the approach on the *Euclid*-like dataset D, which contains the same complexity as dataset C, plus typical distributions of S/N and SEDs. The architecture of the networks kept the same as for dataset C, because of the constant complexity of the problem. Results translate into a performance of $Q \approx 1900$. This seems to indicate that a typical *Euclid* distribution of S/N is not detrimental to training. A few component maps are shown in Fig. 5.27. It also suggests that the breakdown of the stellar population per S/N (Fig. 5.23) does not negatively impact the results. We note that the range of spectral interpolation is somewhat narrower than in dataset C. In that dataset, spectral types span O to M uniformly, whereas dataset D only includes stars F to M, as any other type is negligible. There are two caveats to this promising result: (i) the number of stars from which to construct the PSF is over-estimated, and (ii) training and test samples are drawn from the same model of PSF, and same distributions of stellar parameters. The first can be alleviated by including the time component, or simply reducing the number of stars available. The importance of the second could be explored by varying simulations parameters, and studying the sensitivity of the method. The preliminary results obtained so far do not exhibit clear sensitivity trends. The uncertainty on the sensitivity is very large due to the low number of PSF test samples.

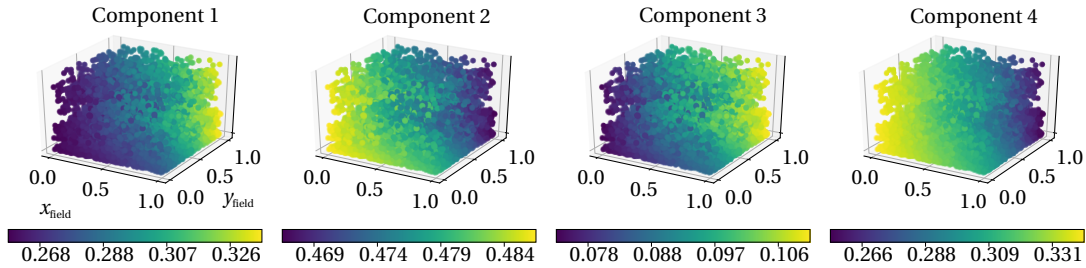


Figure 5.27 – aCAE component maps on the PSF field for dataset D. The third dimension shows the wavelength dependency. Viewing angles are chosen to highlight the smoothness of the representation. The colour bars encode the value of the aCAE components.

5.3.5 Development outlooks

In the above section, we have shown in proof-of-concept conditions, that state-of-the-art auto-encoders can reach *Euclid* targets in terms of PSF reconstruction. Asymmetric inputs paired with convolutional layers allowed to learn a representation of the PSF that can easily be interpolated. Interpolating is supported by the fact that coefficients maps look continuous. While resulting in good performances, the theoretical grounds for interpolating code representations are thin, and future studies should ensure that it is indeed allowed. Variational auto-encoders (VAEs) are showing promising results in the field of generative modelling (see [Doersch 2016](#), and references therein), and a theoretical lead could come from that subfield. VAEs were briefly considered in this study, but their sampling is Gaussian most of the time. This Gaussian prior is incorrect for our application, and finding a prior for each coefficient that varies with position and SED is difficult.

In this study, we exaggerated the number of stars available to reconstruct the PSF. In a realistic setting, there would be a factor five less objects from which to derive a model. The interpolation space of the coefficients can be extended to include the date at which the image was taken. The number of objects would therefore be increased a lot, to the cost of having to learn the PSF ageing effects. Hyperparameters were not systematically optimised, and performance gains could result from such time-consuming activities. The few trials carried out seem, however, to indicate a low sensitivity to fine-tuning. The method is still under development, the code being publicly accessible¹⁴, but could easily be integrated to a real-data pipeline.

At the time of writing, a research paper is being prepared to present the above result to the community. An important step forward will be comparison with competitive methods. Performance comparisons with simple PCA-derived methods show a clear advantage in favour of our method. It will be very interesting to test PSF reconstruction techniques against each others. We are planning to use a large number of test PSF stamps to assess the sensitivity of the method.

The *Euclid* PSF community should work towards producing a translation of the PSF requirements that can efficiently evaluated. With the need of exquisite PSF reconstruction, comes the need for an exquisite PSF reconstruction evaluation tool, which is still elusive. One way to evaluate the effect of non-ideal PSF reconstruction would be to propagate the PSF errors to shear measurement errors and potentially all the way to cosmological parameters.

¹⁴<https://github.com/rameyer/PIANNO>

5.4 A machine-learning technique for shape measurements

In this section, we present a machine-learning approach to galaxy shape measurement. We introduced traditional approaches in Sect. 3.2. In stage IV experiments, like the *Euclid* mission, many previously neglected bias sources become important, for example a wavelength dependency or the binary-induced bias. These new biases must be accounted for and calibrated out. The development of a comprehensive mathematical framework is a complicated and lengthy task. We propose to use a machine-learning technique to yield the accurate measurements necessary for the weak lensing analysis. We want to teach a calibration to the networks, that is an inverse regression from observed quantities to the shear. Machine-learning must be taught how to minimise biases by choosing a suitable cost function and designing a good training set.

As noted by [Hoekstra et al. \(2017\)](#), a shape measurement technique must minimise the sensitivity of its estimates to galaxy parameters. This means that $|\partial\mu/\partial p|$ should be small; where μ is the multiplicative bias and p is any parameter, like the S/N, or the size of galaxies. When both the overall bias and the sensitivity are low, predictions are conditionally unbiased. Minimising conditional biases across all parameters reduces the importance of the training sample distributions. We strive to devise a scheme that achieves low conditional biases. The volume of data generated by stage IV requires a fast processing of galaxy images. Machine-learning techniques are orders of magnitude faster than forward-fitting methods.

Our approach can predict galaxy parameters, like ellipticities, or the directly the shear. Input features are chosen manually and should provide the most informative data to the networks. The MegaLUT scheme also provides a learned weighting scheme to optimally predict the shear. We choose quantities derived from adaptive moment measurements of the light profile as default features. We will demonstrate that this approach can achieve state-of-the-art accuracy on different controlled experiments. We will also highlight that simple networks are sufficient.

We start by describing the concept of our method (sections 5.4.1 and 5.4.2), in particular how we build an accurate regression of noisy features (sections 5.4.3 and 5.4.4). We then lay out our strategies to predict shear, a weighting scheme or galaxy shape parameters in Sect. 5.4.5. After outlining the data simulation (Sect. 5.4.6), we summarise the main results. We end by reproducing the paper presenting the approach (Sect. 5.4.8) and conclude in Sect. 5.4.9.

5.4.1 Early phases of the technique

The idea for this method emerged in the context of the GREAT10 challenge. At that time, it was based on a large look-up table (LUT) of galaxy properties, hence the name MegaLUT ([Tewes et al. 2012](#)). This early version searched for the corrections to shape measurements in a four-dimensional LUT. As presented in Fig. 5.28, the coordinates of its parameter space spanned (i) the elongation of the galaxy, and (ii) of its PSF, (iii) the size ratio galaxy to PSF, and (iv) the orientation of the galaxy relative to its PSF. The LUT was learned from simulated training samples. Training images included realistic noise. Corrections were interpolated

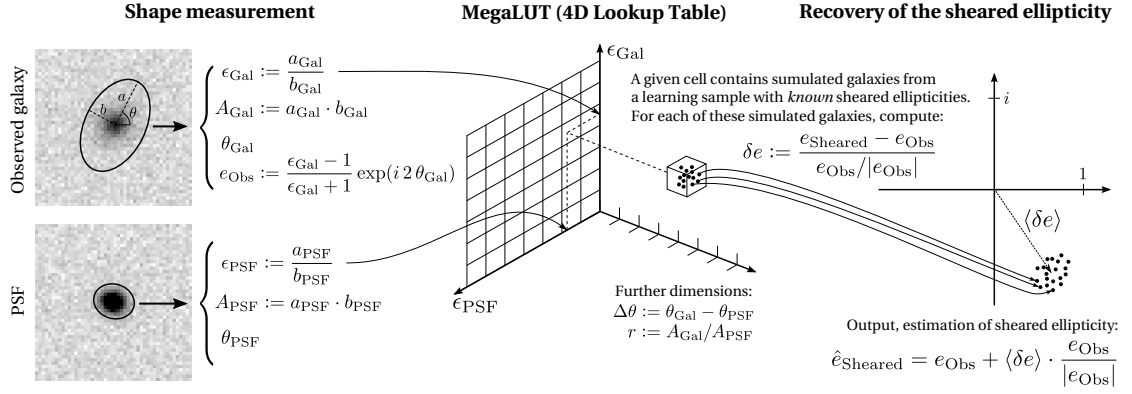


Figure 5.28 – Conceptual structure of the early MegaLUT scheme based on the query of a large lookup table. Adapted from [Tewes et al. \(2012\)](#).

from the training sample. The LUT was divided into hyper-cells of constant corrections. This empirical and conceptually simple scheme did not rely on a specific shape measurement technique. Input features can, *a priori*, be computed by any technique. In the application to MegaLUT to the GREAT10 competition, shape measurements were provided by SExtractor. About nine million galaxies were included in the training sample and divided into 160 000 cells. Competitive performance was achieved, which motivated the development of this approach.

The early MegaLUT suffered from the curse of dimensionality. Increasing the number of input features would require an exponential increase in the training examples. The LUT approach was unsustainable, and yet it showed that a simple data-driven approach could yield accurate estimates of the shape from noisy inputs features. Replacing the LUT by an ANN was envisioned in [Tewes et al. \(2012\)](#), influenced by the [Gruen et al. \(2010\)](#) paper. [Gruen et al. \(2010\)](#) showed that, in the context of weak lensing, accurate predictions of shape parameters could be predicted by networks fed with noisy input features by adapting the cost function.

The next iteration of MegaLUT was spurred by the GREAT3 challenge, and its development has continued since. The author of this thesis joined the team at that point, in late 2013. Based on the lessons learned during the GREAT10 campaign, we implemented an ANN approach. While this version is significantly different than the current implementation, it further showed the potential of the machine-learning approach.

5.4.2 Overview of the measurement concept

Before detailing the different steps, we give an overview of our approach. The concept of MegaLUT is presented in Fig. 5.29. Images of galaxy profiles are measured by the GalSim implementation of adaptive second order moments, and transcribed into a catalogue. Catalogues of the observations contain at least the features necessary for the network’s inputs. In a full-complexity implementation, the number of input features can be estimated to $\mathcal{O}(10)$. In the controlled experiments carried out in this thesis, the number of input features is typically 3 – 5.

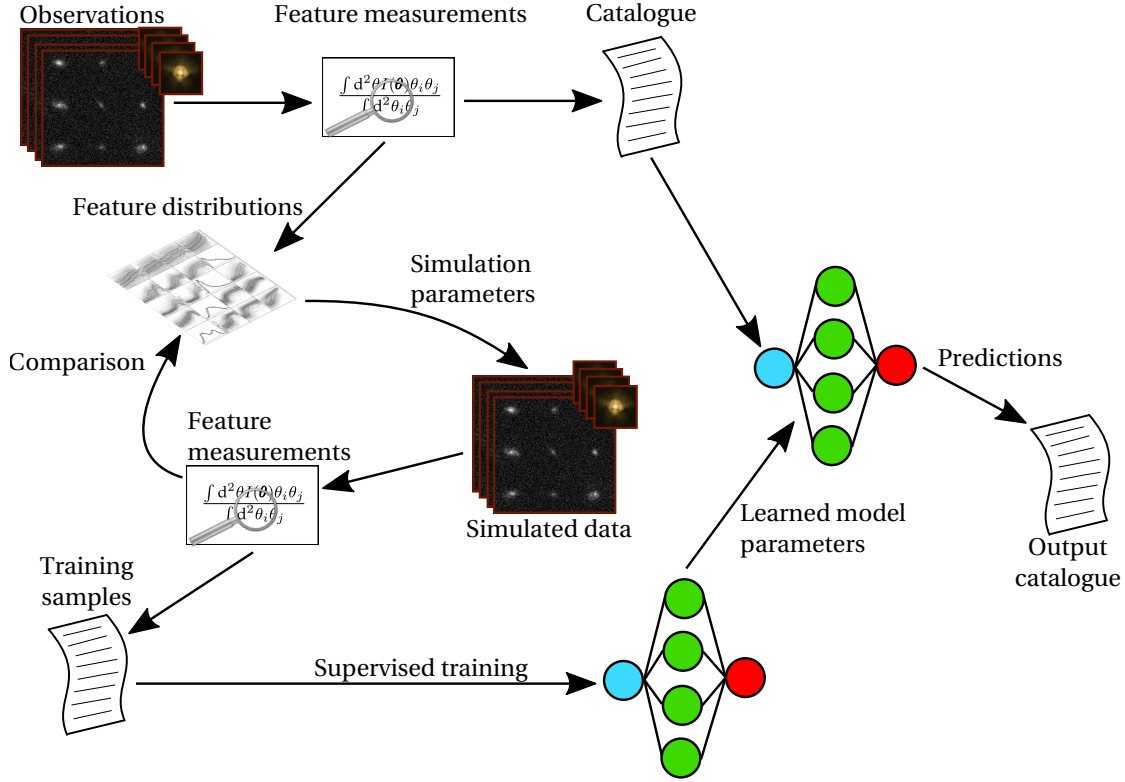


Figure 5.29 – Concept of the current implementation.

The training data are generated such that they cover the same regions of parameter space as the observations. We stress that we want the estimators to be as insensitive as possible to the galaxy properties distributions. We will gauge the sensitivity of the performance to the simulation parameter distributions in subsequent discussions. Training catalogues are built from the feature measurements of the simulated data. True ellipticities and shear values of galaxy images are used as ground truth. Simple models typically entail a two-layered network each with three to ten neurons. While this simplicity in the model could in principle limit the quality of the prediction, it does reduce the training time, and shows a relative diagnostic ease. Models parameters are stored and used to predict the ellipticity or shear of the observations.

In this thesis, the training and the validation datasets are simulated. Having large datasets is beneficial to generalise well from training, but also to reduce the uncertainty on the remaining multiplicative and additive biases. The number of training samples stays relatively low, of the order of the hundreds of thousand or millions. This number of examples is of the same order of magnitude than in the original LUT approach. The current performance, however, is significantly better. Test samples usually numbers in the tens of millions of galaxies.

We note that this implementation, while demonstrating good performance in controlled settings, is not readily integrable to a complete shear measurement pipeline. We do not claim that performances similar to what is presented in the following could be reproduced without

further development on real images. This warning aside, the method is very promising.

5.4.3 Designing an accurate regression estimator in the presence of feature noise

We use an ensemble of networks based on multilayer perceptrons. In this work, we make use of their regression capabilities. We set the activation function, $\sigma(z)$, of the hidden layers to be the hyperbolic tangent, $\sigma(z) = \tanh(z)$. The output layer of the networks, is activated by a linear function, $\sigma(z) = z$. The parameters of the networks are initialized at random, following a centered normal distribution of standard deviation of 0.1. We use the Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) in its `scipy` implementation (Nocedal and Wright 2006, and references therein).

We use the terminology *case* when referring to a single choice of values for the true parameters. This corresponds, in the context of shape measurement, to one particular training galaxy, and in the context of shear measurement to one particular value of shear. Objects in a case share the same ground truth value. The training data consists of n_{case} cases, each of these cases has n_{rea} *realisations*. A realisation is a single observation. Two realisations of a galaxy differ in sub-pixel dither and noise. Measurements on the galaxy profiles are different from one realisation to another. These shape measurements are the input features of the networks. Ergo, input features are noisy: for the same ground truth, there is a distribution of input features. The test set contains only one realisation of the observation. For training and test data sets, we may draw cases which contain the same galaxy viewed with different position angle. This shape noise cancellation (SNC) is introduced to beat down the shape noise, which arises from an insufficient sample of randomly-oriented galaxies (for more on this, see, e.g., Massey et al. 2007a).

We want to design a unbiased estimator in the sense that the error, averaged over many realisations, tends to zero,

$$\lim_{n_{\text{rea}} \rightarrow \infty} \frac{1}{n_{\text{rea}}} \sum_j^{n_{\text{rea}}} \hat{p}_j - p_{\text{true}} \longrightarrow 0, \quad (5.21)$$

where \hat{p} is one estimate for the parameter p , over a population of n_{rea} realisations. We note that this does not imply that errors for individual predictions should be small. The prediction ensemble for a case should be unbiased with respect with the value of the true parameter. The distribution of realisations within each case has an impact on the prediction. The distribution of cases should not influence the overall performance of the ANNs. We now translate the requirement for an accurate estimator into a cost function.

Let o_{jk} be the output of the neural network for the realisation j , and the test case k . The cost function compares these estimates to the target true values t_k of the case k . A typical cost

function in neural networks, is the mean square error,

$$J_{\text{MSE}} \equiv \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \frac{1}{n_{\text{rea}}} \sum_{j=1}^{n_{\text{rea}}} (o_{jk} - t_k)^2, \quad (5.22)$$

which measures the euclidean distance between each prediction and its ground truth. This expression does not make the distinction introduced above between cases and realisations. To correct this, we introduce the mean square bias cost function,

$$J_{\text{MSB}} \equiv \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left(\frac{1}{n_{\text{rea}}} \sum_{j=1}^{n_{\text{rea}}} o_{jk} - t_k \right)^2. \quad (5.23)$$

[Gruen et al. \(2010\)](#) introduced a similar cost function and obtained promising results. The mean square bias (MSB) cost function is similar to the denoising AE cost function eq. (4.23), however dAE were shown only one realisation at a time, here we directly average over the whole population. This cost function imposes that the predictions averaged over the realisations be unbiased. The MSB cost function does not penalise scatter: the network could trade precision for accuracy. The scatter in the predictions could in principle become very large such that the bias measure be artificially small. We have, however, not encountered this extreme trade-off. Throughout this study, we work with shallow and narrow networks. We argue that the limited capacity of our networks prevents the precision blow up. The precision can be increased by the introduction of a weighting scheme when computing the shear.

5.4.4 Input features and shape measurement

We prepare the set of features manually. While the choice of method is *a priori* arbitrary, it is clear that methods achieving lower biases and higher precision are better. Calibration schemes have been developed over decades of shape measurement research (e.g. [Bernstein and Jarvis 2002](#); [Hirata and Seljak 2003](#); [Jee et al. 2016](#); [Huff and Mandelbaum 2017](#); [Fenech Conti et al. 2017](#)). When studying these calibration schemes, it natural to include features such as the size. We therefore prepare an input feature set that comprises biased estimates of the galaxy properties which are used in those calibration schemes.

We choose to work with adaptive second order moments as implement in the HSM method of GalSim ([Bernstein and Jarvis 2002](#); [Hirata and Seljak 2003](#); [Rowe et al. 2015](#)). This method is based on computing second order moments (see eq. 3.1) weighted by a Gaussian function. It is called adaptive because the parameters of two-dimensional weighting Gaussian function are determined iteratively. In the following, we will use mathematical symbol when dealing with true values and represent the measured quantity by a textual symbol with a prefix relating to the measurement scheme. For example, when discussing the true input flux, we use F , but `adamom_flux` when using the measured flux. The `adamom` prefix denotes the use of the GalSim adaptive moments. In most instances, we will use a combination of the following measurements:

5.4. A machine-learning technique for shape measurements

`adamom_g1/2`: the ellipticity component 1 or 2, from the second order moments (see eq. 3.2);

`adamom_sigma`: an estimate of the size, $\sigma = (\det Q)^{1/4}$, where Q is the matrix of the second order moment. This parameter is linked to the half-light radius, R_{hl} , of a Gaussian profile by $R_{hl} = \sqrt{2 \ln 2} \cdot \sigma \simeq 1.1774 \sigma$;

`adamom_rho4`: the weighted fourth moment of the image, ρ_4 . This parameters estimates the kurtosis of the light profile;

`adamom_flux`: the flux is in ADU in the stamp, $F = \int w(\boldsymbol{\theta}) I(\boldsymbol{\theta}) d\boldsymbol{\theta}$, that is the 0th order moment.

In this study, we carefully define the signal-to-noise ratio. As the number of galaxies increases with decreasing S/N, it is beneficial for weak lensing studies to include as low signal galaxies as possible. However, there are many different definitions of the S/N in the literature. We interpret the *Euclid* definition of S/N (Laureijs et al. 2011; Cropper et al. 2016) as

$$S/N \equiv \frac{g \cdot \text{adamom_flux}}{\sqrt{g \cdot \text{adamom_flux} + A_{\text{eff}} \cdot (g \cdot \sigma_{\text{sky}})^2}}, \quad (5.24)$$

where

$$A_{\text{eff}} \equiv \pi \cdot \left(3 \cdot \text{adamom_sigma} \cdot \sqrt{2 \cdot \ln(2)} \right)^2, \quad (5.25)$$

and g is the gain in electrons per ADU. This measure is remarkably close to the widely used SExtractor S/N (See Fig. 5.30), computed as

$$S/N|_{\text{SE}} = \text{FLUX_AUTO} / \text{FLUXERR_AUTO}. \quad (5.26)$$

SExtractor S/Ns are easily obtained from observations, including in weak lensing analyses (e.g. DES, Hildebrandt et al. 2017), hence their popularity. When simulated images are drawn purely with Gaussian noise, they mimic sky-limited observation. This is equivalent to setting a very high value for the gain in eq. (5.24), yielding the approximation,

$$S/N|_{\text{Gaussian}} \equiv \frac{\text{adamom_flux}}{\sqrt{A_{\text{eff}} \cdot \sigma_{\text{sky}}^2}}. \quad (5.27)$$

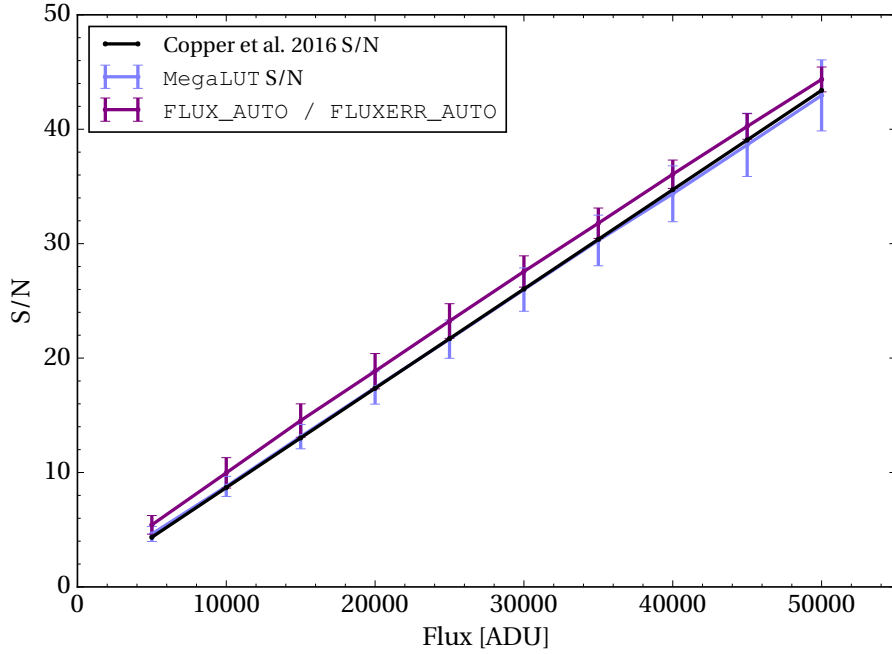


Figure 5.30 – Comparison of different S/N estimates on Gaussian galaxies convolved with Gaussian PSFs with a *Euclid*-like FWHM. The galaxies have an observed half-light diameter of 0.43 arcsec, as per the [Cropper et al. \(2016\)](#) definition.

5.4.5 Methods for shear estimations

Predicting shear estimates rather than ellipticities has one major advantage: the concept of ellipticity does not need to be defined. This quantity may not be possible to define with ambiguous or real galaxy profiles. Note that an ellipticity measure can always be computed from the second order moments of the light profile. In this section, we describe our strategies to predict shear estimates and their associated weights. We can also measure other galaxy parameters with the same conceptual approach. Networks that predict shear or galaxy parameter estimates share the same objective, because we want the predictions to be accurate. The main differences are to be found in the training set. We build training sets to teach networks about the invariances in the data. Different targets (shear, weights or galaxy properties) need different architectures of the training sets. Number choices for the number of cases, realisations and SNC will be different according to the nature of the target.

Predicting shear estimates

Networks that predict shear should be insensitive to the intrinsic shape. The ground truth is the value of the shear. To teach invariance to galaxy shape, each case contains the same profile with many orientations, that is with a high SNC. This entails showing many galaxies with the same profile, but with different true orientation, noise, and sub-pixel positions. NNs estimate a shear for each galaxy, $\hat{g}_{i,jk}$, where i is the shear component, j is the realisation identifier

and k the case number. When the MSB cost function, J , is evaluated the bias is computed for each case k ,

$$J_{\text{MSB}} = \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left(\frac{1}{n_{\text{rea}}} \sum_{j=1}^{n_{\text{rea}}} \hat{g}_{jk} - g_k^{\text{true}} \right)^2, \quad (5.28)$$

where we applied the newly-defined notation to eq. (5.23). A typical training set is shown in Fig. 5.31, left panel. There is no simple relationship between a feature and the shear. Networks must use all of the information in the cases to derive a mapping between the measured galaxy properties and the shear.

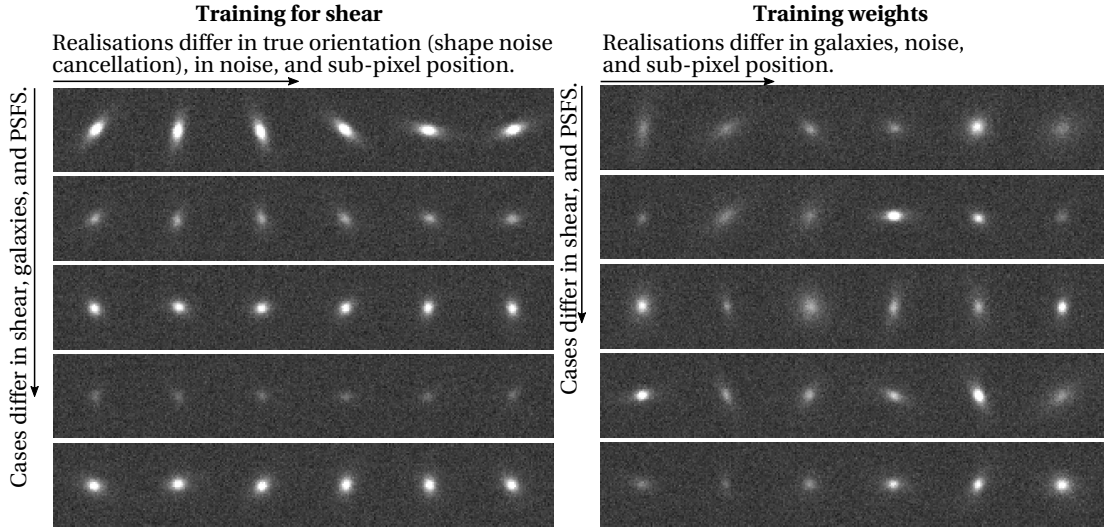


Figure 5.31 – (Left.) Structure of a training set to train a *shear* estimator. (Right.) Structure of a training set to train a *weight* estimator. All galaxies of a case are simulated with the same true shear, but note that typical cosmic shear is too small to be seen in this figure.

Weighting scheme

We designed an accurate estimator of shear in the previous section. However, there is no constraint on the precision of the predictions. The precision is sensitive to galaxy properties. Shear estimates are computed from an observed population of galaxies. Applying a selection function to a population measured with varying precision can result in significant biases, even when individual point estimates are accurate. There are galaxies for which the neural networks trained to predict the shear will inevitably fail, for example because they are barely resolved. The purpose of applying weights, w_i , to the n_{rea} point estimates is to reduce this selection bias (e.g. Kaiser et al. 2000; Bernstein and Jarvis 2002; Fenech Conti et al. 2017). The observed

shear, \hat{g}_{obs} , can be estimated by a weighted average of the shear point estimates, \hat{g}_j ,

$$\hat{g}_{\text{obs}} = \frac{\sum_j^{n_{\text{rea}}} \hat{g}_j \cdot w_j}{\sum_j^{n_{\text{rea}}} w_j}. \quad (5.29)$$

We train networks to predict the weights such that the bias over the observed shear, \hat{g}_{obs} , is minimised. The cost function reads

$$J_{\text{MSWB}} \equiv \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left(\frac{\sum_{j=1}^{n_{\text{rea}}} \hat{g}_{jk} \cdot w_{jk}}{\sum_{j=1}^{n_{\text{rea}}} w_{jk}} - g_k^{\text{true}} \right)^2. \quad (5.30)$$

We note that for this mean square weighted bias (MSWB) cost function to be optimised, we need to feed the predicted value of the unweighted shear, \hat{g} , irrespective of its inclusion in the feature set. The unweighted shear, \hat{g} , is an auxiliary feature: it is typically not included in the feature set, but is necessary for the computation of the cost function. Weights do not depend on the shear, but rather on galaxy properties. Qualitatively, the function to be learned is a cut-off for low S/N or poorly measured galaxies: the network should assign a weight of 0 to low S/N and poorly-measured galaxies and close to 1 to reliably measured galaxies. In weight-predicting networks, we use a variant of the sigmoid function as output activation function, $\sigma(z) = 1/(1 + \exp(-4z))$.

The training set must contain the selection function of the survey. This implies that a case is made up of many different galaxy profiles, with only one realisation each. The SNC should not be included here, otherwise the networks will learn to assign large weights only to bright and large galaxies. They could incorrectly learn that the shear can be measured from the brightest galaxies only. Such a training set, illustrated in Fig. 5.31, right panel, enables the networks to learn a well-controlled selection function: to down-weight faint galaxies and cancel any biases due to the original selection function.

Measuring galaxy parameters

Predicting accurate galaxy parameters is similar to the shear strategy. Contrary to predicting shear or their associated weights, the networks predict galaxy parameters with very informative input features: features `adamom_g1` and `adamom_g2` directly estimate the galaxy ellipticity. Galaxy size is linked to `adamom_sigma`. The networks need to calibrate the remaining biases. These can depend on many other parameters of the light profile, including the concentration or the flux. Networks must be trained with different cases, grouping galaxy properties. Each of the case must have enough realisations to beat feature noise. An example of a training set for ellipticity is shown in Fig. 5.32. The cost function to be optimised is similar to eq. (5.28),

$$J_{\text{MSB}} = \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left(\frac{1}{n_{\text{rea}}} \sum_{j=1}^{n_{\text{rea}}} \hat{p}_{jk} - p_k^{\text{true}} \right)^2, \quad (5.31)$$

where p is the parameter to be predicted.

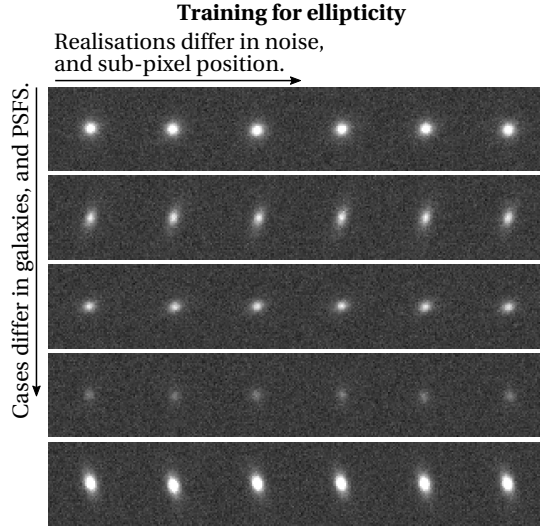


Figure 5.32 – Structure of a training set to train a *ellipticity* estimator.

5.4.6 Galaxy image and PSF simulations

All of the training simulations and most of the test galaxies were simulated with a single Sérsic profile. This profile was introduced by Sérsic (1963) and is a generalisation of an exponential profile. The surface brightness $I(r)$ of an axisymmetric profile varies as

$$I(r) = \frac{F}{2n\pi\Gamma(2n)r_0^2} \exp\left\{-(r/r_0)^{1/n}\right\}, \quad (5.32)$$

where F is the total flux, n is the Sérsic index, $\Gamma(\cdot)$ the Gamma function, and r_0 a scale radius.

Galaxy parameters, like its flux or size, are drawn from distributions that depend on the experiments. Associated PSFs are included when simulating images, during the `GalSim` image process. We convolve the galaxies with their PSF to yield the observed profiles. We will apply the scheme to three different datasets. Details of the distributions are shown in Paper IV.

First, a fiducial dataset with simplistic distribution of parameters will be used to demonstrate the different strategies. The distributions of Sérsic indices, n , half-light radii, R_{hl} , and surface brightness, S_B are drawn from uniform distributions. The flux is computed from the surface brightness as $F = \pi R_{hl}^2 \cdot 10^{-S_B}$. Ellipticities are drawn from Rayleigh distributions, which were not finely tuned to mimic any observed distribution. We inject sky-limited noise, neglecting Poisson noise. Stationary PSFs are simple Gaussians. We demonstrate the following processes:

1. predict accurate point-estimates for ellipticities;
2. produce shear estimates;
3. compute associated weights;
4. and account for spatially variable PSFs.

PSF profiles in variable PSF experiments are still Gaussian, but varying in size and ellipticities according to the *Euclid* requirements (Laureijs et al. 2011).

Second, we generate an *Euclid*-like dataset by simulating galaxies observed by the Galaxy Evolution from Morphologies and SEDs (GEMS) survey (Rix et al. 2004) and using *Euclid*-like image characteristics. This dataset, while still simple, allows to gauge the performance of the method on data comparable to what can be expected from the *Euclid* survey, both in terms of predicting weighted shear and galaxy ellipticities.

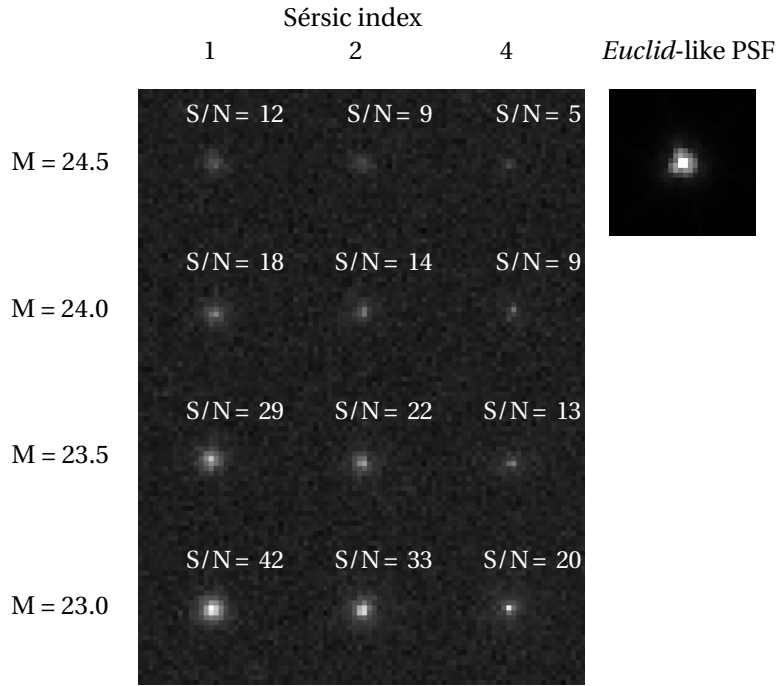


Figure 5.33 – Galaxies with the same observed size, but at different magnitudes and Sérsic indices. The observed size is a half light diameter of $0.43''$. Values of the S/N are averaged on many realisations of the (Gaussian and Poisson) noise. The profiles are convolved with the PSF shown in the top right panel. We show only the 30 central pixels for all stamps.

Distributions of galaxy parameters are fitted to the GEMS observations following the work of Hoekstra et al. (2017) and the PSF is space-like, simulated with GalSim (see Paper IV for details). We carefully select simulations parameters to produce image characteristics similar to VIS-like images (once again, we refer to Paper IV). Most of the parameters, like exposure time, read-out noise or sky background, can be found in the literature. The value of the zero-point, however, could not be found, such that we adjust its value to meet the *Euclid* requirement. The definition of the *Euclid* S/N is given in Cropper et al. (2016) and reproduced in substance: “a 0.43 arcsec extended source of 24.5 VIS AB mag is measured at 10σ in a 1.3 arcsec diameter aperture from three exposures of 565 seconds.” We interpret the 0.43 arcsec extension to mean a half-light diameter of $0.43''$, and measure the S/N with our observational definition of the S/N (eq. 5.24). As the *Euclid* definition of S/N does not require a extended object profile, we

will average over Sérsic indices to get the zero-point. A profile with a high Sérsic index has a high concentration such that two galaxies with the same flux and same radius will not have the same S/N, as shown in Fig. 5.33. We also average over the ellipticities, but this is a second order effect with respect to the Sérsic indices. We made the important simplifying assumptions in the *Euclid*-like dataset as follow.

- (i) There are no blends in the stamps. Galaxies are produced on a grid.
- (ii) We neglect the wavelength-dependence of the galaxy and PSF profiles.
- (iii) We produce images whose exposure time is three times the frame exposure. That is, we assume perfect frame co-addition.
- (iv) The PSF is stationary to reduce the computing time. When describing the *Euclid*-like results, we will have demonstrated our PSF correction scheme (see Sect. 5.4.7).

Third, we used the GREAT3 datasets to (i) provide comparable performance metrics, and (ii) make an order estimate of the sensitivity to more complicated profiles. We trained our networks with single Sérsic only, but applied them to the bulge plus disk Sérsic models and real HST images. We drew our own training sets, from distributions of parameters that were estimated from the data. We kept this study as blind as possible by not using any information that was not at the disposal of the participants at the time of the GREAT3 challenge. We trained networks with our training simulations to directly predict shear and its associated weight. We restrict ourselves to the branches with stable PSFs and only one exposure.

5.4.7 Summary of the results and lessons learned

The requirements for Stage IV experiments were derived by Cropper et al. (2013) and Massey et al. (2013). The *Euclid* requirements on the shape measurement accuracy are summarised as: (i) a multiplicative bias $|\mu| < 2 \cdot 10^{-4}$ and (ii) an additive bias of $|c| < 1 \cdot 10^{-3}$ (Laureijs et al. 2011; Duvet et al. 2015). We will use those next-generation targets as goals for our method.

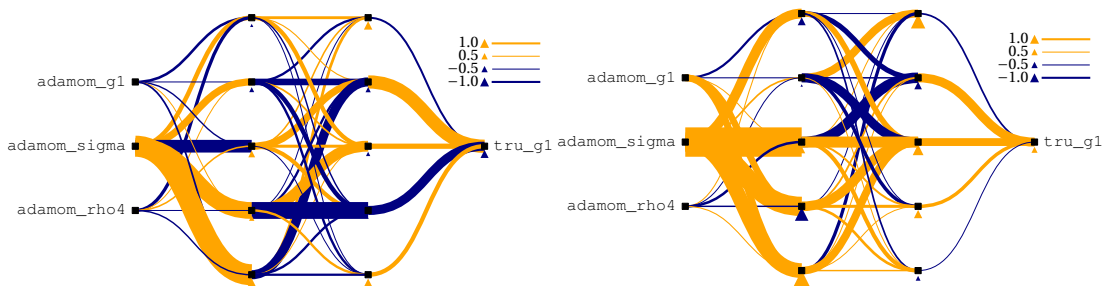


Figure 5.34 – Two examples of networks belonging to the same committee, trained on the same dataset (fiducial and noisy), and achieving similar performance.

Predicting ellipticities of galaxies to get an estimate of the shear is the conceptually simpler approach. We demonstrate in Paper IV that our method is an accurate estimator. The predictions meet the requirements, irrespective of the experiment. Conditional biases are observed, in particular in networks trained with noiseless training examples. Networks were trained with noiseless data to reduce the simulation and training times. To reduce conditional biases, we show noisy examples to the networks. Conditional biases are indeed lowered, however, the training time soars from typically a few hours to a few days. In addition, noise-trained networks are more precise, as measured by the root-mean-square deviation of the shear bias. In Fig. 5.34, we show the topography of two members of the same committee trained with noisy data in the fiducial experiment. Their performance is similar, but slightly worse than the accuracy of their committee. Studying the topography of the networks suggests that the size of the galaxy is the most important feature. This is not surprising as analytical calibration schemes depend strongly on the size of galaxies.

The importance of the galaxies parameter distribution was tested in *Euclid*-like simulations by training on GEMS or uniform distributions. In the latter set, all galaxy parameters, like size and ellipticity, are drawn from an uniform probability distribution. The two sets are very different from one another. We only made sure that the uniform distribution covers at least the same range of parameters as the *Euclid*-like test galaxies. The resulting performances are remarkably similar, suggesting that the distribution of the training cases is indeed not important. Quantifying the sensitivity to the distributions is difficult as it requires training many committees on a spectrum of distributions. However, when predictions are conditionally unbiased, the performance will not depend on the details of the distributions. Conditional biases are low in our method, also suggesting a low dependence on the parameter distributions.

Directly predicting shear point estimates is biased at the level of $\mu \sim \mathcal{O}(2\%)$ for low S/N galaxies. Shear estimates are accurate when measuring high S/N galaxies only. The weighting scheme should be introduced to extract the maximum amount of information. Weight predicting networks reproduce the expected behaviour: they down-weight low S/N galaxies, and highlight reliable galaxies, with a fast transition from small to high weights. Remaining weighted shear biases are small, in most of our experiments, they were consistent with shape measurement requirements.

In the original GREAT3 challenge, while we used a significantly different approach, we saw a hint of model bias in the real galaxies branches (Mandelbaum et al. 2015). This model bias is still present in our most recent results, at the level of $\mu \sim \mathcal{O}(0.5\%)$. As during the challenge, the networks are trained with single Sérsic profiles, and tested on real galaxy profiles. This small but significant bias could be reduced by increasing the complexity of the simulations to two-component galaxies.

Analytical PSF corrections are not applied to the features in the current implementation. Ergo, networks have to learn to correct for the PSF. When spatially-variable PSF fields are considered, corrections depend on the galaxy position in the field. Dealing with this new complexity is

fairly straightforward for the networks: we include the coordinates of the galaxy as a new input feature. These position-aware networks learn a representation of the PSF simply based on the galaxy coordinates. We tested the position-aware strategy on four different PSF fields with increasingly complex fields on the fiducial dataset, resulting in low conditional and overall biases. Networks which did not have access to the positional information exhibit a strong sensitivity to PSF parameters, multiplicative and additive.

Shallow and narrow networks are sufficient to calibrate shear measurements. The introduction of further complexity, like two-component galaxies images, or the wavelength dependence will likely increase the required network capacity. Networks architectures should however remain fairly simple, hence keeping the training and prediction to short time periods and keeping the overfitting risk low.

Even if the distribution of parameters are irrelevant for the performance, the training set should be “clean.” This entails avoiding impossible cases, like unresolved galaxies for which features measurements are meaningless or outliers, which can prevent the optimisation from converging.

5.4.8 *Paper IV: Weak-lensing shear measurement with machine learning: Teaching artificial neural networks about feature noise*

In this Paper IV, we present in detail the methodology and results of the MegaLUT approach. We begin by a description of how we teach networks about feature noise. Then, we turn to the three datasets and study the performance of the method. We found that simple and narrow networks, with the right cost function and training dataset organisation can predict accurate weighted estimates of shear and ellipticity. By extension, such approach could be used to accurately predict any galaxy parameters, like the size. We also propose possible solutions to the challenges which remain to be solved.

The author of this thesis contributed to the project since late 2013 by implementing parts of the method, carrying out many tests and writing sections of the paper. The work was hence shared with the principal author of this article, Malte Tewes. Reiko Nakajima and Frédéric Courbin advised us on important aspects of the methods and helped perfecting the concept. The paper is to be submitted to the journal *Astronomy & Astrophysics* and is reproduced below.

(See next page.)

Weak-lensing shear measurement with machine learning

Teaching artificial neural networks about feature noise

M. Tewes¹, T. Kuntzer², R. Nakajima¹, and F. Courbin²¹ Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany² Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

December 8, 2017

ABSTRACT

Cosmic shear, that is weak gravitational lensing by the large-scale matter structure of the Universe, is a primary cosmological probe for several present and upcoming surveys investigating dark matter and dark energy, such as *Euclid* or WFIRST. The probe requires an extremely accurate measurement of the shapes of millions of galaxies based on survey imaging data. Crucially, the shear measurement must address and compensate for a range of interwoven nuisance effects related to the instrument optics and detector, noisiness of the images, unknown galaxy morphologies, colours, blending of sources, and selection effects. In this paper, we explore the use of supervised machine learning as a tool to solve this inverse problem. We present a simple architecture that learns to regress shear point estimates and weights via shallow artificial neural networks. The networks are trained on simulations of the forward observing process, and take combinations of moments of the galaxy images as inputs. A challenging peculiarity of the shear measurement task, in terms of machine learning applications, is the combination of the noisiness of the input features and the requirements on the statistical accuracy of the inverse regression. To address this issue, the proposed training algorithm is designed to minimize both bias over many realizations of an observation, and sensitivity to properties of the sample of source galaxies. We first introduce key aspects of our approach using toy-model simulations, and then demonstrate its potential on more realistic images drawn to mimic *Euclid*-like data. Finally, we analyse measurements on data from the GREAT3 challenge, and obtain highly competitive multiplicative and additive shear biases despite the use of a simple training set. From these promising results we conclude that the further development of suited machine learning approaches is of high interest to meet the stringent requirements on the shear measurement in current and future surveys. We make a python implementation of our technique publicly available.

Key words. methods: data analysis – gravitational lensing: weak – cosmological parameters

1. Introduction

Images of distant galaxies appear slightly distorted, typically at the percent level, as light bundles reaching the observer are differentially deflected due to gravitational lensing by massive structures along the line of sight. As galaxies come in a variety of intrinsic shapes, inclinations, and orientations, these weak distortions are not identifiable on individual sources. In this sense, galaxies give us only a very noisy view of the distortion field. However, despite this intrinsic “shape noise”, the weak lensing (WL) effect imprints spatial correlations on the apparent galaxy shapes. Observing these spatial correlations, ideally as a function of redshift, allows us to infer properties of the large-scale matter structure of the Universe, and how this structure has grown over time.

This probe, known as cosmic shear, is one the main scientific drivers for surveys poised to explore dark matter and dark energy, such as KiDS¹ (de Jong et al. 2015), the Dark Energy Survey (DES, The Dark Energy Survey Collaboration et al. 2016), the ESA *Euclid*² mission (Laureijs et al. 2011), and NASA’s Wide Field InfraRed Survey Telescope WFIRST³. Kilbinger (2015) and Mandelbaum (2017) provide recent re-

views on the field, with a particular focus on the analysis methods to interpret the data from wide field surveys.

The statistical uncertainty of cosmic shear measurements, which is related to the finite number of galaxies probing the shear field, decreases with the increasing size of the surveys. To make full use of large surveys, the accuracy of the data analysis methods must therefore be high enough to avoid that systematic errors dominate the cosmological inference (Refregier 2003). For *Euclid*, surveying 15’000 square degrees of extragalactic sky, the resulting accuracy requirements are unprecedented. These requirements flow down, on the observational side, to (1) the determination of photometric redshifts and (2) the measurement of galaxy shapes. The cosmology community is working intensively on both aspects and on the required algorithmic improvements, often addressing effects that could previously be neglected due to the limited survey size.

Regarding the problem of photometric redshift determination, “empirical” and machine learning methods are now considered as at least equivalent to traditional template-fitting methods in terms of precision and accuracy. They are also complementary, as they are based on fundamentally different principles and assumptions. Several applications of artificial neural networks (NN) yield highly competitive results, especially when predicting redshift probability distributions (e.g., Bilicki et al. 2017; Bonnett 2015). Furthermore, D’Isanto & Polsterer (2017) demonstrate how deep convolutional NNs can infer redshifts by

¹ <http://kids.strw.leidenuniv.nl/>² <http://www.euclid-ec.org/>³ <http://wfirst.gsfc.nasa.gov>

directly processing multi-band image data at the pixel level, as compared to using fluxes measured in apertures.

The shear measurement problem has not yet seen a similar evolution towards machine-learning methods. The problem of shear measurement is also referred to as “shape measurement” in the literature, as the shape (more precisely the ellipticity) of galaxies yields an estimator for the lensing shear. There are two traditional categories of shear measurement techniques: (1) methods based on the measurement of weighted quadrupole moments of the observed light profile and (2) methods that forward-fit a model. Mathematically, these categories share strong similarities (Simon & Schneider 2017), and both have to tackle the same sources of biases in order to serve as accurate shear estimators.

The most prominent observational issues are the deformation of the sheared galaxy light profile by the telescope optics (often seen as convolution by a point-spread function PSF), the pixelation of the image by the detector, and the pixel noise. Information about the original galaxy shape and the shear is lost by each of these effects, while systematic errors are introduced. Even for space-based instruments, the shape of the PSF varies over the field of view and in time, and the PSF for each galaxy must therefore first be reconstructed with high fidelity. In addition, the low signal-to-noise ratio (S/N) of the galaxy images leads to biases that have to be accounted for, notably as shear estimators are not linear functions of the image pixel values (e.g. Refregier et al. 2012). A large variety of shear measurement methods have been developed to deal with these effects, notably in context of the public Shear Testing Programme (STEP) and the GRavitational lEnsing Accuracy Testing (GREAT) challenges (Heymans et al. 2006; Massey et al. 2007; Bridle et al. 2010; Kitching et al. 2012; Mandelbaum et al. 2015). Today’s state-of-the-art shape measurement methods involve various forms of simulation-based calibration to account for different biases (e.g., Fenech Conti et al. 2017; Huff & Mandelbaum 2017; Jee et al. 2016), yet without embracing a full machine-learning approach. The computational cost of the shape measurement process is also of importance, with *Euclid* set to observe about 1.5 billion galaxies. Rigorously testing a method will typically imply applying it to simulations larger than the survey itself, underlining the need for fast algorithms.

In this paper, we use supervised machine learning (ML) to address the problem of shear measurement, building upon the few previous applications of ML to this specific problem (Gruen et al. 2010; Tewes et al. 2012; Graff et al. 2013). Specifically, we simulate noisy and PSF-convolved galaxy images with known shear, and train NNs to regress shear estimates based on features of these images, *so to minimize shear prediction biases rather than shear errors*. With an earlier version of this approach, we participated in the GREAT3 challenge under the name *MegaLUT* (Mandelbaum et al. 2015). The development of a machine learning approach is motivated by

1. the unavoidable need for some form of shear calibration via image simulations, for any state-of-the-art technique,
2. the low CPU cost of machine learning predictions, as compared, e.g., to iterative forward-fitting methods (either frequentist or Bayesian),
3. the potential of simulation-driven methods to easily embrace further complex bias sources not identified at the moment, without affecting the initial formalism,
4. the possibility to control and penalize the tradeoff between *sensitivity* of the method to parameters affecting bias, and the bias itself.

A distinctive aspect of this ML application is the noisiness of the data. For the low-S/N galaxies of interest to cosmic shear studies, the unavoidable uncertainty on their ellipticity is larger than the shear distortion we wish to recover accurately. The cost function of the training algorithm and the structure of the training data must therefore be adapted so that the neural networks can learn to correct for biases resulting from the propagation of noisy inputs through them.

To ease the analysis and comparison with other methods, the present work is limited to the prediction of point estimates and weights for each component of the shear. This was also the format adopted by GREAT3 (Mandelbaum et al. 2014). The large number of source galaxies in weak lensing surveys led the community to (so far) favor these over probability distributions. Furthermore, traditional shape measurement methods only produce point estimates, and they are also easier to analyse, for example when computing correlation functions. Note that this situation is however changing, with most current methods adopting some more descriptive probabilistic formalisms (Bernstein & Armstrong 2014). We see the implementation presented in this paper as a stepping stone towards a machine-learning probabilistic approach.

This article is organised as follows: we introduce the required formalism of WL in Sect. 2. In Sect. 3, we describe how NNs achieve accurate predictions in the presence of noise in their inputs. The input features, as measured on galaxy images, are presented in Sect. 4. We then detail how we connect these steps to form a shear measurement method in Sect. 5. We demonstrate and apply our method on simple simulations in Sect. 6, on more realistic *Euclid*-like simulations in Sect. 7, and on GREAT3 data in Sect. 8. We offer perspectives in Sect. 9 and summarise in Sect. 10.

2. Formalism of weak gravitational lensing

In the following we give minimal definitions of the formalism of weak lensing and its estimation. Recent reviews include Kilbinger (2015) and Bartelmann & Maturi (2017) and a comprehensive introduction can be found in Schneider et al. (2006).

2.1. Shear and ellipticity

The weak-lensing distortion seen in a given field of view can locally be approximated as a linear transformation between the “true” unlensed coordinates and the observed coordinates, expressed by a Jacobian matrix. This local transformation is often written as

$$\begin{pmatrix} x^{\text{true}} \\ y^{\text{true}} \end{pmatrix} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \begin{pmatrix} x^{\text{obs}} \\ y^{\text{obs}} \end{pmatrix}, \quad (1)$$

where g_1 and g_2 are the two components of the (reduced) shear, causing a change in the ellipticity of observed galaxies, and where κ is the convergence, describing the change in their apparent size. It is often convenient to write the shear as a complex number $g = g_1 + ig_2$.

Most traditional methods to measure the lensing shear deal with expressions for the *ellipticity* of galaxy, as for example KSB (Kaiser et al. 1995; Hoekstra et al. 1998). Instead, the method we develop yields a direct estimator of the shear signal g as defined above, and does not require any mathematical description of the ellipticity at any stage of the forward process. Indeed, the notion of the ellipticity of a galaxy is not trivial, as real galaxies have complex morphologies without simple elliptical isophotes,

not even mentioning pixellation and noise. For idealized galaxies with elliptical isophotes, we do however define an ellipticity in the following, as we will use an ellipticity in the analysis of the sensitivity of your method and for some experiments. For such a galaxy, with semi-major axis a and semi-minor axis b , we follow the notation of the GREAT3 challenge (Mandelbaum et al. 2014) and define the ellipticity ε as a complex number of modulus $|\varepsilon| = (1 - b/a)/(1 + b/a)$ and a phase determined by the position angle ϕ of the major axis such that $\varepsilon_1 = |\varepsilon| \cos(2\phi)$ and $\varepsilon_2 = |\varepsilon| \sin(2\phi)$. With this definition, and considering only weak shear, i.e. $|g| \ll 1$, the observed ellipticity ε^{obs} of an idealized lensed galaxy is related to its true intrinsic ellipticity, $\varepsilon^{\text{true}}$, by $\varepsilon^{\text{obs}} \approx \varepsilon^{\text{true}} + g$. The average ellipticity of a group of galaxies submitted to the same shear is an unbiased estimator of the shear, $\langle \varepsilon \rangle \approx g$, under the assumption that these source galaxies are intrinsically randomly oriented, i.e. $\langle \varepsilon^{\text{true}} \rangle = 0$.

2.2. Biases and sensitivity of shear estimation

Biases of a shear estimator \hat{g} are commonly quantified using a linear bias model following Heymans et al. (2006), decomposing the bias into a multiplicative part, μ , and an additive part, c , for each component,

$$\hat{g}_i - g_i^{\text{true}} = \mu_i \cdot g_i^{\text{true}} + c_i + \text{noise}. \quad (2)$$

Given shear measurements on simulations with known true shears, estimates of these biases μ and c are obtained by fitting a line to the shear estimation residuals $\hat{g}_i - g_i^{\text{true}}$ against the true shear value. The commonly used components $i = \{1, 2\}$ of the shear and the biases μ_i and c_i are defined by the coordinate grid used in Equation 1, usually the image pixel grid. The first (second) component describes deformations along the axes (along the diagonals) of this grid. In addition, following GREAT3 conventions (Mandelbaum et al. 2015) and in line with Fenech Conti et al. (2017), we use the indices $i = \{+, \times\}$ to relate to components in a frame rotated to be aligned with the anisotropy of the PSF. The estimation of these PSF-oriented biases is done on simulations with variable orientation of the PSF⁴.

In line with the above linear bias model, the numerous *sources* for bias are also often categorized into “multiplicative” and “additive” (Mandelbaum 2017). For example, the width of the PSF and the noise in the images are sources for multiplicative bias, as both effects tend to make galaxies look rounder, i.e., less sheared (see, e.g., Melchior & Viola 2012). A typical source for an additional bias is the imperfect correction for an anisotropic PSF, leading to a net shift in the measured galaxy ellipticity. We refer to Massey et al. (2013) for a more comprehensive list of biases and studies of their propagation into cosmic shear results. Stage IV experiments require multiplicative (additive) biases and the uncertainty on these biases to be of the order of $|\mu| \lesssim 2 \cdot 10^{-3}$ ($|c| \lesssim 2 \cdot 10^{-4}$).

In this paper, we will perform evaluations of μ and c in different bins of “true” parameters potentially affecting the bias, such as the intrinsic size of the galaxies. This is made possible by carrying out numerical experiments using simulated data. It is crucial to be aware that any binning or selection according to some noisy “observed” parameters might lead to shear estimation biases due to selection effects. For example, the estimate of the size or the signal-to-noise ratio of a galaxy can in practice

⁴ More precisely, to estimate those biases, one first rotates the components 1 and 2 of \hat{g} and g^{true} by -2θ , where θ is the position angle of the PSF anisotropy, and then performs the linear regressions on these rotated components.

depend on the orientation and magnitude of the shear. For a discussion of selection biases, see, e.g., Fenech Conti et al. (2017).

As mentioned by Hoekstra et al. (2017), an important goal for a shear measurement method should be to minimize the sensitivity $|\partial\mu/\partial p|$, to any parameter, p , potentially affecting the multiplicative bias μ of a measurement. A tradeoff between this sensitivity, and the overall bias, will have to be made. Let us consider some extreme examples. Suppose that a simple given method shows a strong multiplicative bias on a given set of simulations. Applying a plain multiplicative scaling to all its shear estimates will apparently remove this overall bias. However, the sensitivity of this method to the galaxy population and simulation parameters might be increased by this rescaling. On the other hand, a method strongly driven by a prior on the galaxy profiles might show low sensitivity to the details of a galaxy population. But also, such a method would have a suppressed response to the shear affecting the observed galaxy profiles, and therefore have a large overall bias. When designing a shear measurement method, both sensitivity and integrated biases should therefore be kept under control simultaneously.

3. Accurate regressions from artificial neural networks in presence of feature noise

In this section, we describe how we train neural networks to perform accurate regressions despite noisy input features, building upon ideas from Gruen et al. (2010). We keep this part generic to any inverse problem, and will introduce the particular application to weak-lensing shear measurement in Section 5.

3.1. The inverse regression problem

A standard feedforward neural network (NN) with N input nodes and one output node can be seen as a “free-form” fitting function of $\mathbb{R}^N \rightarrow \mathbb{R}$ (see, e.g., Tagliaferri et al. 2003, for an introduction to NNs and applications to astronomy). As such, the property of a NN to be non-linear in its inputs (also called *features*) is explicitly desired, to allow for flexibility of the fitting function. A natural consequence of this non-linearity is that if noisy realizations of input data are to be propagated through the NN, the resulting distribution of outputs might well differ from the noise distribution of the inputs. This is a property shared by any non-linear estimator. In particular, the expectation value of the output can be offset from the output which would be obtained from noise-free or less noisy inputs, leading to a net noise bias.

Let us consider a NN of sufficient capacity, i.e. flexibility, for a given regression problem. The shape of the NN regression, i.e. the fitting function, is then entirely determined by the *training* of the network. This training consists in optimizing the network parameters, i.e. the weights and the biases of the nodes, so as to minimize a cost function which compares network predictions to some known truth, i.e. the “target” values. A simple and common choice for such a cost function is the mean square error (MSE) between the network predictions and the target values, in analogy to an ordinary least squares or maximum likelihood method. When fitting a model to noisy observations that depend on noiseless explanatory variables, the MSE does lead to the usually desired fitting curve (or hypersurface, in case of many-input nodes). The latter traces, in the limit of many observations, the average values of the observed variable in bins of the explanatory variables.

In this work, our use of NNs is however “inverse”. We want to regress estimates for the explanatory variable (the NN target)

based on noisy observations of the dependent variables (the NN inputs), a problem known in statistics as an *inverse regression* or *calibration*. As mentioned by Gruen et al. (2010), it is counter-productive in such a situation to train a NN to minimize a MSE expressed between targets and individual predictions of the explanatory variable. We can however formulate other cost functions which explicitly favor accuracy in the predictions of the explanatory variable, when facing noise in the observed dependent variables. For this, the training data has to be structured so that the neural network can experience several realizations of the noise in the dependent variables for each value of the target explanatory variable.

3.2. Training with realizations and cases

To structure our training data, we introduce the distinction between *realizations* and *cases*:

- A training *realization* is a single observation of the noisy dependent variables, for a particular (known) value of the explanatory variable. Note that measurements of the dependent variables resulting from a physical process give us such realizations, except that the value for the explanatory variable is usually not known.
- A training *case* is an ensemble of realizations obtained for the same value of the explanatory variable. In other words, for our application of NNs, it is an ensemble of (input, target) pairs all sharing the same target value. When the training data is entirely simulated, cases can easily be generated to contain as many realizations as desired.

The training data therefore consists of an ensemble of cases, each containing an ensemble of realizations. Cost functions can now take advantage of this structure. We define the mean square bias (MSB) cost function, which penalizes the estimated prediction bias over the realizations in each case, as

$$\text{MSB}(\mathbf{p}, \mathbf{D}, \mathbf{t}) \doteq \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left[\frac{1}{n_{\text{rea}}} \sum_{i=1}^{n_{\text{rea}}} o(\mathbf{p}, \mathbf{D}_{i,k}) - t_k \right]^2, \quad (3)$$

where \mathbf{p} groups all the parameters (weights and biases) of the NN, \mathbf{D} represents the training inputs, containing the input vector $\mathbf{D}_{i,k}$ for each of the n_{rea} realizations in each of the n_{case} cases. $o(\mathbf{p}, \mathbf{D}_{i,k})$ is the NN output for each realization, and \mathbf{t} the training targets (with the target t_k of each case). Note that in the same notation, the classical MSE cost function, making no distinction between realization and cases, would be written

$$\text{MSE}(\mathbf{p}, \mathbf{D}, \mathbf{t}) \doteq \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \frac{1}{n_{\text{rea}}} \sum_{i=1}^{n_{\text{rea}}} [o(\mathbf{p}, \mathbf{D}_{i,k}) - t_k]^2. \quad (4)$$

The apparently small difference between MSB and MSE is therefore that the MSB averages the NN outputs over the realizations in each case before comparing them to the target values. For both cost functions, the NN still learns how to predict one output for each realization. Let us note some consequences of the MSB cost function, which plays an important role in this paper.

First, in the limit of a sufficiently large number of realizations per case, the MSB does not penalize scatter in the predictions. A network trained to minimize MSB will, as desired, trade precision for accuracy, but it could potentially go beyond the optimal use of information and introduce additional unnecessary noise in

its predictions. In practice, one can control this behavior, as well as potential overfitting to the training data, by limiting the capacity of the NN, typically by limiting the number of nodes and layers.

Second, one has to acknowledge that an inverse regression problem might simply not have an “accurate” solution, in the sense of a solution with vanishing MSB. If the observed dependent variables (the NN input features) do not carry information about the explanatory variable (the NN target) the corresponding target values will not be accurately estimated. And even if this information is still there, given a finite number of realizations and cases, a sufficiently strong noise in the input features will lead to biased predictions. Note that this might affect some “difficult” realizations only, while other regions of parameter space allow for sufficient accuracy.

More generally, not only the accuracy but also the achievable precision of the predictions might vary from one realization to another. In situations where the noisiness of an observed realization can be estimated from the observation itself, we can therefore further mitigate the effect of noise and extract more information by going beyond the prediction of point estimates.

3.3. Predicting weights

In this paper, we explore the simplest extension to the prediction of point estimates, by including the prediction of weights. For this, we propose the use of a separate NN, in parallel to the NN predicting the point estimates. The two networks are trained successively. In a first step, the NN yielding point estimates is trained using the MSB cost function. Then, the second NN is trained to predict an optimal weight for each realization, in order to increase the accuracy of each case. For this second NN, with parameters \mathbf{p}_w and exclusively positive outputs w , we define the mean square weighted bias (MSWB) cost function

$$\text{MSWB}(\mathbf{p}_w, \mathbf{O}, \mathbf{D}, \mathbf{t}) \doteq \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left[\frac{\sum_{i=1}^{n_{\text{rea}}} o_{i,k} \cdot w(\mathbf{p}_w, \mathbf{D}_{i,k})}{\sum_{i=1}^{n_{\text{rea}}} w(\mathbf{p}_w, \mathbf{D}_{i,k})} - t_k \right]^2, \quad (5)$$

where \mathbf{O} contains the predicted point estimates $o_{i,k} = o(\mathbf{p}, \mathbf{D}_{i,k})$ obtained through the first NN. A peculiarity of this cost function is that no explicit target values for the weights is given. Furthermore, by construction, the weights w minimizing the MSWB might have an arbitrary scale. In practice, we can easily impose both the positivity and an upper bound to the weights by using an activation function $\mathbb{R} \rightarrow (0, 1)$ for the output layer of this second NN.

Note that the training data (\mathbf{D}, \mathbf{t}) for the weight training can have a different structure of realizations and cases than the training data for the point estimates. It is always possible to obtain the point estimate predictions \mathbf{O} from the first NN by running it on the training data of the second NN. We can thus make use of two training datasets, each optimized for its purpose.

We will further discuss the properties and behavior of NNs trained with the MSB and MSWB cost functions and the importance of the distributions of cases and realizations in Section 5, in the context of the practical application to weak lensing shear estimation.

3.4. Neural network implementation, training optimization algorithm, and committees

In the following, we briefly summarize details and default settings of the NNs. All results of this paper are obtained using an

experimental custom NN library implemented in `python`, which we make publicly available (see Appendix A).

For both types of networks (point-estimates and weights), we use small fully-connected NNs with typically 2 hidden layers of 5 nodes each. All input and hidden nodes use the hyperbolic tangent $f(x) = \tanh(x)$ activation function. For the output layer, we use an identity activation function for the prediction of point estimates, and a variant of a sigmoid, $f(x) = 1/(1 + \exp(-4x))$ for the weight-predicting networks. We follow the conventional practice to deal with highly heterogeneous feature scales, and prepend a normalization (or whitening) of the input data vectors to our networks (Graff et al. 2013). This normalization independently scales and shifts the features seen by each node of the input layer, so that, for the training data, all inputs cover the interval $[-1, 1]$.

Instead of using the conventional back-propagation (Rumelhart et al. 1986), we train our networks with a Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative optimization algorithm (Nocedal & Wright 2006, and references therein) in its `scipy` implementation⁵. The use of an algorithm that is agnostic of the network details, and therefore computes all required gradients numerically, allows for easy experimentation with cost functions and also with unconventional nodes, such as product units (Durbin & Rumelhart 1989; Schmitt 2002). To increase the efficiency of the training, we implement a caching mechanism for the results computed by each layer of the network. We also use so-called mini-batch optimization (see, e.g., Nielsen 2015), i.e. we randomly select a “batch” of typically 25% of the training cases, perform several (typically 30) optimization iterations on this batch, and iteratively pursue with the next randomly selected batch.

We start the training iterations from a randomized initial parameter state, with network weights and biases drawn from a centered normal distribution with a standard deviation of 0.1. Due to this random initialization as well as the mini-batch optimization, networks trained on exactly the same data yield different estimators. We exploit this stochastic behaviour to increase the robustness of our training procedure, by systematically using so-called *committees* of typically 8 NNs in place of individual NNs. After the parallel training of such a committee, and a repeated evaluation of the performance of each member on an independent validation dataset during the training, we retain the best half of the members to form our final estimator. This allows in particular to reject badly converged optimizations, and to verify the overall stability of the training procedure (see also Zhou et al. 2002). We take averages of the predictions made by the retained committee members as output of a committee⁶.

Finally, we note that our implementation allows to individually mask realizations of each case, which is important to handle failures of the input feature measurements, discussed in the next Section.

4. Feature measurement on galaxy images

The raw data of a weak-lensing study consists of survey images. In this section we describe how we measure a small set of features based on moments of the observed galaxy light profiles

⁵ <https://www.scipy.org>

⁶ For the weight-predicting NNs described in Section 3.3, the unconstrained scale of the predicted weights could potentially require a prior normalization. In practice, we observe however that the use of the sigmoid output activation function results in members predicting weights of very similar scales.

from which the shear is to be inferred. Those features will serve as input to the machine-learning algorithm, potentially together with information from a PSF-model, multiband photometry, or other relevant parameters. Note that for this exploratory work we deliberately opt for a small number of selected features describing the galaxy images, to ease experimentation, efficiency, and also to set a benchmark. Deep-learning approaches with convolutional NNs, which directly learn filters to extract optimal galaxy features from image pixels are an obvious alternative (Tuccillo et al. 2017; D’Isanto & Polsterer 2017). However, we expect that few simple “hand-crafted” features⁷ are sufficient to capture a very large fraction of the shear information from the noisy galaxy images of interest to a weak lensing analysis, especially on simple simulations.

4.1. Adaptive weighted moments

To describe the galaxy shapes we use statistics based on moments computed with an adaptive elliptical Gaussian weight function (in contrast to the circular weight function used in Tewes et al. (2012), which we observe to yield less precise results). We employ the well-tested and efficient implementation offered by the HSM module of the `GalSim` software package (Bernstein & Jarvis 2002; Hirata & Seljak 2003; Mandelbaum et al. 2012; Rowe et al. 2015, and references therein). The same or a very similar moment computations are used in other shape measurement techniques, such as DEIMOS (Melchior et al. 2011) and the methods directly implemented within `GalSim`.

To stress the computational nature of these features and connect them with the HSM implementation, we denote them in a fixed-width font. We define the following moment-based features.

`adamom_flux` corresponds to the total source flux of the best-fit elliptical Gaussian profile (`ShapeData.moments_amp` in `GalSim`), expressed in ADU. Note that this is a biased estimate of the flux of any realistic (i.e., non-Gaussian) galaxy profile, but that such biases have no direct consequences for machine-learning features.

`adamom_g1` and `adamom_g2` are components of the observed ellipticity (`ShapeData.estimated_shape.g1/2` in `GalSim`), which would correspond, for a simple elliptical Gaussian profile and without PSF, noise, and pixellation, to the ellipticity defined in Section 2 as an estimator for shear.

`adamom_sigma` gives a measurement of the average radial extension of the profile, in units of pixels (`ShapeData.moments_sigma`). In the case of a circular Gaussian profile, it would estimate its standard deviation.

`adamom_rho4` gives a weighted radial fourth moment of the image, measuring the concentration, i.e. a kurtosis, of the light profile (`ShapeData.moments_rho4` in `GalSim`).

4.2. Noise measurement and Signal-to-noise ratio

The signal-to-noise ratio (S/N) of galaxy images is a key quantity when assessing the quality of a shear measurement. A scientific analysis of a shear catalogue will tend to include galaxies with a S/N as low as tolerable, for a given shear measurement technique. Unfortunately, S/N measurements mentioned across the literature are often difficult to compare, as the observational

⁷ In machine learning, *hand-crafted* features are statistics designed and selected by an expert, versus *learned* features which are developed by an algorithm based on training data.

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

definition of a S/N is not trivial and not always fully described. In the following, we present the simple observational S/N that we use to evaluate our method.

First, we quantify the background pixel noise for each target galaxy using a rescaled median absolute deviate (MAD) to estimate the standard deviation (see, e.g. [Rousseeuw & Croux 1993](#))

$$\sigma_{\text{sky}} \doteq 1.4826 \cdot \text{median}(|\xi_i - \text{median}(\xi_i)|), \quad (6)$$

where ξ_i are the pixel values (in ADU) along the edge of a “stamp” of sufficient size centered on the target galaxy. Generalizations of this procedure, for better precision, are easily conceivable if required. The robust MAD statistic has the advantage, over a plain standard deviation, that potential field stars, galaxies, or image artifacts on the stamp edge have a reduced impact.

In the second step, we combine this background pixel noise measurement with the results from the adaptive moment measurements described above to obtain a S/N. Naturally, our definition of S/N follows from the CCD equation (see, e.g., [Chromey 2010](#)), and we choose a circular aperture with a radius of three times the measured half-light radius of the source as effective area for the background noise contribution. More precisely,

$$S/N \doteq \frac{G \cdot \text{adamom_flux}}{\sqrt{G \cdot \text{adamom_flux} + A_{\text{eff}} \cdot (G \cdot \sigma_{\text{sky}})^2}}, \quad (7)$$

where

$$A_{\text{eff}} \doteq \pi \cdot (3 \cdot \text{adamom_sigma} \cdot \sqrt{2 \cdot \ln(2)})^2, \quad (8)$$

and G is the gain in electrons per ADU. For a Gaussian profile, the numerical factor $\sqrt{2 \cdot \ln(2)} \approx 1.1774$ would rescale the standard deviation into the desired half-light radius. This choice of effective aperture A_{eff} has a strong influence on the S/N, and might seem arbitrary as galaxy light profiles are not Gaussian. We observe however that this definition gives results within a few percents of the widely-used ratio $\text{FLUX_AUTO}/\text{FLUXERR_AUTO}$ given by the **SEXtractor** software ([Bertin & Arnouts 1996, 2010](#)), for all simulations considered in this paper.

To mimic “sky-limited” observations, simulated images are sometimes drawn purely with a stationary Gaussian noise. In this approximation, Eq. (7) simplifies to

$$S/N_{\text{Gaussian}} \doteq \frac{\text{adamom_flux}}{\sqrt{A_{\text{eff}} \cdot \sigma_{\text{sky}}^2}}. \quad (9)$$

We show some simulated sources with Gaussian profiles for different S/N and sizes in Fig. 1 (see also Fig. 12 for an illustration with PSF-convolved elliptical Sérsic profiles).

Note that for a machine-learning shear-measurement, a measured S/N is potentially an interesting input feature of each galaxy, especially if the number of features needs to be small ([Tewes et al. 2012](#)). However, in the following, we will not use the S/N as input feature, but provide instead separately the more fundamental flux and size measurements to the machine-learning algorithm, complemented by a sky noise measurement if required. This use of flux instead of S/N allows, in particular, for testing a single training on test sets with different noise levels, or for training on data with a lower noise than the actual observations. Nevertheless, we will extensively use the observed S/N defined above in the analysis of shear estimation biases.

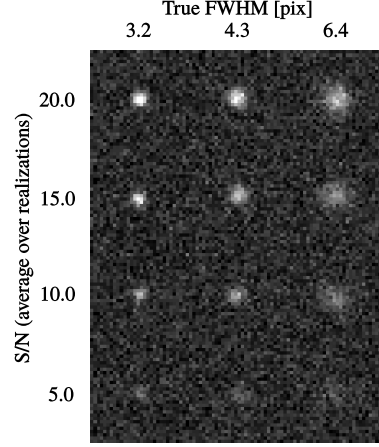


Fig. 1. Illustration of the S/N on simple Gaussian profiles with Gaussian pixel noise. The fluxes are chosen so that the average S/N, measured on many realizations of each source, matches the scale given on the left.

5. Machine-learning shear estimation

We now describe how we use and train NNs to predict an estimate for the shear of each galaxy, using the NN cost functions and the input features introduced in the previous sections. We focus on the core principles of the machine-learning approach, and defer for now the numerous complications that a full shear measurement pipeline has to face.

Recall that we consider here the prediction of point estimates of the shear components \hat{g}_i , $i \in \{1, 2\}$, and associated weights which we denote w_i . We will predict these point estimates and weights with independent NNs that are trained with different cost functions. For the sake of simplicity, we also distribute the predictions related to the two components to independent networks, instead of considering networks with multiple output nodes. We therefore train 4 scalar estimators, each consisting of a committee of several NNs.

Depending on the conditions in which the shear estimation method is to be applied, such as ground- or space-based data, variability of the sky background, instrumental effects in the data, selection of the source galaxies, accuracy to be achieved, different ways to setup and train these estimators can be considered. In the following, we present and motivate one simple fiducial approach in generic terms, using two different “training sets”, that is forward-simulations of observed galaxies with known shear. Alternatives will be discussed and explored later in this paper.

5.1. Step I: shear estimates with low conditional bias

We start by training the shear point estimates \hat{g}_i . A simple toy-model choice of the input features could be measures of the ellipticity components, the flux, the size of the observed galaxy image, the noise of the sky background, and the ellipticity and size of the PSF model⁸ at the location of the considered galaxy.

⁸ Note that a very different approach to inform the ML about the PSF is presented in Section 7.

These 8 input values allow the shear estimator to roughly correct for the PSF shape and noise bias.

We use the MSB cost function (Eq. 3), which, muting the explicit dependency on the training data, takes the form

$$\text{MSB}(\mathbf{p}) = \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left[\frac{1}{n_{\text{rea}}} \sum_{j=1}^{n_{\text{rea}}} \hat{g}_{jk}(\mathbf{p}) - g_k^{\text{true}} \right]^2 \quad (10)$$

for each component of \mathbf{g} , and with \mathbf{p} designating the parameters of the estimator.

The structure of the training set, that is the number of cases and realizations, is the next most important choice. To train the estimator to be both accurate and as insensitive as possible to the distribution of “true” galaxy properties, we aim at penalizing its *conditional* bias, that is its bias in any subregion of this true parameter space. In other words, we aim at a potential estimator which would be accurate for any PSF, and any true galaxy size, elongation, flux, etc.

We generate a training set as illustrated in Fig. 2. Within each case k , the realizations share the same true shear g_k^{true} (the target value for the training), but also the same value for other explanatory variables that we can request the estimator to attempt to become insensitive to, given the information it obtains from its input features. Consequently, each case contains only one “true” galaxy combined with one particular PSF, always seen under the same shear. While other aspects of the data, such as the position angle of the galaxy, its exact position on the pixel grid, and the realization of the pixel noise do have a direct influence on the shear estimate, they have to be dealt with statistically. Indeed, a shear estimator cannot be insensitive to the intrinsic orientation of a galaxy, which is degenerate with the shear. This orientation acts as a form of unavoidable noise (so-called shape-noise) for the shear measurement. Therefore, within each case, we draw n_{rea} realizations of these noise sources, and train the estimator to yield unbiased predictions despite this noise and pixellation.

Note that the required value n_{rea} to sufficiently average-out the noise effects with respect to a significant bias can be reduced by noise cancellation techniques. With such techniques, a controlled ensemble of compensating samples is taken, to improve the precision on the bias of a case beyond what would be achieved by randomly drawing the realizations. In Fig. 2, the intrinsic orientations of the galaxies are rotated in regular intervals on a ring in the (ϵ_1, ϵ_2) -plane, so that the average intrinsic ellipticity within each case exactly vanishes (following Nakajima & Bernstein 2007). Such techniques have become known as shape noise cancellation (see, e.g., Mandelbaum et al. 2014, and references therein).

Let us consider again the cost function. If a hypothetical estimator would achieve a zero MSB cost, for an infinite amount of realizations per case, this estimator could be said to be fully insensitive to the distribution of galaxy and PSFs it is presented with, among the population it was trained on. It is important to acknowledge that this is not possible in practice for all regions of this “true” parameter space: consider the example of an intrinsically small, “unresolved” galaxy, whose observed shape will not carry shear information. The PSF, the noise, and the pixellation lead to a loss of information which cannot be compensated by the estimator.

The role of the second step is to build a function that down-weights galaxies from which an unbiased estimate cannot be obtained.

Training for shear

Realizations differ in true orientation (shape noise cancellation), in noise, and sub-pixel position.

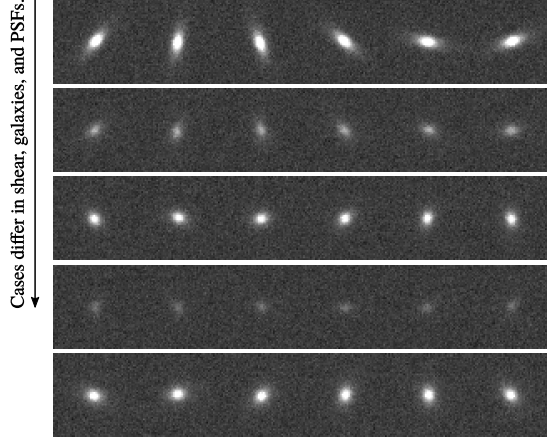


Fig. 2. Illustration of the structure of a training set to train a shear estimator \hat{g}_i with an MSB cost function. The horizontal frames correspond to different *cases*, each containing different *realizations* of a galaxy. All galaxies of a case are simulated with the same true shear, and the same PSF. Note that despite the circular symmetry of the PSFs used in this illustration, the typical cosmic shear is too small to be visible by eye.

Training weights

Realizations differ in galaxies, noise, and sub-pixel position.

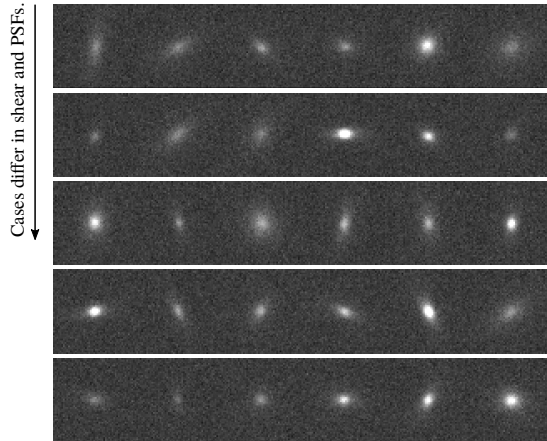


Fig. 3. Structure of a training set to train a weight estimator, w_i , with a MSWB cost function. Within each case, this training informs the method about approximated distributions of properties of the source galaxies and selection functions.

5.2. Step II: weight prediction

Given the estimators \hat{g}_i , $i \in \{1, 2\}$, we now train independent NNs to predict the associated weights w_i . We use an MSWB cost function from Eq. (5), which can be written, separately for each component i of the shear and its estimators, as

$$\text{MSWB}(\mathbf{p}_w) = \frac{1}{n_{\text{case}}} \sum_{k=1}^{n_{\text{case}}} \left[\frac{\sum_{j=1}^{n_{\text{rea}}} \hat{g}_{jk} \cdot w_{jk}(\mathbf{p}_w)}{\sum_{j=1}^{n_{\text{rea}}} w_{jk}(\mathbf{p}_w)} - g_k^{\text{true}} \right]^2. \quad (11)$$

We recall that the w are constrained to the interval $(0, 1)$ by design of the NNs, and that the estimates \hat{g}_i are to be computed ahead of the training of the weight-predictor, for each galaxy in the training data.

Again, putting aside technical details of the machine learning algorithm, we consider the choice of input features and the structure of the training data.

Regarding the input features, it could seem intuitive that a small set of features, describing for example the observed size and flux, is sufficient to optimally down-weight low-S/N source galaxies for which an unbiased shear estimation cannot be achieved. After all, if the \hat{g}_i achieve low conditional biases, the act of removing intrinsically small and faint galaxies from the sample cannot introduce any additional biases. This reasoning is however wrong, as we don't have access to any "true" galaxy parameters, which are unrelated to the shear. Instead, the input features, including the measurement of size and flux, are based on the observed galaxy and will inevitably show dependencies on the shear, at some level. Using such a small set of features would lead to a shear-dependent weighting, and thereby lead to biases even if the \hat{g}_i itself is accurate. Weighting acts in this regard exactly as any selection function, leading to *selection biases* (Miller et al. 2013; Kaiser 2000; Bernstein & Jarvis 2002).

Two conclusions can be drawn from this observation. First, it is justified to maintain the full set of features when training the weight-estimator, so that the weights can exploit the full information from each source to counter selection effects. Second, if selection biases prior to the shape measurement are affecting the data, this step of training the weights is a natural place to inform the ML-algorithm about the selection function.

We structure the training data for the weights as illustrated in Fig. 3. The realizations within a case still all share the same true shear and PSF, but now also sample ideally the full population of observed galaxies. By this structure, we therefore aim at predicting weights so that, for any shear and any PSF, the overall shear prediction error (both statistical and bias) gets minimized.

We stress that the introduction of these weights, estimated on the noisy observations, to the shear estimation formalism will potentially re-introduce some small conditional biases that we attempted to minimize in Section 5.1. For example, the different realizations of a galaxy shown in Fig. 2 will get slightly different weights. Given the loss of information in the observation process, it is expected that a shear estimator cannot be fully insensitive to the true galaxy parameters. But the approach presented here attempts to minimize this sensitivity to smallest achievable level.

Furthermore, we stress that the trained weight-estimator *does* depend on the distribution of galaxy properties in the training data. To pick again an extreme example for the purpose of illustration, the prevalence of unresolved galaxies (or mis-identified stars) in the source population will influence how conservative the rejection of small observed galaxies needs to be in order to avoid biases. Also, a shape-noise cancellation scheme cannot be

used for the weight training, as the cancellation would be more efficient on high-S/N than on low-S/N galaxies, and would thus exaggeratedly favor the former.

5.3. On the estimation of ellipticity, size, magnification or other parameters

To describe how the parameter of an idealized galaxy, as it would be seen by an infinite-resolution and noise-free imaging system, can be estimated, we consider the example of the ellipticity of a source. This example is of particular interest, as it can serve as a shortcut for shear estimation. The set of input features remains unchanged from the previous sections, and the NNs are trained with an MSB cost function, using the ellipticity components of an idealized galaxy defined in Section 2.1 as target values⁹. Figure 4 illustrates a simulation structure to train such a point-estimator for "true" ellipticity (that is ellipticity before PSF-convolution and noise). Cases cover a variety of galaxies and PSFs, and the realizations within each case differ only in noise and their exact positioning on the pixel grid. No shear is added to the training simulations. Under the hypothesis of the idealized galaxy morphology, the resulting point-estimate of the ellipticity can be seen as an estimate for the shear. Associated weights can be trained exactly as done in Section 5.2. The advantage of this shortcut, with respect to training a shear estimator, is that less realizations per case are needed to achieve comparable results, as the shape noise has been removed from the problem.

Variants of these steps can be used to construct estimators for other galaxy properties, as for example the angular size. The availability of an accurate size measurement is mandatory for galaxy size-magnification studies (Schmidt et al. 2011), which suffer from the same instrumental bias sources as shear studies. An approach to directly predict a magnification estimator could also be explored, in analogy to the shear estimator presented in this paper. Doing so, the ML algorithm could learn to exploit physical correlations between galaxy properties (see, e.g., Huff & Graves 2013) while compensating for the *observed* correlations introduced by the measurement process on noisy images. Discussing these estimators in more detail is, however, beyond the scope of this paper.

5.4. Practical notes on the training convergence and data

The successful training of supervised ML algorithms typically requires some experimentation with hyperparameters, such as the size of the NNs and the size of a training set. In the following, we briefly list important observations and advices which ease the methodical optimization of the architecture and training of the neural networks. While some of these suggestions might seem elementary to ML-practitioners, we detail them in the particular context of the presented galaxy shape measurement problem. We assert that these principles are useful for any ML shape measurement approach.

1. *Validation set*: arguably the most important idea is to always use a separate validation dataset to evaluate a training performed on some training dataset. This validation set can be simulated in the similar way as the training set, but should otherwise be independent (i.e., contain different cases and

⁹ Recall that *real* galaxy profiles have no simple ellipticity. However, if the use of galaxies with complex morphology is required, shape estimates previously obtained from other techniques can serve as target values.

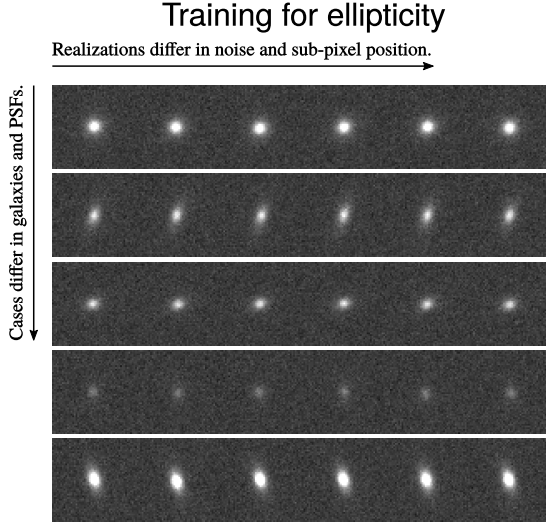


Fig. 4. Structure of a training set to train a *ellipticity* estimator, which can be useful to act as a shortcut to the shear estimator training.

- realizations). Monitoring the cost function value on both validation and training sets during the training allows the detection of training convergence and potential overfitting of the ML algorithm. Overfitting could happen if the training set is too small (in terms of n_{rea} and n_{case}) and/or if the NN-capacity is much higher than required. Validation sets with both structures shown in Fig. 2 and 3 are useful. The first one to test the achieved quality of the point shear estimator, and the second one to test the overall shear estimate including the predicted weights.
2. *Start with small NNs*: the experimentation should start with few simple features and a very low-capacity network, to obtain a benchmark solution. Before adding features, or increasing the NN capacity by adding nodes, a validation set of sufficient size on which one can clearly visualize the limitation of the benchmark solution should be available. Note that the shear estimation problem, based on the input features described in this paper, does not require a large capacity. Indeed, the dependency of the shear estimate on the observed features is smooth, and no discontinuities are expected.
 3. *Training set adjustments*: it is often advantageous to use *different* source galaxy property distributions in the training set than in the data one wishes to process. In particular, when training the shear point estimator with MSB (Section 5.1 and Fig. 2), cases from which no unbiased shear estimates can be expected should not be in the training set. A typical example is unresolved galaxies, with an intrinsic extension much smaller than the PSF, whose features will carry no shear information. Even a small number of these cases can dominate the cost function value, and lead the NNs to overfit and yield biased estimates also on much easier cases. For the same type of training, it can also be beneficial to fill the true parameter space relatively uniformly, and to extend the range of true simulation parameters (e.g., shear, galaxy flux and size) beyond what the real-sky data contains. Note that the training

set for the weight-prediction, and also an overall validation set, *should* mimic the real-sky data and therefore include all problematic cases in a representative way, so that the weight-prediction and actual performance of the trained NNs can be assessed.

4. *Amount of simulated data*: ideally, the size of a training set should be increased up to a point at which no further improvement of predictions made on an even larger validation set is seen. As a rule of thumb, the validation set size needed to probe biases to some desired accuracy gives a good indication of the required training set size. For example, as the different shear “cases” of a constant-shear GREAT3 branch contain 10 000 galaxies each (including shape noise cancellation), one needs a training set with about as many realizations to obtain satisfactory results.

6. Demonstrating the method on simple simulations

We first apply the approach on a set of simulations with limited complexity. We demonstrate strategies to

1. predict accurate point-estimates for ellipticities,
2. produce shear estimates,
3. compute shear associated weights,
4. and account for spatially variable PSFs.

In an effort to minimize the size of the training set (thus reducing the training time) and control over-fitting, we impose two restrictions. First, the number of input features is limited, and, second, the architecture of the network is small (narrow and shallow). By studying narrow and shallow networks on reasonably simple data, we can better identify important trends in the behaviour of the networks and their training.

6.1. Data simulation and distribution of galaxy properties

Table 1. Parameter distributions of the training sets generated for fiducial experiments.

Set to learn the prediction of:	Ellipticity	Shear / Weight
Shear components g_1, g_2	0	$\mathcal{U}(-0.1, 0.1)$
Intrinsic ellipticity amplitude e	$\mathcal{R}(0.2)_{[0, 0.6]}$	$\mathcal{R}(0.2)_{[0, 0.6]}$
Sérsic index ^a n	$\mathcal{U}(0.3, 4)$	$\mathcal{U}(0.3, 4)$
Surface brightness S_B [a.i.]	$\mathcal{U}(0.8, 1.2)$	$\mathcal{U}(0.8, 1.2)$
Half-light radius R [pix]	$\mathcal{U}(1.5, 15)$	$\mathcal{U}(1.5, 4)$

Notes. We use $\mathcal{U}(a, b)$ to denote the uniform distribution between a and b , and $\mathcal{R}(\sigma)$ for a Rayleigh distribution with mode σ . Intervals in subscript denote the range to which we clip a distribution, so that no sample falls outside of the given interval. a.i. stands for arbitrary units.

^(a) In practice, we grid the values for the Sérsic index instead of drawing them randomly. This significantly speeds up the galaxy stamp generation, as GalSim can reuse cached Sérsic profiles.

The parameters of the simulated galaxies are drawn from simple distributions, presented in Table 1. Small galaxies with large fluxes and conversely are avoided by computing the flux F from a combination of the surface brightness S_B and the half-light radius R_{hl} ,

$$F = \pi R_{hl}^2 \cdot 10^{-S_B}. \quad (12)$$

We use GalSim (Rowe et al. 2015) to generate all simulations. Sub-pixel dither and, if applicable, Gaussian noise are included

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

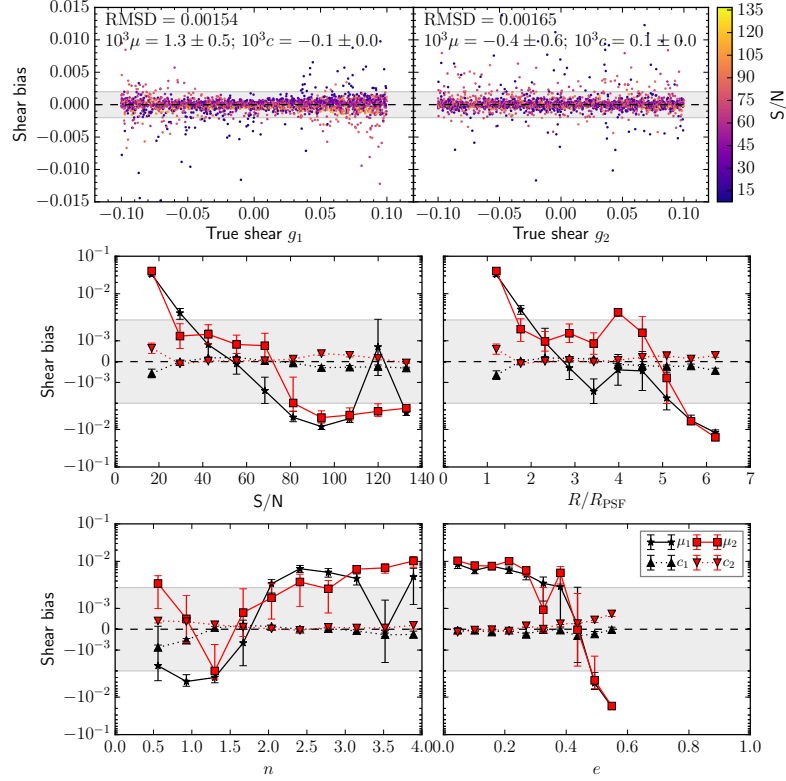


Fig. 5. Shear biases for test galaxies as function of the true shear components (top panel) and the properties of the galaxy profiles (bottom panels). The shear is predicted by a committee of six networks that were trained on noise-less images. The grey band depicts the region which complies with the bias requirements of stage IV weak gravitational lensing surveys. The bottom panel plots are shown in with a linear scaling in the grey band and logarithmic elsewhere. The error bars show the $1-\sigma$ confidence interval for the multiplicative bias μ and the additive bias c . The parameters of the galaxy profiles are defined in Tab. 1.

in the 48×48 pixels galaxy stamps. Because we include only Gaussian noise, we place ourselves in the sky-limited regime, hence the S/N ratio is computed according to eq. (9). To meet the sky-limited condition with the flux induced by the choice of surface brightness, we set the noise level to $\sigma_{\text{sky}} = 0.008$ ADU. For simplicity, we choose a pixel size of $1''$ and use a round Gaussian PSF with a standard deviation of $2''$.

6.2. Predicting accurate ellipticities estimates

We begin by predicting point estimates of the ellipticity components. We prepare $n_{\text{case}} = 10\,000$ noiseless galaxy images for training, each with 20 different dither realizations. This dataset is typically generated in 50 2.70 GHz CPU hours. The absence of noise considerably reduces the required size of the training sample and the training time. Networks are shown three features:

1. the ellipticity component (admom_g1 or admom_g2).
2. the size (admom_sigma).
3. the concentration of the light profile (admom_rho4).

Additional features do not decrease the cost function. We note that this statement is only valid for the fiducial dataset. If the

simple relationship between flux and size (eq. 12) is no longer valid, introducing the measured flux as a feature will be beneficial to performance.

We train 6 networks, which form a committee. The architecture of the networks is 2 hidden layers each containing 5 neurons. We train a committee for e_1 and one for e_2 . 20% of the training examples are set aside to prepare a validation set to independently estimate the training error. The parameters are optimized using a mini-batch training of 20% of the original size of the set. We run the optimizer for 200 iterations before changing to another batch. This procedure is repeated 50 times. The training time is of the order of 4 hours. We note that our neural network implementation is not optimized for speed. Computing time is shown for comparison between the experiments in this paper. Three networks are selected based on their training performance¹⁰ to predict an averaged value of the ellipticities.

We generate 50 millions galaxies for the test set, with $n_{\text{case}} = 5\,000$ and $\text{SNC} = 10\,000$. Each case is sheared by a random value, chosen to be between -0.1 and 0.1 . We remove from the

¹⁰ Networks are selected by the value of their cost function on a randomly-selected validation set.

M. Tewes et al.: shear measurement with machine learning

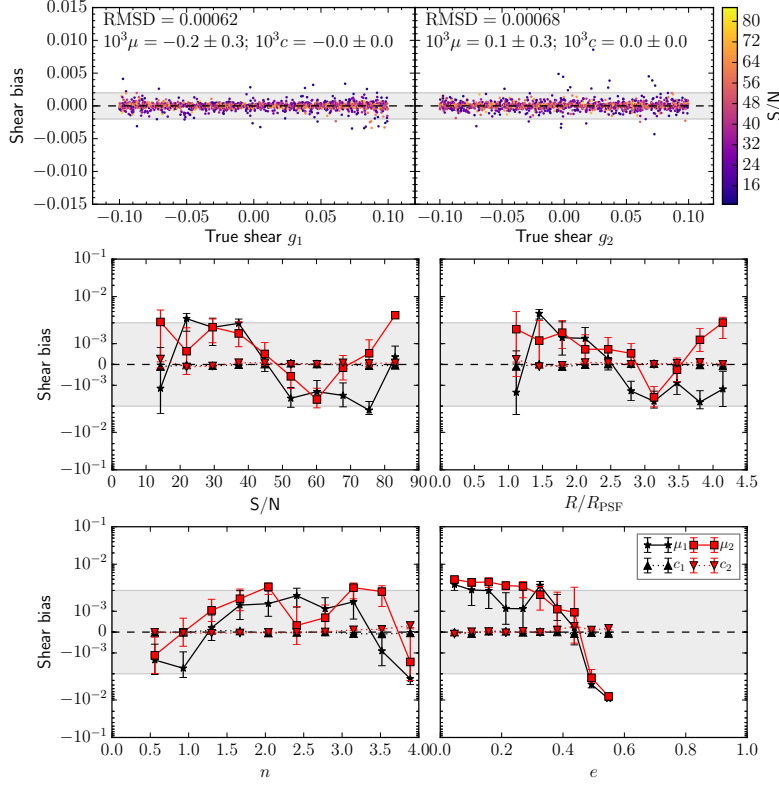


Fig. 6. Remaining shear biases measured on the test set, but networks were trained with noise. Same axes as in Fig. 5.

test set any case with more than 1% failed shape measurements. This removes less than 10% of the cases.

As shown in Fig. 5, predictions for the ellipticity components are accurate. We consider a bias significant if it is larger than the requirements on stage IV weak lensing surveys. There are, however, a number of significant outliers. Outliers are either at low or high S/N . They stem from very small and very large galaxies. When the size of the galaxy becomes too large, a significant fraction of the light falls outside of the stamp. Removing badly-cropped galaxies is a straightforward option to mitigate the large-size bias. Another method would be to add more such galaxies in the training set. Small galaxies in this fiducial dataset have a low S/N . Training with many realizations of noisy galaxies would reduce the low S/N bias. We choose to include low S/N stamps in the training set and to discard too large galaxies.

We create another, noisy, training set. Noise and image properties are the same than the test set. Training on a noisy data set requires more realizations of the images. We multiply the number of cases by two and realizations by a factor of 25. This leads to a training set of one million examples, and a simulation time of about 200 CPU hours. We keep the same model as in the noiseless case. The training time dramatically increases from a few hours to ~ 24 hours per network. In the training set, we set aside all cases with more than 1% of the realizations could not be measured.

The resulting multiplicative bias for low S/N galaxies is small, consistent with the requirements (Fig. 6). The important sensitivities to S/N and radius observed in the noiseless training have disappeared. The precision has increased with respect to the noiseless case. Noise can be included in the data set from the start of the learning procedure, but it can also be added later. Including noise late in the training procedure lead to similar performances.

We have demonstrated that we can predict accurate ellipticities for this simple fiducial dataset. Training with noisy inputs is longer, but translates into less conditional biases and better precision.

6.3. Predicting accurate shear estimates

Instead of predicting an ellipticity component, the networks now predict the shear, as in Sec. 5.1. We still use the same input features. The structure of the training set is substantially changed compared to the ellipticity approach, however we still use noise-free examples. Every training example is now sheared by a value between -0.1 and 0.1. We prepare 10 000 cases of different shear. S/N is increased from one to 1 000. The total number of training images is 10 millions. The architecture of the networks and the training procedure remains the same as for the ellipticity. The test set comprises 5 million images of galaxy differentiated in $n_{\text{case}} = 5\,000$ cases each with a S/N of 1 000.

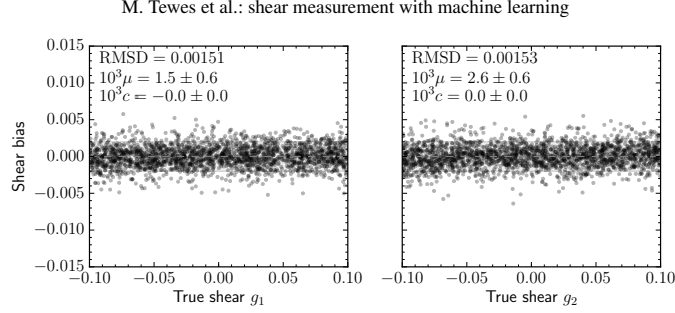


Fig. 7. Remaining biases for the direct shear measurement without applying any weighting scheme. Galaxies with $S/N < 30$ were manually cut out.

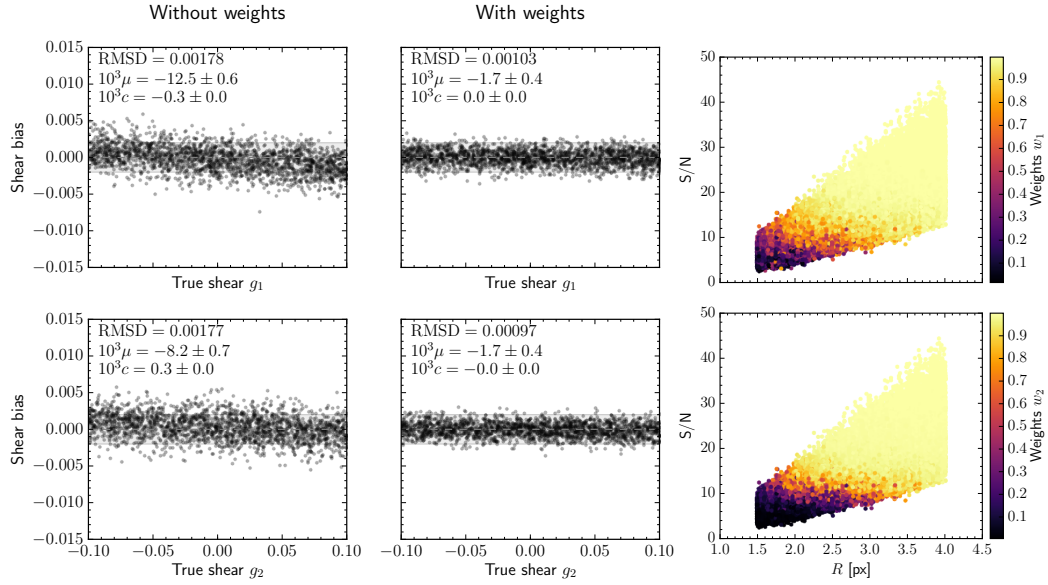


Fig. 8. The four left panels show shear bias in function of the true shear for the component one and two when averaging without or with weighting. The rightmost panels show the weights for the two components in function of the true flux and the half light radius. Predictions in the left most panels are obtained with data similar to Fig. 7, without manually weighting the galaxies.

We pre-process the test set by filtering out badly measured galaxies (by the same procedure as in the ellipticity approach) and by removing galaxies with $S/N < 30$. The remaining biases are shown in Fig. 7. The multiplicative and additive biases are consistent with the requirements.

We stress that this selection cuts off a significant fraction of the test set. In this experiment, we showed that we can successfully measure the shear directly, without defining the notion of ellipticity. However there are two important caveats in this experiment: (1) a case is made up of several realizations of the same galaxies: the position angle, sub-pixel dither and noise are changed. (2) We had to manually select the rejection thresholds. The former is physically unrealistic and the latter requires much manual inspection and corrections. We will solve both issues in the next experiments by introducing a weighting scheme.

6.4. Predicting weights for shear

For this experiment, we generate a test set which is more realistic than in the previous sections. A case of shear now consists of 2 500 different true shapes, with SNC of 4 and a single realization of the stamps. We test for $n_{\text{case}} = 1\,000$ different cases leading to 10 million galaxy training images. We first train a committee of networks to estimate the shear. The predictions of this network are accurate on the training set, but significantly biased on the test set. The multiplicative bias is of the order of 1-1.5%, an order of magnitude above the target.

The weights are learned in a committee of 6 networks following Sect. 5.2. The feature set is two-dimensional: `admom_flux`, and `admom_sigma`. The architecture is shallow: only one hidden layer with three neurons.

We generate 100 cases of shear with each 100 000 galaxies per case, or 10 million stamps, requiring ~ 400 CPU hours. A network is typically trained in ~ 8 hours.

The committee learns the expected cut-off weight function as a function of the S/N as shown on Fig. 8. We note that the features are biased estimates of the true flux and size. Applying the weights yields a non-significant shear bias of $-0.17 \pm 0.04\%$. The weighted predictions for g_2 are similar. The additive bias is consistent with zero for both shear components. The precision of the predictions is substantially increased when the shear are weighted.

We have shown that the networks can learn a weighting scheme. The resulting shear predictions are accurate and precise. We stress that the weight training set has to represent a realistic case, to teach the network about selection biases.

6.5. Inclusion of a spatially-variable PSF

Previous experiments were performed using a spatially constant Gaussian PSF. We now turn to predicting point estimates for the ellipticities with a smoothly varying Gaussian PSF.

Our strategy is to teach the networks that the PSF is varying across the field. We generated galaxies at random positions in a field of arbitrary dimensions. We vary the PSF ellipticities in the range specified for the *Euclid* mission (see Laureijs et al. 2011).

We extend the feature set by including the coordinates of the galaxies in the PSF field. The architecture of the network is otherwise kept as is. We call these networks position-aware networks. The remaining biases are compared to networks trained without galaxy positions. We refer to these networks as naive networks. The approach is tested on increasingly complicated PSF fields (illustrated on on Fig. 9):

1. Only the PSF e_1 component is varied and depends on the x position in the field. Other parameters are constant.
2. PSFs are radially oriented towards the centre of the field and their size increase with the distance to the center.
3. Component PSF e_1, e_2 are radial, pointing outward and increasing with the distance to the center. The size smoothly varies in the field.
4. An *Euclid*-like asymmetric field.

No significant overall or conditional biases are observed as a function of the galaxy parameters (Fig. 10). This statement is valid for position-aware networks only. Significant conditional biases arise as a function of the PSF parameters for naive networks (Fig. 11). Naive networks introduce both significant additive and multiplicative biases. Additionally, the precision of the predictions is reduced.

In the position-aware networks, we increased the size of the feature set, without increasing the size of the networks. Results show that the PSF is corrected, and that the overall performance is similar to the stationary PSF experiments. The above results support our approach to gradually add complexity to simulations and teaching these new effects to the networks, only increasing the network capacity when necessary.

7. Application to *Euclid*-like simulations

We turn to simulations mimicking realistic galaxy properties and *Euclid*-like imaging characteristics. This dataset, while still based on single Sérsic profiles and galaxy stamps without blends, will allow for some first quantitative best-case evaluation of the method in a pixellation and noise regime comparable to the *Euclid* survey.

We apply our method in two ways:

1. Extracting shear from predicted ellipticities, like in Sect. 6.2.
2. Predicting a weighted shear, like in Sect. 6.4.

We also study the sensitivity of the performance to the input parameter distributions.

7.1. Image simulations

As in the previous section, we simulate the images with `GalSim`. The galaxies get sheared, convolved with the PSF, pixelized, and finally noise Poisson shot noise and Gaussian read-out noise is added using `GalSim.CCDNoise()`. We describe parameters of this simulation procedure in the following.

7.1.1. GEMS-like galaxy properties

Galaxy parameter distributions are based on the Galaxy Evolution from Morphologies and SEDs (GEMS, Rix et al. 2004) survey observed with the *Hubble Space Telescope* (HST). The survey fitted single Sérsic profiles to the observations, resulting in a parameter catalog. We use this GEMS catalog with the same methodology as in Hoekstra et al. (2017). Galaxies in the magnitude range $20 < M < 24.5$ are generated, to mimic the VIS selection function. We neglect the effect of light profile blends, therefore there is no need to simulate fainter objects. The simulated, GEMS-like, galaxy catalogues contain five parameters.

1. Magnitude, M . The number density of simulated galaxies per magnitude is a power law. We identify the F606 magnitude with the VIS band.
2. Sérsic index, n . The observed distribution is discretised to speed up the image generation process.
3. Half-light radius, R_{hl} . Half light radius are drawn from a skewed normal distribution, whose radius and width linearly depend on the magnitude of the objects.
4. Ellipticity¹¹, e . In a simplification effort, and in accordance with Hoekstra et al. (2017), ellipticities are assumed to be uncorrelated with magnitudes. We draw ellipticities from a Rayleigh distribution,

$$P(e, e_0) = \frac{e}{e_0^2} \exp\{-e^2/2e_0^2\}. \quad (13)$$

We set $e_0 = 0.25$, and apply a ellipticity cut-off at 0.9.

5. Position angle, θ . We assume no intrinsic alignment in our simulations, therefore the position angle is assumed to be uniformly distributed.

Details of the parameter distributions and their comparisons with the GEMS distributions are presented in the appendix B. We neglect the wavelength dependence of the *Euclid* galaxy profiles.

7.1.2. *Euclid* VIS-like PSF

We choose to prepare a constant VIS-like PSF with a G5V template spectrum, selected from the Pickles library (Pickles 1998). We simulate an under-sampled PSF with characteristics similar to *Euclid* with the `GalSim` software. As demonstrated in Sect. 6.5, variable PSFs can be processed with good performance. In order to remain in a relatively simple setting to keep the simulation and training times reasonable, we assume a

¹¹ More precisely, the polarization $e = (a^2 - b^2)/(a^2 + b^2)$

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

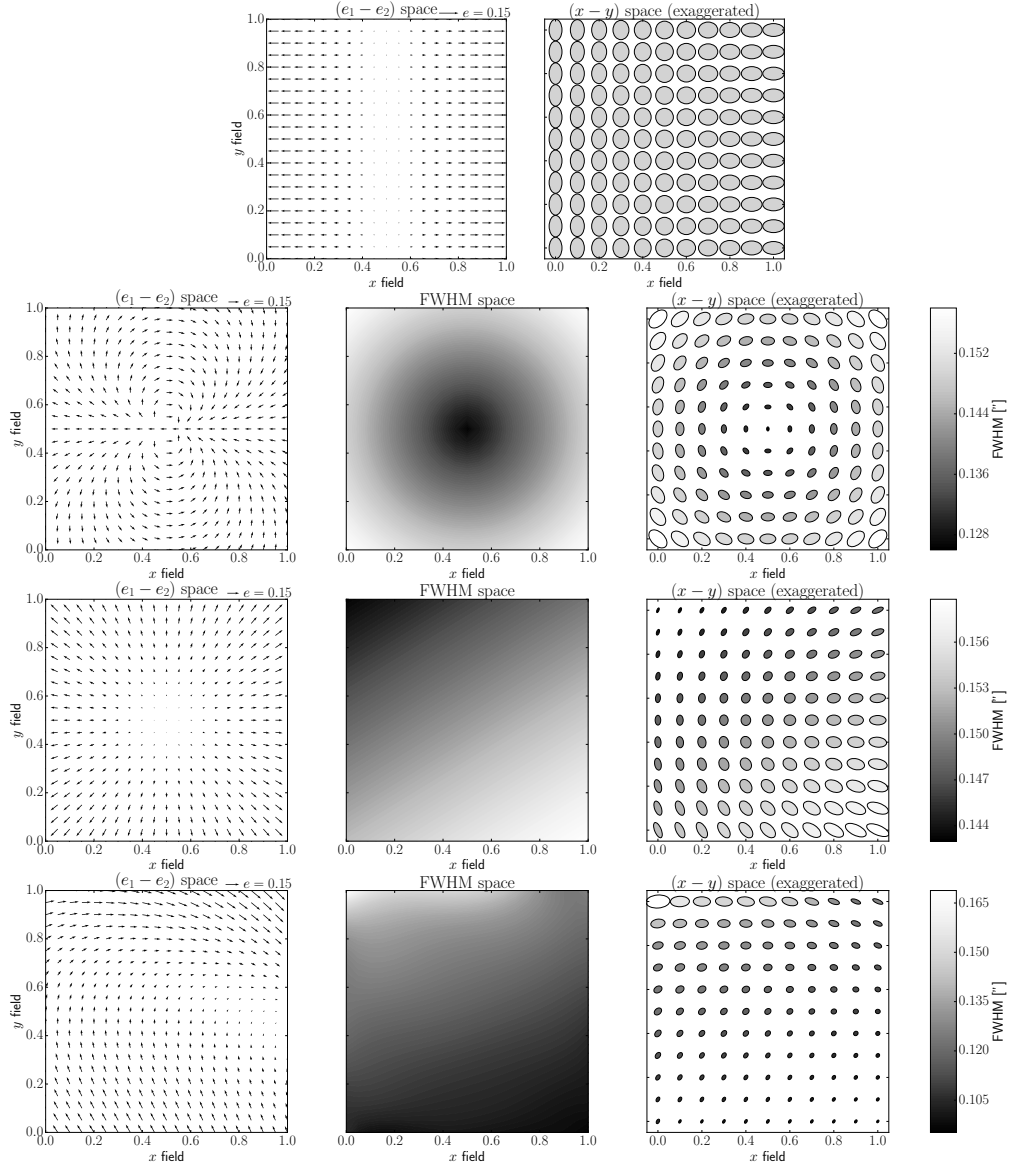


Fig. 9. The fields used for the variable PSF fiducial case. From top to bottom: only e_1 component varying, radially-oriented PSF, e_1 and e_2 components radially-oriented and *Euclid*-like fields.

spatially-stable PSF. We select randomly a PSF whose characteristics meet the *Euclid* PSF profile requirements (Laureijs et al. 2011). The PSF is almost round and has a FWHM of $0.18''$. PSF parameter choices are listed in appendix C. We note that we do not aim at reproducing a realistic *Euclid* PSF, but rather a PSF with similar characteristics. A fast substitute for this *Euclid*-like PSF is a Gaussian profile with a standard deviation of 0.75 pixels, corresponding to a FWHM of 1.77 pixels.

7.1.3. Preparing *Euclid*-like images

To make our results easily reproducible and comparable, we fully describe the *Euclid*-like simulations.

- Pixel size l_{pix} of $0.1''$ (Laureijs et al. 2011).
- Stamp size, $n_{\text{pix}} = 48$ pixels on-a-side.
- Exposure time, $t_{\text{exp}} = 3 \times 565$ s (Laureijs et al. 2011). Individual VIS exposures will have an exposure time of

M. Tewes et al.: shear measurement with machine learning

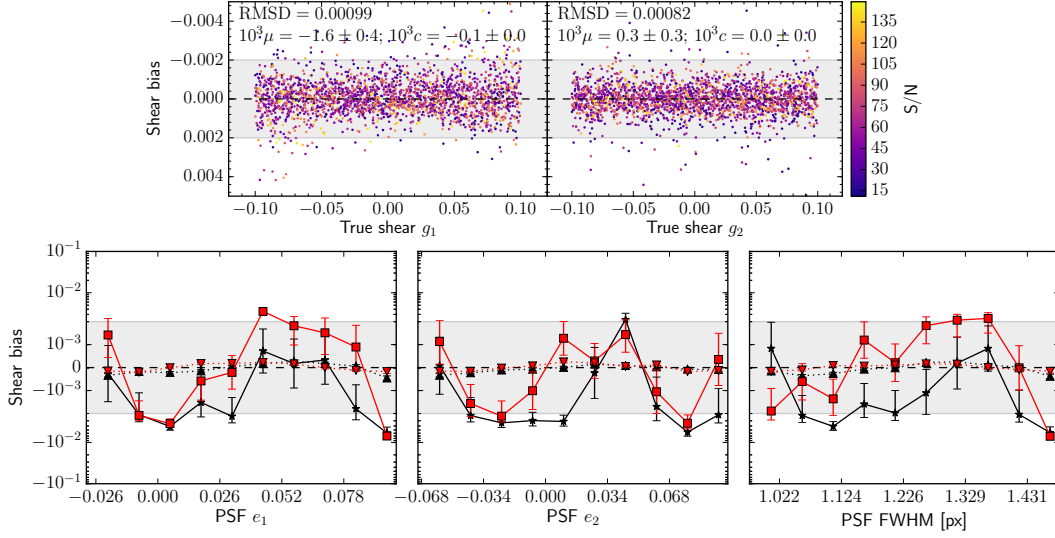


Fig. 10. Shear bias for a variable *Euclid*-like PSF field experiment. Committees are trained with galaxy spatial positions. The bottom panel shows the biases in function of the PSF shape.

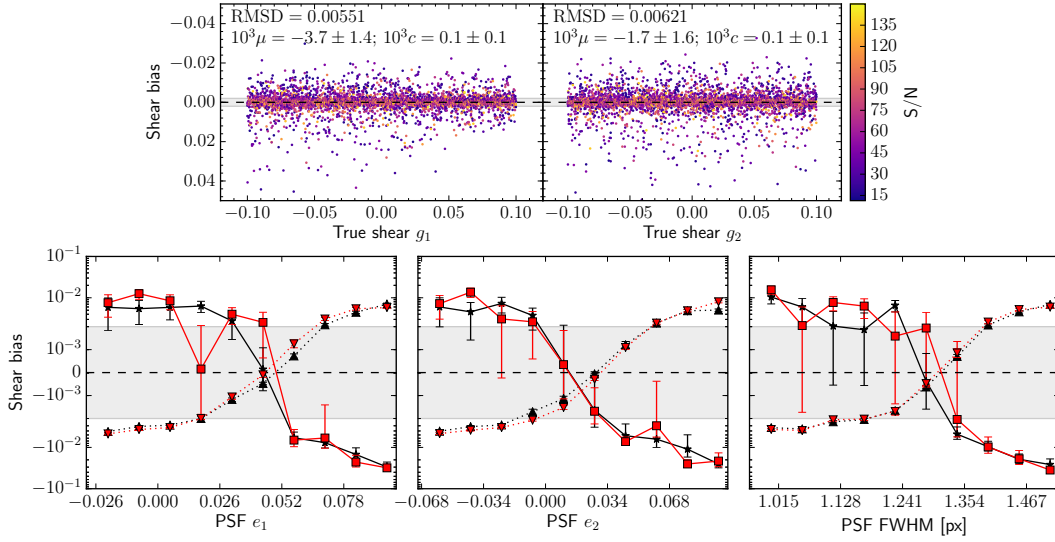


Fig. 11. Shear bias for a PSF variable experiment. Committees are trained without PSF or galaxy spatial position. The PSF field is an asymmetric, *Euclid*-like, field of Fig. 9. We highlight that the scale of the top plot is one order of magnitude larger than when networks are trained with spatial information (see Fig. 10).

- 565 s, but here we directly simulate the depth of 3 exposures as a single acquisition.
- Gain, $G = 3.1$ electrons/ADU (Niemi et al. 2015).
 - Read-out noise of 4.2 electrons (Cropper et al. 2016).
 - Sky background of $M_{\text{sky}} = 22.35$ mag arcsec $^{-2}$ (Réfrégier et al. 2010), which corresponds to a brightness of $2.5 \cdot 10^{-18}$ erg s $^{-1}$ cm $^{-2}$ Angstrom $^{-1}$ arcsec $^{-2}$, assuming a tophat VIS

- spectral sensitivity¹². We assume that we are dominated by the Zodiacal background.
- Zero-point, $Z_p = 24.6$ mag (see below).

¹² Through this paper, we use the AB magnitude system.

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

To generate the profile with GalSim, the flux, in ADU, of a source of magnitude M is given by

$$F[\text{ADU}] = \frac{t_{\text{exp}}[\text{s}]}{G} \cdot 10^{-0.4(M-Z_p)} \quad (14)$$

and the sky level, in ADU per pixel, by

$$F_{\text{sky}}[\text{ADU/pixel}] = (l_{\text{pix}}[''])^2 \cdot \frac{t_{\text{exp}}[\text{s}]}{G} \cdot 10^{-0.4(M_{\text{sky}}-Z_p)}. \quad (15)$$

A sample of simulated galaxies are shown on Fig. 12.

We adjust the value of the instrumental zero-point empirically, such that our simulations meet the expected *Euclid* performance. More precisely, we aim at obtaining an average S/N of 10 on simulated 3×565 -second exposures of elliptical Sérsic galaxies with parameters drawn from the GEMS distributions for a magnitude of 24.5 in the original resolved F606W filter, and convolved with a *Euclid*-like PSF. The S/N of a source depends strongly on its extension, however we observe that limiting the observed size of sources to a half-light-diameter of $0.43''$, following the *Euclid* performance description given in (Cropper et al. 2016), would yield a similar zero-point (within 0.1 mag). Our estimate of the VIS zero-point is certainly only a rough approximation, we provide it solely to make our simulations easily reproducible.

7.1.4. Lensing properties

In the *Euclid*-like simulations, just like in the fiducial experiments, we assume a constant value per case for the shear. We draw shear values from an uniform distribution between -0.1 and 0.1. We assume no magnification in our dataset.

7.2. Model architecture and input features

We keep our networks narrow and shallow. Six networks with two layers and five neurons per layers are trained per ellipticity component. The networks learn to minimize the mean square bias cost function. We use the same training and performance validation procedure as in Sect. 6. The committees retain the three best members for prediction.

The features set contains four items:

1. the ellipticity component 1 or 2 (admom.g1 or 2).
2. the size (admom.sigma).
3. the concentration (admom.rho4).
4. the flux (admom.flux).

The flux is a conceptually motivated new feature. In the *Euclid*-like dataset, the distribution of flux and radius are not straightforwardly related. Contrary to the fiducial dataset, including an estimate of the flux increases the information on the galaxy shapes.

7.3. Shear from ellipticity predictions

We estimate galaxy ellipticities with the same procedure as in Sect. 6.2 and measure the remaining biases on a large test catalogue of 40 million galaxies. The test contains $n_{\text{case}} = 4000$ cases, each galaxy generated with $\text{SNC} = 10000$. Similarly to the fiducial case, we apply selection cuts on the test set to remove cases for which adaptive moments failed for more than 1% of the galaxies, and galaxies with S/N lower than 10. These criteria set aside less than 10% of the catalogue.

We train networks on three different datasets:

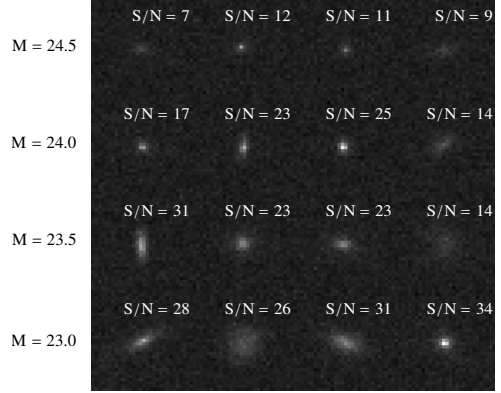


Fig. 12. Stamps of *Euclid*-like galaxies whose parameters are drawn from GEMS distributions. Only the central 30 pixels are shown.

1. A noiseless *Euclid*-like training set, with galaxy parameters distributed according to our GEMS model.
2. A noisy *Euclid*-like training set, with GEMS distributions.
3. A noiseless simple uniform distribution of the parameters. We draw uniform distributions in ellipticities, Sérsic indexes, sizes and half light radius. Half-light radii are uniformly drawn between 0.1 and 1.9 arcsec.

The differences in input distributions and features are shown in Fig. 13. The uniform distribution covers a somewhat larger region of parameter space.

We first train the committees with a noiseless dataset based on the GEMS distributions. The total training catalogue lists $n_{\text{case}} = 200000$ cases, each with $n_{\text{rea}} = 3$. This training dataset requires of the order of 50 CPU hours to generate. Trainings occur on the same machines as for the fiducial networks, completing their tasks in about 3 hours.

The remaining biases measured on the test dataset are non-significant: both additive biases c_i are consistent with zero, and the multiplicative factors, μ_i are consistent with the requirements. This suggests that our simple models have sufficient capacity to deal with the increased complexity of the dataset. The sensitivity $|\partial\mu/\partial p|$ of the method is small for all parameters, except a for hint of higher biases at high Sérsic indices.

Secondly, to reduce the conditional biases in the predictions, we train a committee on a noisy GEMS dataset. We prepare $n_{\text{case}} = 30000$ cases, each with $n_{\text{rea}} = 2000$ realizations of the noise and dither, leading to a large 60-million samples training set. As for the fiducial case, we remove hopeless training cases. The number of cases is reduced to simulate this dataset is a time of ~ 1000 CPU-hours. Training time substantially increases to ~ 72 hours. Overall remaining biases are consistent with zero for the multiplicative and additive components (See Fig. 14). There are no consistent trend in the sensitivity, except for a dependency on the Sérsic index.

Thirdly, we train our models with noiseless galaxies generated from uniform distributions of the parameters. The committees are designed and trained exactly as the previous experiments. We keep the training set size as in the noiseless training, to $n_{\text{case}} = 200000$ cases, with $n_{\text{rea}} = 3$. The performances on the test set exhibit negligible additive biases, and the multiplica-

M. Tewes et al.: shear measurement with machine learning

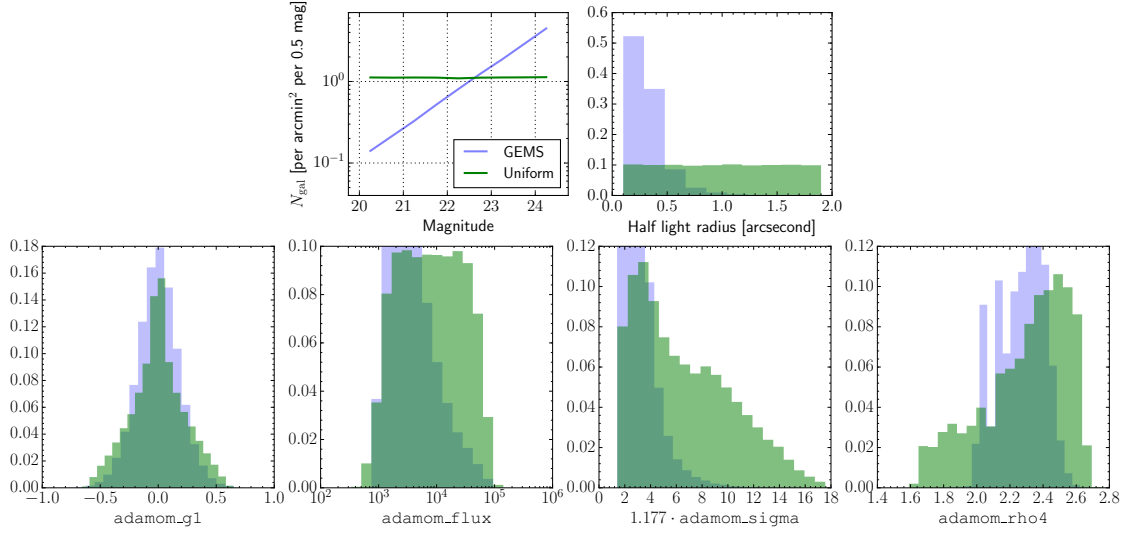


Fig. 13. *Top.* Comparison of the input distributions for dataset catalogues in the *Euclid*-like application. *Down.* Histogram comparison of the input features, normalised by the number of samples. We show only `adamom_g1` as the second ellipticity component is similar. $1.177 \cdot \text{adamom_sigma}$ corresponds to the half light radius of a Gaussian galaxy, and in pixels. The difference of shape in the distribution of half light radius in the uniform distribution can be explained by selection bias. Large but faint galaxies cannot be measured, and removed from the datasets.

tive biases are also consistent with the requirements. Sensitivity to the galaxy parameters are small as well and do not demonstrate consistent trends, except for a small conditional bias with size. The results suggest that the method is not very sensitive to the training distributions, and that there is no need to fine-tune the prior distribution, which is a significant advantage of the method.

7.4. Predicting weighted shear

We turn to predicting weighted shear for a *Euclid*-like dataset. As in Sect. 6.4, we prepare $n_{\text{case}} = 1000$ cases of shear each with 2500 different galaxies viewed with $\text{SNC} = 4$ different orientations for testing. The unweighted shear is learned by the committees described in Sect. 7.2 on a noiseless dataset with $n_{\text{case}} = 5000$, $\text{SNC} = 100$. Weights are learned by two committees of six networks. The weight training set has the same characteristics as in Sect. 6.4: $n_{\text{case}} = 100$, with 100 000 different galaxies, and with only one realization of image. The networks have the same architecture as the unweighted shear networks. They minimize the mean square bias for weights, MSBw. We increase the weights feature set to four inputs:

1. the ellipticity component (`adamom_g1` or `adamom_g2`).
2. the size (`adamom_sigma`).
3. the concentration (`adamom_rho4`).
4. the flux (`adamom_flux`).

Figure 15 shows that the unweighted predictions are significantly biased. As in the fiducial experiment, taking a weighted average of the shear yields more accurate and precise estimates. The remaining biases are non-significant. Weights are distributed mainly as a function of the measured concentration and the measured size. Measurements of galaxies with a high

concentration, i.e. a high Sérsic index, are more biased. The weighting schemes favors well measured galaxies. We reach performances similar to those of the fiducial experiment: small biases and increased precision.

We have shown that predicting a weighted shear directly is possible and reaches the requirements on simple *Euclid*-like data.

8. Application to GREAT3

In this section, we apply our approach on simulations from the GREAT3 challenge, to demonstrate its performance on a reference dataset.

8.1. GREAT3 data

GREAT3 is the latest data analysis challenge for weak-lensing shear measurement algorithms (Mandelbaum et al. 2014, 2015). While the blind phase ended in April 2014, the challenge simulations still provide a welcome benchmark for the evaluation of algorithm, and are publicly available¹³. The GREAT3 data is structured in several branches of somewhat successive complexity. In this work, we focus exclusively on simple branches for which the PSF of each galaxy is provided (in form of a pixelated images), and which mimic single epoch observations. To ease the bias analysis, we also limit ourselves to the so-called “constant-shear” branches, which contain 200 “subfields” with 10 000 galaxy “stamps” and one stationary shear each. We stress that our ML approach draws no advantage from this stationary shear situation, and that exactly the same algorithm can equally well be applied to variable-shear branches. These choices leave

¹³ <http://great3challenge.info>

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

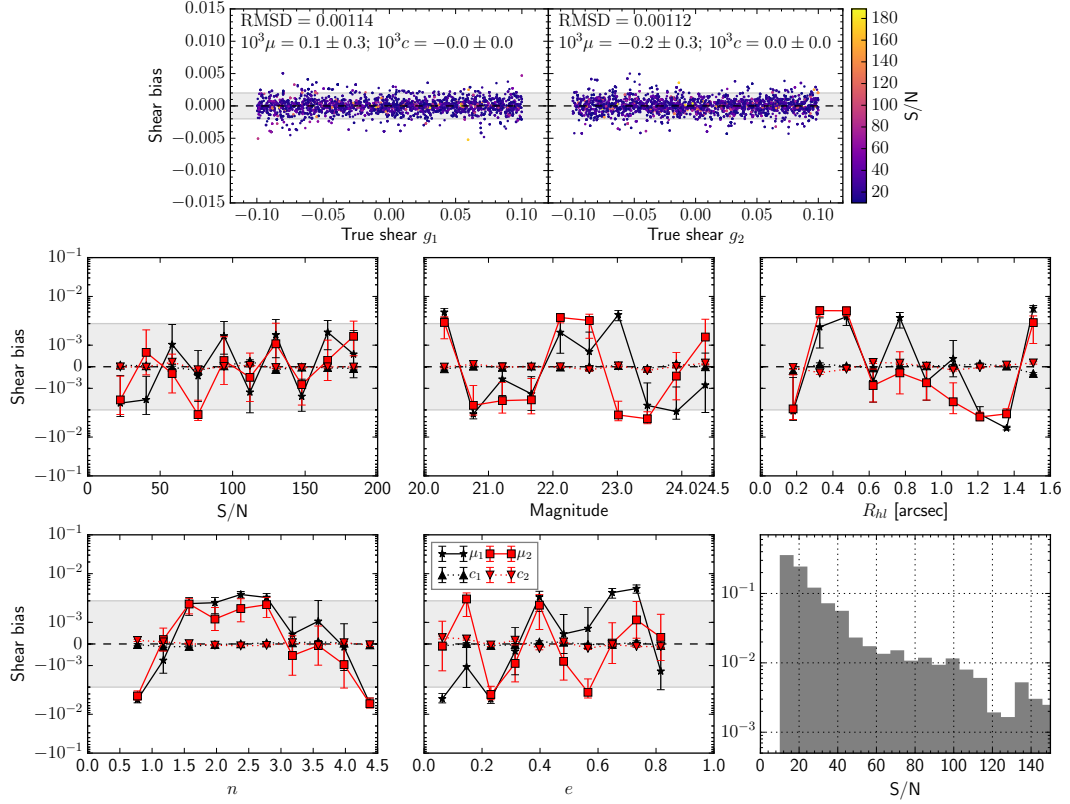


Fig. 14. Remaining shear biases in the case of predicting ellipticities on a *Euclid*-like dataset. Networks were trained on a noisy and *Euclid*-like dataset. The bottom right panel shows the logarithmic normalised distribution of test galaxies S/N.

us with 4 branches, corresponding to the combinations of either ground- or space-based observations of either “control” or real galaxies. The “control” galaxies were drawn for GREAT3 as combinations of two Sérsic profiles (bulge+disk) based on model fits to *Hubble Space Telescope* (HST) observations, while the “real” galaxies use actual HST images. Following the standard GREAT3 nomenclature, the 4 branches we analyse are *control-ground-constant* (CGC), *control-space-constant* (CSC), *real-ground-constant* (RGC), and *real-space-constant* (RSC).

Naturally, to apply our algorithm, we do not use any information that was hidden to the participants during the challenge. In particular, we perform no iterative adjustments based on bias evaluations on the GREAT3 data. To adjust the machine learning setup, we only employ our own internal validation datasets. We therefore stay blind in this experiment.

In the following, we describe how we generate training sets using simple single Sérsic profiles (Sérsic 1963), for all 4 considered branches. The performance on both the parametric bulge+disk models and the real galaxies of GREAT3 therefore gives a first handle on how sensitive our machine learning method is to the accuracy of the training sample. This is a key motivation for using the GREAT3 data.

8.2. Training

Within each GREAT3 branch, the PSF model varies widely between subfields. For example, in space-based branches, subfields have PSFs of different telescopes, with different number of spikes. Attempting to train individual NNs to correct for this large diversity of PSFs within a branch has no motivation from any potential application of the method to real survey data, and would require a large number of features characterizing the PSF shape. We therefore opt for an individual ML training for each subfield, with dedicated training simulations using the PSF of this same subfield. With this choice, we no longer have to inform the NNs about the PSF, as it is identical for every galaxy in a subfield.

For the point-estimate prediction \hat{g}_i and the weights w_i , we use the same 4 input features `adamom.gi`, `adamom.sigma`, `adamom.flux`, and `adamom.rho4`. Indeed both PSF and noise properties remain constant within each subfield, and the ML algorithm therefore does not require galaxy-specific information about these. Compared to a real-life application, this corresponds to the assumption that the ML is optimally informed about the PSF and background noise.

Given that we train the method using galaxy profiles with simple elliptical isophotes, we can simplify the first step of the algorithm, by training a point-estimator for ellipticity instead of

M. Tewes et al.: shear measurement with machine learning

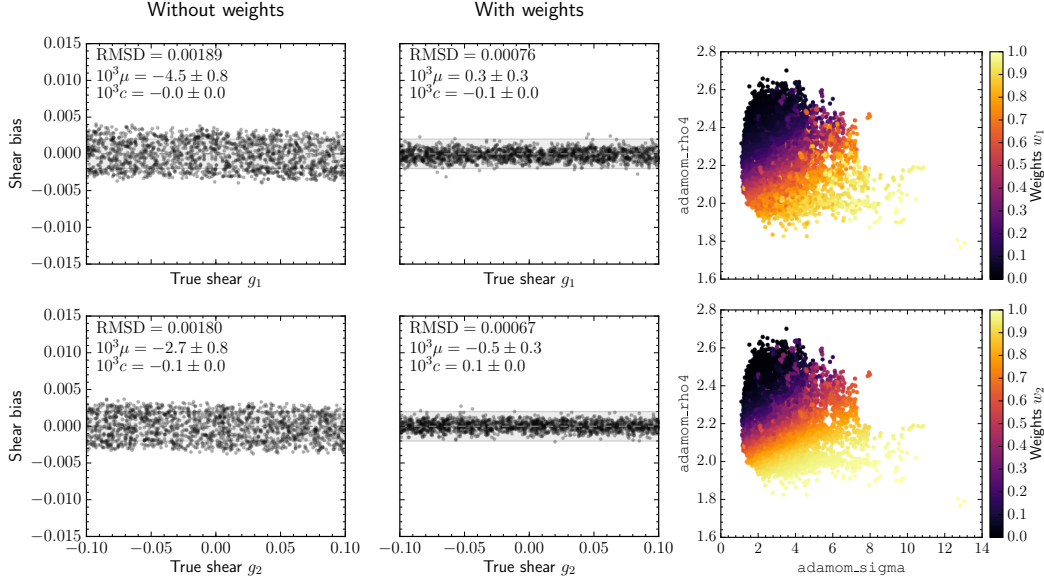


Fig. 15. Weighted shear predictions for a *Euclid*-like dataset. The four left panels show shear bias in function of the true shear for the component one and two when averaging without or with weighting. The rightmost panels show the evolution of the weights.

shear, as described in Section 5.3. For each subfield (i.e., each PSF) of the branches CGC, RGC, CSC, and RSC, we generate a training set with a structure as illustrated in Fig. 4, with $n_{\text{case}} = 1000$ cases and $n_{\text{rea}} = 10$ realizations per case. To train the weights, simulations with the full noise level are required, demanding for a larger training set. We use 200 cases with 1000 realizations. We add stationary Gaussian noise to the stamps used for weight training and validation, with a variance measured on the GREAT3 simulations of the subfield (see Equation 6).

Following the algorithm outlined in Section 5, for each subfield and for each of the two shear components, we train two NN committees: one for the point-estimator, and one for the weight. All committees have 8 members.

Simulation parameters are summarized in Table 2, and some resulting distributions of measured features in 16.

8.3. Results

Figures 17 and 18 show the individual shear measurement errors on the 200 subfields of the CGC and CSC branches, respectively. For each subfield, the predicted g_i is computed as the (weighted) average of the shear estimations obtained for each galaxy. In particular for the ground based branch CGC, the use of weights drastically improves the precision on these average g_i . The PSFs of CGC have a wide range in quality (seeing). For the subfields with very wide PSF, galaxies with small intrinsic extension carry little shear information. The weight-training successfully adapts to these PSFs. No significant multiplicative and additive biases are observed from linear fits to the residuals shown in these figures.

Table 3 presents the complete set of bias measurements, obtained in the very same way. The results labeled with fGC and fSC are obtained from internal validation sets, which

share exactly the same properties and PSFs as the CGC and CSC branches of GREAT3, but use the same single-Sérsic parametrization of the galaxies as used for the training. These results are also visualized in Fig. 19. Arguably, a moderate model bias can be seen, with a tendency towards negative μ for the real galaxy branches.

We show in Fig. 20 an analysis of the quality of the PSF correction achieved in this work, for the CGC branch, to allow a direct comparison with the recent work by (Huff & Mandelbaum 2017). By down-weighting galaxies for which no accurate shear predictions can be obtained, the algorithm achieves a correction with no significant residuals.

To conclude on this application to GREAT3, we recall two reasons why residual biases might remain in the analysis:

1. For both the “real-galaxy” and “control” branches, the galaxy models used for the training (simple Sérsic profiles) differ significantly from the galaxies in the GREAT3 data.
2. The PSF information used for the training is taken from the pixelated images of the PSF model, which have the same (rough) sampling than the galaxy images.

8.4. Evolution of the code since the GREAT3 challenge

We participated in the original GREAT3 in 2014 with an earlier version of the approach, under the name *MegaLUT*. Results presented in this paper are based on a substantial evolution of the algorithm since these original submissions to the GREAT3 challenge. We briefly summarize the main differences between the present work, and the code used in Mandelbaum et al. (2015).

For the GREAT3 challenge participation, we optimized an MSE cost function (with no distinction between realizations and cases), targetting the ellipticity of a galaxy. We did not train any weight prediction. To improve the performance despite this

M. Tewes et al.: shear measurement with machine learning

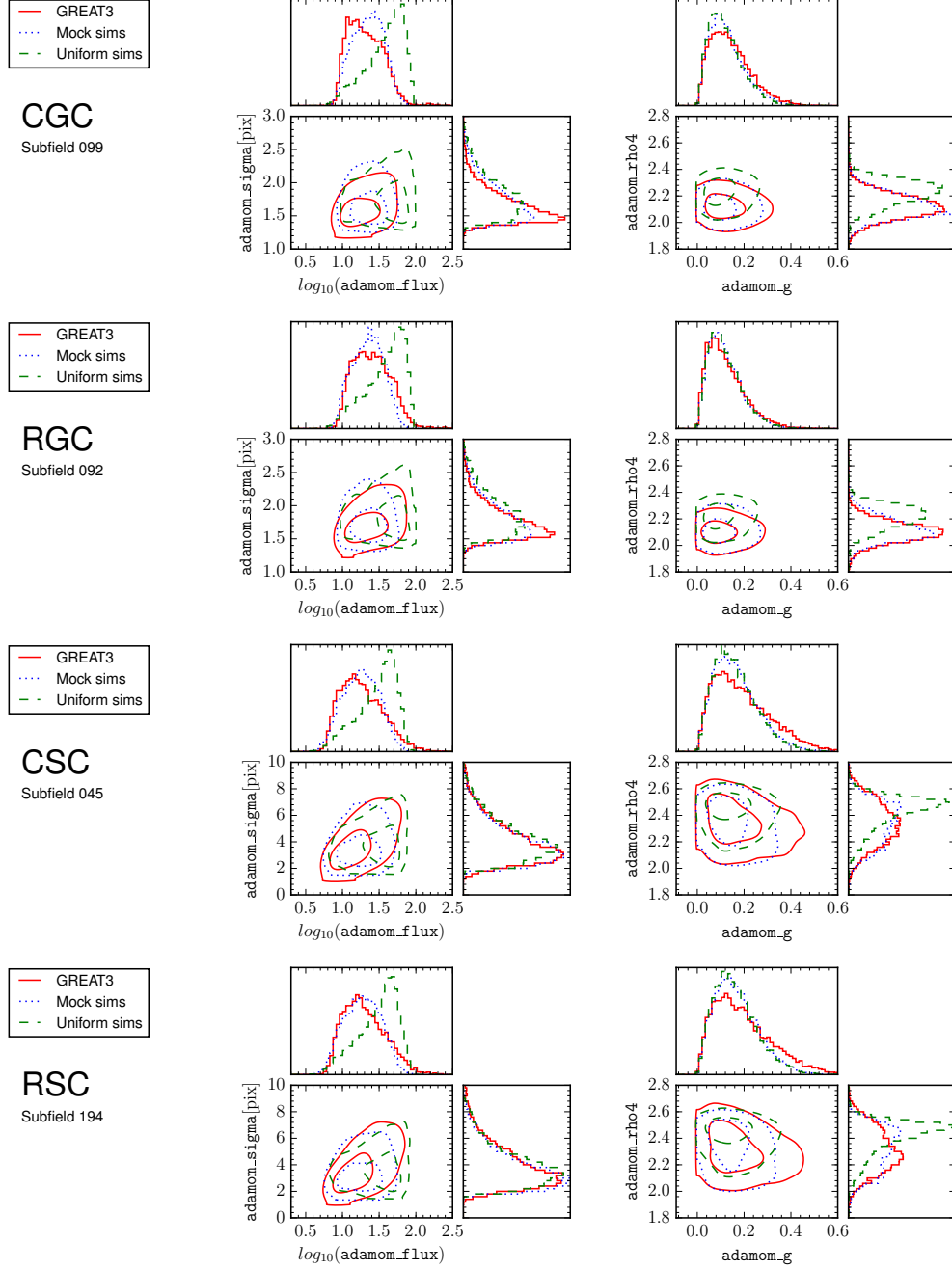


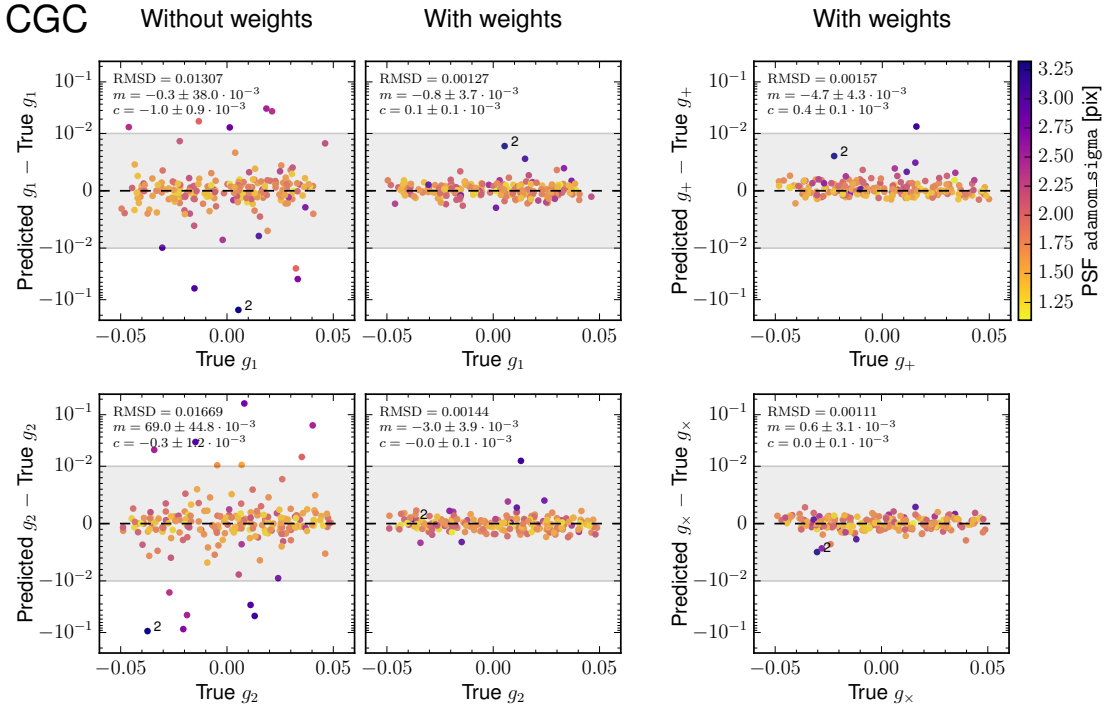
Fig. 16. Measured distributions of galaxy properties in our training simulations and the GREAT3 data, for the subfield with the sharpest PSF in each of the 4 considered branches.

Table 2. Parameter distributions of the training sets generated for each GREAT3 subfield

Branch type:	Ground-based		Space-based	
Training set to learn the prediction of:	Point-estimate	Weight	Point-estimate	Weight
Simulation type:	“Uniform”	“Mock”	“Uniform”	“Mock”
Shear components g_1, g_2	0	$\mathcal{U}(-0.1, 0.1)$	0	$\mathcal{U}(-0.1, 0.1)$
Intrinsic ellipticity modulus e	$\mathcal{R}(0.2)_{[0, 0.7]}$	$\mathcal{R}(0.2)_{[0, 0.7]}$	$\mathcal{R}(0.2)_{[0, 0.7]}$	$\mathcal{R}(0.2)_{[0, 0.7]}$
Sérsic index ^a n	$\mathcal{U}(0.5, 4)$	$\mathcal{U}(0.5, 2.5)$	$\mathcal{U}(0.5, 4)$	$\mathcal{U}(0.5, 4)$
Flux F [counts]	$\mathcal{U}(10, 100)$	$\mathcal{N}(15, 20)_{[10, 200]}$	$\mathcal{U}(10, 100)$	$\mathcal{N}(0, 30)_{[10, 200]}$
Half-light radius R [pix]	$\mathcal{U}(0.75, 3.0)$	$\mathcal{N}(1.0, 0.8)_{[0.75, 3.0]}$	$\mathcal{U}(1.25, 10.0)$	$\mathcal{N}(2.5, 3.5)_{[1.25, 10.0]}$

Notes. We use $\mathcal{U}(a, b)$ to denote the uniform distribution between a and b , $\mathcal{N}(\mu, \sigma)$ to denote a normal distribution with mean μ and variance σ^2 , and $\mathcal{R}(\sigma)$ for a Rayleigh distribution with mode σ . Intervals in subscript denote the range to which we clip a distribution, so that no sample falls outside of the given interval.

^(a) In practice, we grid the values for the Sérsic index instead of drawing them randomly. This significantly speeds up the galaxy stamp generation, as GalSim can reuse cached Sérsic profiles.


Fig. 17. Shear estimation errors on the CGC branch.

flawed approach, we deliberately generated training sets with a higher S/N than the challenge data.

9. Outlook

A number of effects were not included in this study. We list some of the most important in the following, and give our point of view on how to address them in future work.

Color dependency: galaxy colours give rise to important biases in *Euclid*-like images, which PSF is almost purely dominated by diffraction. Spatially-variable SED in the image of galaxies leads to bias multiplicative factor of $\mu \lesssim 0.5 \cdot 10^{-3}$ (Voigt

et al. 2012; Semboloni et al. 2013; Cropper et al. 2013). This additional bias is significant with the next generation of surveys. Ellipticities measured on galaxies with the same shape parameters, but different SED do not necessarily match. For this effect to be taken into account by the ML, it should be included in training simulations. As for the networks accounting for the spatial variability of the PSF across the field, additional input features should be given. Such features could be a representation of the SED of the object, or any colour-related estimate.

Blends: galaxy profiles can be blended with faint undetectable objects, with other detected galaxies or with stars. Light profile blends are affecting the measurement of the shear by

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

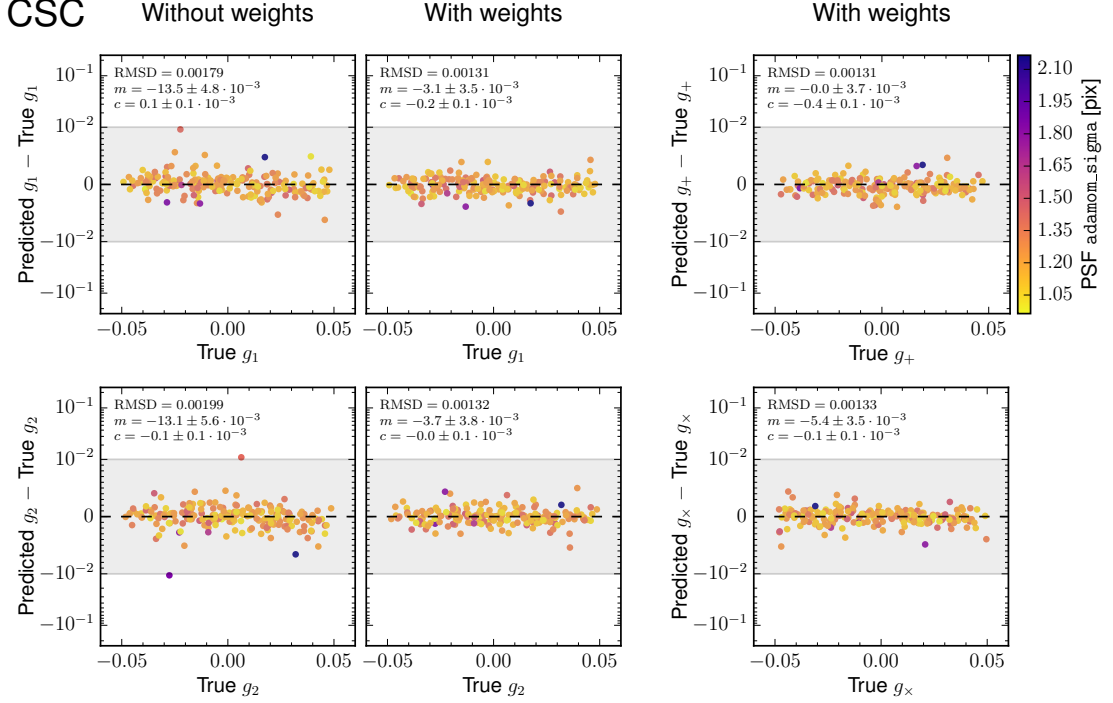


Fig. 18. Shear estimation errors on the CSC branch.

Table 3. Results on GREAT3 constant-shear branches

Branch	$\mu_1 \cdot 10^3$	$\mu_2 \cdot 10^3$	$\mu_+ \cdot 10^3$	$\mu_\times \cdot 10^3$	$c_1 \cdot 10^3$	$c_2 \cdot 10^3$	$c_+ \cdot 10^3$	$c_\times \cdot 10^3$
fGC	-1.50 ± 2.78	$+3.18 \pm 2.04$	$+1.97 \pm 2.28$	-0.30 ± 2.62	$+0.02 \pm 0.08$	$+0.01 \pm 0.06$	$+0.08 \pm 0.07$	-0.02 ± 0.07
CGC	$+4.34 \pm 3.52$	$+1.99 \pm 3.03$	$+1.58 \pm 3.43$	$+3.70 \pm 3.06$	$+0.02 \pm 0.09$	-0.11 ± 0.08	-0.03 ± 0.09	-0.16 ± 0.08
RGC	-5.42 ± 3.82	-4.72 ± 3.42	-3.84 ± 3.34	-7.98 ± 3.79	$+0.12 \pm 0.09$	-0.11 ± 0.08	$+0.29 \pm 0.08$	-0.02 ± 0.09
fSC	-0.97 ± 1.49	-1.87 ± 1.50	-1.55 ± 1.41	-1.18 ± 1.57	-0.04 ± 0.04	-0.00 ± 0.05	$+0.07 \pm 0.04$	-0.02 ± 0.04
CSC	-3.11 ± 3.55	-3.73 ± 3.78	-0.04 ± 3.73	-5.37 ± 3.52	-0.19 ± 0.09	-0.00 ± 0.09	-0.35 ± 0.09	-0.09 ± 0.09
RSC	-5.49 ± 3.15	-6.40 ± 3.15	-3.77 ± 3.08	-7.70 ± 3.25	$+0.04 \pm 0.08$	$+0.04 \pm 0.08$	-0.01 ± 0.08	-0.08 ± 0.08

changing the ellipticity distribution (Dawson et al. 2016), which leads to non-negligible biases (See the impact in DES, Samuroff et al. 2017). Masking could be applied to recognisably blended objects. However, this approach removes a large fraction of objects, since blending affect a large proportion of the population. Moving from a moment measurement approach to a fast multi-object fitting of simple elliptical profiles to the observed images (without forward-modelling any PSF convolution) could be a possible solution, aside an overall statistical calibration.

Image stacking: in this study, we assumed that three *Euclid* exposures could be perfectly co-added. This simplification should be lifted. Co-addition of the single exposures could be implemented, or measurements on the individual images could be processed simultaneously. This latter proposition could entail straightforward averaging features or, e.g. measuring the moments of the different exposures with the same weighting function. Image co-addition should be carefully carried out, notably in terms of PSF quality. Bernstein & Jarvis (2002) discuss the

advantages of the two different approaches, and recommend processing at the catalog level rather than at the image level, because of the PSF effects.

One option is to include the combination process itself in the simulated training set, and to optimize iteratively the hyperparameters. A possible approach could be the control loops described in Refregier & Amara (2014).

Astrometric distortions: these are not currently handled by the code. A solution could be to measure moments in a world coordinate system (WCS) rather than in pixel space, or to convert the shear predictions from pixel to WCS a posteriori.

Realistic training set: training galaxies are modelled with simple Sérsic profiles. More realistic profiles could be introduced to reduce model bias. We detected a possible model bias when applying our scheme to the real galaxy branches of the GREAT3 challenge. More realistic profiles, or simulations based on high-S/N actual observed galaxies could be considered to reduce model bias.

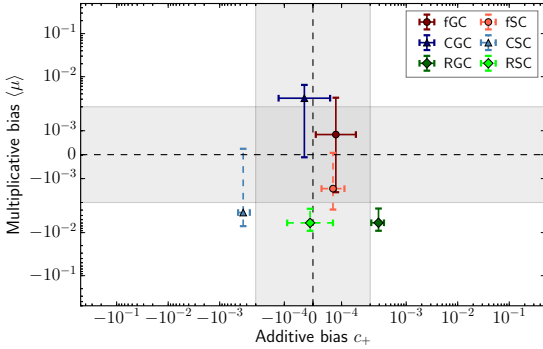


Fig. 19. Multiplicative bias μ averaged over the components 1 and 2, against the additive bias c_+ defined in the coordinate system of the PSF anisotropy. This figure can be directly compared with Fig. 17 of the GREAT3 result paper (Mandelbaum et al. 2015). Darker shades of colors encode ground branches and lighter tones space branches. Space branches are shown with dashed error bars.

Working with probability distribution functions: rather than predicting point estimates of the shear, the method could return probability distributions. The formalism of point estimates and weights, as developed in this paper, is not encoding fully the shear information that can be extracted from galaxy images. To exploit the shear information to its maximum, prediction of continuous shear probability distributions, as represented by Gaussian mixture models, is the next major development stage.

Input features: the current algorithm implements only crude measurements, i.e. we apply no calibration scheme before feeding the features to the networks. In particular, we leave the task of the PSF correction to the networks. If instead we were to apply fast and robust analytical calibrations to the second order moments, the network would only have to correct for the post-calibration biases.

Deep learning: in this study, we decided on the set of input features. These sets were motivated by analytical calibrations schemes developed for adaptive moments (Bernstein & Jarvis 2002; Hirata & Seljak 2003). This manual process could be replaced by the introduction of deep learning (Tuccillo et al. 2017, and references therein): we would let the algorithm decide which feature to use to derive an accurate shear prediction.

10. Summary and conclusions

In this paper, we apply neural networks to the problem of weak lensing shear measurement. This is a calibration problem: we want to retrieve the shear from input features measured on noisy images.

We designed an accurate estimator for the shear in the presence of feature noise. To that end, we have implemented networks which minimize a different cost function than the conventional MSE: the mean square bias. The MSB cost function penalizes the bias in the training set.

The second important aspect of our approach is the construction of the training set: realizations of the same NN ground truths are grouped by case. The bias is measured over the realizations in a case. A learned weighting scheme is devised to foster the insensitivity of the method to distributions of the galaxy param-

eters and to the selection function of the experiment, while increasing the precision. Definitions of the cases and number of realizations are key hyperparameters of the method. The architecture of all of networks is kept small for all applications, only two layers of a few neurons are sufficient to produce state-of-the-art shear measurements.

We applied the method to three experiments: a simple fiducial dataset, *Euclid*-like simulations and the GREAT3 challenge data. In each of the experiments, we showed that we can limit the remaining overall and conditional biases to the stringent targets of the Stage IV experiments. Shear estimates are accurate on resolved galaxies, in particular when the networks are trained on many noisy realizations of the images. The weights associated with each galaxy allow to adjust their importance in the shear measurement. As expected, unresolved galaxies are down-weighted.

We do not foresee any major obstacle in the further development of the method. The complexity of the galaxy profiles and the networks can be increased to accommodate additional effects that were not considered in the present research.

To summarize, our method is advantageous in several respects over other methods: (1) it can return shear estimate without defining the concept of ellipticity, (2) it is fast, (3) it is relatively insensitive to the sample properties of the source galaxies, and (4) new effects can be treated seamlessly as long as we can find sensitive input features. The limitations mainly reside in a careful design of training sets and of realistic enough simulations.

We stress that the aim of this paper is to demonstrate the feasibility of a shape measurement calibration by neural networks rather than to describe a full and finalized pipeline. Obtaining good results on an idealized challenge such as GREAT3 does not allow us to conclude on the readiness for real data. Designing sufficiently good image simulations (in particular, with crowded stamps) for the calibrations to be trustworthy is a separate problem. However, a robust conclusion of the present work is that there is no fundamental and technical barrier in achieving this goal with our new calibration framework.

Acknowledgements. MT acknowledges support by a fellowship of the Alexander von Humboldt Foundation and the DFG grant Hi 1495/2-1. TK is supported by the Swiss National Science Foundation (SNSF). We are thankful to the organizers of the GREAT3 challenge, as well as to all developers of GalSim, for their thorough work. It is also a pleasure to acknowledge heavy use of the community-developed *scipy* (Oliphant 2007), *matplotlib* (Hunter 2007), and *astropy* (Robitaille et al. 2013) packages.

References

- Bartelmann, M. & Maturi, M. 2017, *Scholarpedia*, 12, 32440
- Bernstein, G. M. & Armstrong, R. 2014, *Advances in High Energy Physics*
- Bernstein, G. M. & Jarvis, M. 2002, *Astronomical Journal*, 123, 583
- Bertin, E. & Arnouts, S. 1996, *Astronomy and Astrophysics Supplement*, 117, 393
- Bertin, E. & Arnouts, S. 2010, *Astrophysics Source Code Library*
- Bilicki, M., Hoekstra, H., Amaro, V., et al. 2017
- Bonnett, C. 2015, *MNRAS*, 449, 1043
- Bridle, S., Balan, S. T., Bethge, M., et al. 2010, *MNRAS*, no
- Chromey, F. R. 2010, *To Measure the Sky*
- Cropper, M., Hoekstra, H., Kitching, T. D., et al. 2013, *MNRAS*, 431, 3103
- Cropper, M., Pottinger, S., Niemi, S., et al. 2016, *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, 9904, 99040Q
- Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, *ApJ*, 816, 11
- de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, *A&A*, 582, A62
- D'Isanto, A. & Polsterer, K. L. 2017, eprint arXiv:1706.02467
- Durbin, R. & Rumelhart, D. E. 1989, *Neural Computation*, 1, 133
- Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, 467, 1627
- Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. N. 2013, arXiv, 790

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

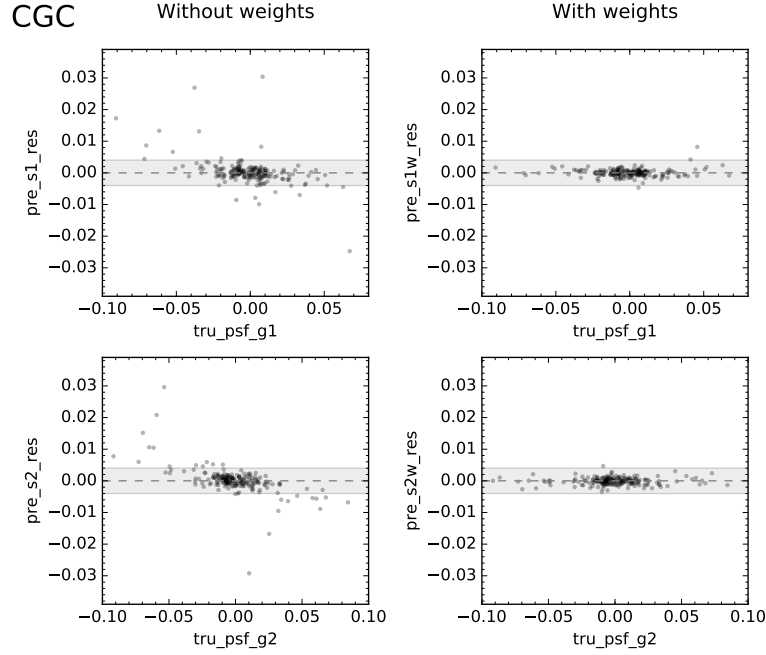


Fig. 20. Shear estimation residuals as function of PSF shape, for the CGC branch. This can be compared directly with Fig 3 from [Huff & Mandelbaum \(2017\)](#).

- Gruen, D., Seitz, S., Koppenhoefer, J., & Riffeser, A. 2010, *ApJ*, 720, 639
- Heymans, C., van Waerbeke, L., Bacon, D., et al. 2006, *MNRAS*, 368, 1323
- Hirata, C. M. & Seljak, U. 2003, *Advances in High Energy Physics*, 343, 459
- Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, *ApJ*, 504, 636
- Hoekstra, H., Viola, M., & Herbonnet, R. 2017, *MNRAS*, 468, 3295
- Huff, E. M. & Graves, G. J. 2013, *ApJ*, 780, L16
- Huff, E. M. & Mandelbaum, R. 2017
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90
- Jee, M. J., Tyson, J. A., Hilbert, S., et al. 2016, *ApJ*, 824, 77
- Kaiser, N. 2000, *ApJ*, 537, 555
- Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, 449, 460
- Kilbinger, M. 2015, *Reports on Progress in Physics*, 78, 086901
- Kitching, T. D., Balan, S. T., Bridle, S., et al. 2012, *MNRAS*, 423, 3163
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *arXiv*, arXiv:1110.3193
- Mandelbaum, R. 2017
- Mandelbaum, R., Hirata, C. M., Leauthaud, A., Massey, R. J., & Rhodes, J. 2012, *MNRAS*, 420, 1518
- Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, *MNRAS*, 450, 2963
- Mandelbaum, R., Rowe, B., Bosch, J., et al. 2014, *The Astrophysical Journal Supplement*, 212, 5
- Massey, R. J., Heymans, C., Berge, J., et al. 2007, *MNRAS*, 376, 13
- Massey, R. J., Hoekstra, H., Kitching, T. D., et al. 2013, *MNRAS*, 429, 661
- Melchior, P. & Viola, M. 2012, *MNRAS*, 424, 2757
- Melchior, P., Viola, M., Schäfer, B. M., & Bartelmann, M. 2011, *MNRAS*, 412, 1552
- Miller, L., Heymans, C., Kitching, T. D., et al. 2013, *MNRAS*, 429, 2858
- Nakajima, R. & Bernstein, G. M. 2007, *Astronomical Journal*, 133, 1763
- Nielsen, M. A. 2015
- Niem, S.-M., Cropper, M., Szafraniec, M., & Kitching, T. D. 2015, *Experimental Astronomy*, 39, 207
- Nocedal, J. & Wright, S. 2006, *Numerical Optimization* (Springer)
- Oliphant, T. 2007, *Computing in Science & Engineering*, 9, 10
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *Astronomical Journal*, 124, 266
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, *Astronomical Journal*, 139, 2097
- Pickles, A. J. 1998, *PASP*, 110, 863
- Refregier, A. 2003, *Annual Review of Astronomy and Astrophysics*, 41, 645
- Refregier, A. & Amara, A. 2014, *Physics of the Dark Universe*, 3, 1
- Refregier, A., Amara, A., Kitching, T. D., et al. 2010, *arXiv*:1001.0061
- Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, *MNRAS*, 425, 1951
- Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, *The Astrophysical Journal Supplement Series*, 152, 163
- Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, *A&A*, 558, A33
- Rousseeuw, P. J. & Croux, C. 1993, *Journal of the American Statistical Association*, 88
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, 10, 121
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, *Scientific American*, 323, 533
- Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2017
- Schmidt, F., Leauthaud, A., Massey, R. J., et al. 2011, *ApJ*, 744, L22
- Schmitt, M. 2002, *Neural Computation*, 14, 241
- Schneider, P., Kochanek, C. S., & Wambsganss, J. 2006, *Saas-Fee Advanced Course*, Vol. 33, *Gravitational Lensing: Strong, Weak and Micro*, ed. G. Meylan, P. Jetzer, & P. North (Springer Verlag)
- Semboloni, E., Hoekstra, H., Huang, Z., et al. 2013, *MNRAS*, 432, 2385
- Sérsic, J. L. 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 6, 41
- Simon, P. & Schneider, P. 2017, *A&A*, 604, A109
- Tagliaferri, R., Longo, G., Milano, L., et al. 2003, *Neural Networks*, 16, 297
- Tewes, M., Cantale, N., Courbin, F., Kitching, T. D., & Meylan, G. 2012, *A&A*, 544, A8
- The Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *Physical Review D*, 94, 022001
- Tuccillo, D., Huertas-Company, M., Decenciere, E., et al. 2017
- Voigt, L. M., Bridle, S. L., Amara, A., et al. 2012, *MNRAS*, 421, 1385
- Zhou, Z.-H., Wu, J., & Tang, W. 2002, *Artificial Intelligence*, 137, 239

Appendix A: Python implementation

We make publicly available the code that we developed in the scope of this paper, including scripts and configuration files generating the presented results and figures, at <https://astro.uni-bonn.de/~mtewes/ml-shear-meas/>.

The core code comes in the form of two separate python packages:

tenbilac is an artificial neural network library for noisy features, implementing the peculiar distinction between training cases and realizations. It is implemented in python and numpy, and agnostic about the particular application to galaxy shape measurement.

megalut is a library providing a toolbox for experimenting with shear and shape estimators, build around GalSim¹⁴ and astropy¹⁵. It includes a simple wrapper to process GREAT3 data, and an interface to **tenbilac**.

Appendix B: Galaxy parameters from GEMS

We prepare parametric models of the distributions by fitting the GEMS galaxies. five parameters are required to determine our *Euclid*-like galaxies: (1) the magnitude, (2) the Sérsic index, (3) the half light radius, (4) the ellipticity, and (5) the position angle. Galaxy count per magnitude are selected from a power-law distribution with a slope of 0.375 and an intercept of -8.250. Sérsic indices are uncorrelated from the other parameters. The log normal Sérsic distribution has a scale $s = 1.2$, and width $\sigma = 1.0$, clipped to a minimum of 0.3 and a maximum of 4.5. Half light radii are drawn from a skewed normal distribution. The parameters distribution of half light radii depends on the magnitude of the object, M . We model the width, σ , and location, ℓ , parameters by linear approximations,

$$\sigma \simeq -0.021 \cdot M + 0.690; \quad \ell \simeq -0.108 \cdot M + 2.074. \quad (\text{B.1})$$

The ellipticity is drawn from a Rayleigh distribution, as explained in eq. (13), and is uncorrelated with the other parameters. Positions angles are uniformly distributed.

Figure B.1 show comparisons between the GEMS survey and our simulated catalogue. We note that the galaxy number count in GEMS sharply decreases for magnitudes fainter than ~ 26 . GEMS is incomplete at these deep magnitudes, but this region is beyond the scope of this paper as we draw galaxies in the range $20 \leq M \leq 24.5$. For our approach, the density number of galaxies is not important as we can generate an arbitrary number of galaxies stamps to train the methods. Generally, machine-learning applications taught with more data translate to better performance. We could train on an ensemble of galaxies larger than the *Euclid* sample first to learn to minimize the biases optimally. The predictions of a smaller ensemble would still conserve the unbiased properties. A large number of galaxies is also needed to reduce the uncertainties on the biases.

Appendix C: *Euclid* VIS-like PSF simulation parameters

In the interest of durable reproducibility, we provide the chosen GalSim parameters. The PSF is simulated with the GalSim.OpticalPSF() method. We pass the following non-zero parameters:

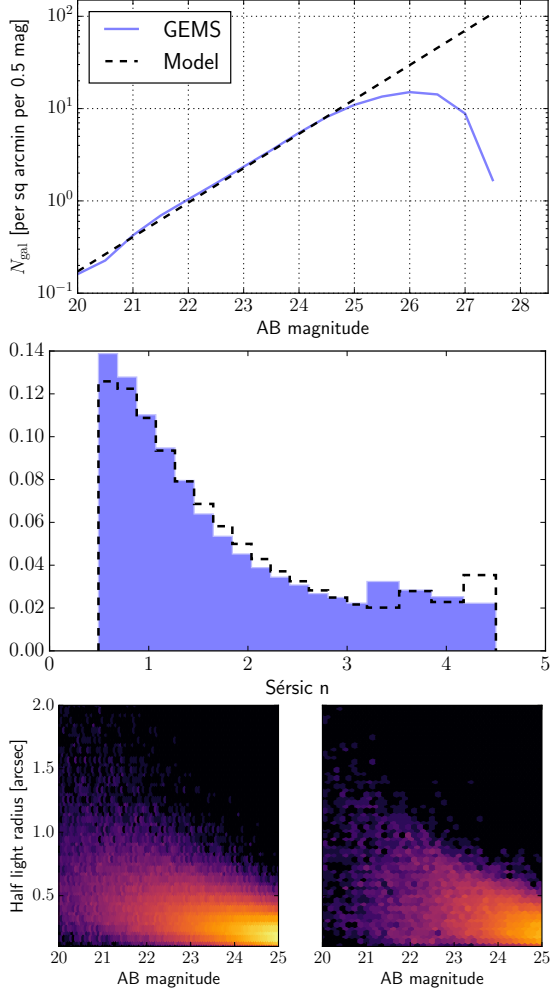


Fig. B.1. Comparisons between the GEMS survey and our simulated catalogue. *Top.* Number density of galaxies per galaxy bins. The model is a simple power law. We note that the number of galaxies in the model is arbitrary in this study as we can simulate an arbitrarily large number of galaxies for training and test. It is selected here to show the goodness of the fit between magnitudes 20 and 25. *Middle.* Fraction of Sérsic indices of the galaxies. GEMS galaxies were modelled as single Sérsic profiles and measured by GALFIT (Peng et al. 2002, 2010). The GEMS indices were cut at 0.5 and 4.5. *Bottom.* Half light radius number density as a function of magnitude. *Bottom left.* Model. *Bottom right.* GEMS survey.

- diam = 1.2
- obscuration = $D_{M2}/D_{M1} = 0.35/1.2 = 0.29$
- nstruts = 6
- strut.thick = 0.015
- strut.angle = 105°
- astig1 = -0.04

¹⁴ <https://github.com/GalSim-developers/GalSim>

¹⁵ <http://www.astropy.org>

5.4. A machine-learning technique for shape measurements

M. Tewes et al.: shear measurement with machine learning

- astig2 = -0.025
- trefoil1 = -0.015
- trefoil2 = 0.008

Astigmatism and trefoil values were selected randomly.

The PSF image is the sum of profiles generated by `GalSim.OpticalPSF()` evaluated every $\lambda_m = 2.5$ nm between 550 and 900 nm, weighted by the spectral energy distribution template of a G5V star from the Pickles library (Pickles 1998). The resulting PSF image is shown in Fig. C.1.

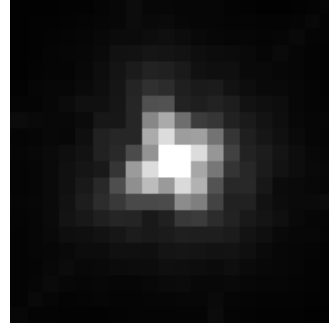


Fig. C.1. GalSim-generated PSF used for the *Euclid*-like simulations. We show the central 20 pixels only at the *Euclid* resolution. The image is displayed in logarithmic scale.

5.4.9 Outlooks

A machine-learning approach, such ours, is promising for the extremely demanding (not so) future surveys. This method is well set on its path to be integrated in the nominal *Euclid* reduction pipeline. Even if its development is still actively ongoing, the proof-of-concept work presented in Paper IV shows the potential of the method. We note that in the 2013-2014 version of the method, the one used for GREAT3, networks were trained with a mean square cost function, and resulted in biases of the order of $\mathcal{O}(0.5 - 1\%)$. The introduction of MSB has allowed to reduce this bias to below the *Euclid* requirements. Other refinements of the optimisation, like addition regularisation terms, could be explored.

Future investigations will concentrate on increasing the complexity of the model (by, e.g. producing probability distributions rather than point estimates), the simulations (by including more effects or introducing more realistic galaxy profiles), and the features (in particular by analytically correcting for the PSF before feeding the feature set to the networks). The custom ANN library we coded to implement the MSB cost function should be improved to take full advantage of modern computing techniques. The list of the future work presented in Paper IV is long, and will require important computing resources. We do not foresee any insurmountable obstacles however: the most important effects that are still be included (like colour dependencies and blends) can be readily simulated with libraries like `GalSim`.

6 Finding galaxy-scale strong lenses

Overview

By design, large surveys in Stages III and IV cover a large fraction of the sky. Weak-lensing requires deep and high resolution images. *Euclid* alone covers about a third of the sky with a pixel scale of 0.1 arcsecond. In the previous chapters, we extensively discussed the exquisite quality of the PSF, so there is no need to brag about it once more. High quality images are will be systematically searched. This search yields the material for weak lensing: galaxies and stars. With the right tools, this immense database can deliver many other interesting objects, such as galaxy-scale strong gravitational lenses.

This chapter can be seen as special topic. Its focus is on a different regime of gravitational lensing: strong. In particular finding galaxy-scale strong lenses in forthcoming data. Even though we change paradigm, we still make use of the same tools: artificial neural networks. Ever since the impressive results of CNNs on the ImageNet challenges (see Sect. 4.3.4), deep learning has gain traction in many communities. We describe a strong lens (SL) discovery challenge, which we entered with several implementations of CNNs, discuss their performance and propose a path forward.

6.1 A data avalanche

The number of known (strong) lenses is small, of the order of a few hundreds. Most of them were serendipitously discovered as a human stumbled upon lens images. Confirming the strongly lensed nature of a source is not trivial; it requires spectral comparison of the different light bundles thought to be originating from the same source. The images of the lensed sources and the profile are often angularly close. Hence, discovering new strong lenses demands high resolution. One of the largest catalogue of SL contains a mere 131 objects, discovered in the Sloan Lens ACS Survey (SLACS) ([Bolton et al. 2008](#)).

[Collett \(2015\)](#) predicted the number of new lenses harvested by large-area surveys. DES is foreseen to yield about $2.5 \cdot 10^3$ lenses. Numbers of discoverable lenses in future surveys are

even larger: LSST $\sim 1.2 \cdot 10^5$ and *Euclid* $\sim 1.7 \cdot 10^5$. These estimates are based on realistic simulations of SL populations and should be interpreted as upper limits. They still need to be found in the images. The weak-lensing survey in *Euclid* is designed to observe thirty to forty galaxies per arcmin square. The number density of lenses is roughly $10^{-4} - 10^{-5}$ smaller than the galaxy number density. Two conclusions can be drawn from this ratio: (i) the number of objects to analyse is very large, and (ii) strong lenses are rarities. In data analysis terms, it translates to (i) a requirement to be fast to process and (ii) extremely skewed classes, an issue similar to confirming rare diseases. A highly skewed class can simply be completely ignored by a traditional classifier, because it will still be rewarded with a very low cost. Possible solutions to imbalance in a dataset (He and Garcia 2009) include (i) re-sampling the training set, by artificially augmenting the number of strong lenses and using less negative images. This solution is chosen by the lensing community as we can simulate as many images as wanted. (ii) Changing the cost function to the analysis of the confusion matrix or the ROC. Finally, (iii) treat SL searches as an anomaly detection problem. Pre-processing of the data could exclude unlikely candidates. The imbalance problem is not limited to SL searches; finding pulsars¹ suffers from the same issues. Some of the possible solutions listed previously were applied to the pulsar classification techniques (see, e.g., Lyon et al. 2016). As in pulsar searches, SL detection methods focus on reducing the FPR.

Scientific motivations for this hunt for data are outlined in Sect. 2.5.1. The impressive growth of the SL catalogue will enable more precise science, transforming the field from a pioneering era to maturity. The avalanche of new objects will eventually cascade to follow-up proposals and data analysis pipelines. An example: time-delay measurements. Until now, the number of lenses whose time delay is well characterised is limited to $\mathcal{O}(10^1)$ (Treu and Marshall 2016). With the foreseeable dramatic increase in volume (and in depth), the current observing facilities will not be sufficient. There are plans for proposing an ESA M5 (medium mission, to be launched in ~ 2030) primed for probing dark matter properties, studying the nature of compact object and fathoming the solar neighbourhood for Earth-like planets. The *Theia* mission² (Theia Collaboration 2017) could dedicate a significant fraction of its time to staring at lens systems. Light curves with no interruption would represent an important increase in data quality. On the data analysis front, the determination of time delays between the different sources images will have to be automated. A community challenge (Strong Lens Time Delay Challenge; Dobler et al. 2015) was launched in 2015 to assess the quality of such algorithms. Lastro entered the competition with a team running the PyCS curve-shifting method (Tewes et al. 2013). To ensure good performances of the algorithm, an initial guess was required. Since then, crude delay determination techniques were designed, but at the time it was decided to estimate this delay by visual inspection (Bonvin et al. 2016). Each volunteer (amongst them, the author of this thesis) spent $\sim 30 - 60$ s per curve on average for the 5000 simulated curves. Many teams entering this challenge, including ours, had the vivid realisation that automatic methods were an absolute necessity to deal with the data volume of the not-so-future surveys.

¹Pulsars are fast rotating neutrons stars. They emit energy in very focused beams. When the beam path crosses the Earth, the received signal has a lighthouse-like form, extremely constant in frequency.

²<http://theia.phyip3.dur.ac.uk/>

6.2 Automated search methods

The development of galaxy-scale finding methods mirrors somewhat the stages of machine-learning research (Sect. 4.1). Methods worked their way up from rule-based algorithms to deep learning. In this short section, we give some historical context to automated search methods.

6.2.1 Early methods

The epoch of visual inspection by human experts screening through thousand of images nears its end. While still a legitimate technique to vet candidates, it is no longer widely used to discover new lenses in data. In this section, we look at two examples of methods that have been applied to CFHTLS legacy data and both revealed new lenses. The first one is very much based on selecting objects based on empirical rules. The second takes a more data-driven approach.

Early systematic searches used the fact that a lensed source is brighter because of lensing magnification bias. A pre-processing technique was to look for multiple sources around visually bright quasars (e.g. [Surdej et al. 1987](#)). RingFinder ([Gavazzi et al. 2014](#)) is typical of early methods. The method first selects bright elliptical galaxies, which are seen as efficient gravitational lenses. Galaxy type is estimated by SED fitting. The bulk of the lens galaxy is then removed. Subtracting the bulk is carried out with a simple technique: getting the image of the bulk in a filter sensitive mostly to the bulk, scaling it to the original filter and removing. If the foreground galaxy is the lensing object, blue features should appear. Cuts in colour are applied as criteria to detect a potential lens. If the detected sources resembled rings or arcs, the galaxy was flagged as a potential strong lens. The method returned more than 2 500 candidates for CFHTLS. Visual inspection by multiple experts assessed the probability of being a lens. The list was reduced to 330 possible lenses, amongst them about seventy new objects. Follow-up confirmation campaigns yielded thirty-three gravitational lenses, but not all possible lenses were tested. Extrapolating to the whole survey, CFHTLS data should contain about 220 lenses. This method was later extended from a two-band analysis to a five-band lens modelling ([Brault and Gavazzi 2015](#)). While having successfully found new objects, the authors claim that this method can only be a stepping stone towards automated searches, as it would require too much human resources, at least ~ 5000 person hours to go through the dataset once.

The algorithm by [Joseph et al. \(2014\)](#) proposed to build a representation of selected galaxies using PCA decomposition. Modelled profiles can subsequently be subtracted from the image. The goal of this stage is the same than in the previous method: to reveal lensing features in the residual images. [Paraficz et al. \(2016\)](#) applied the PCA technique to CFHTLS data using a different method for identifying these features. They used topometric clustering algorithms to identify potential rings and arcs. Identified objects were fitted to assess if the objects could be lensed ([Tramacere et al. 2016](#)), introducing physics in the data treatment. 1 098 candidates met the lens criteria, that we later visually inspected. A fraction of $\sim 10\%$ of the candidates

were classified as definite or probable lenses.

When comparing the results of the different lens finders (see [Paraficz et al. 2016](#), and references therein), it appeared that the catalogues did not match. This means that some algorithms are more sensitive than others to certain types of lenses. There are many traps for the algorithms, like ring galaxies or artefacts, that look very much like lenses. Combining classifications from different approaches, which are not sensitive to the same features, could mitigate misclassification. In each of the methods presented here, there is a definite lack of versatility when searching for lensing features. This issue will be corrected in the next section. A citizen-science project³ was launched to propose to a large number of people to classify lenses. This initiative could also be a solution not to overlook peculiar lens configurations.

6.2.2 Detecting strong lenses in upcoming surveys

While the two above methods did uncover new lenses, they quickly showed their limitations: model bias and too much human supervision. Astronomers were convinced to use what they first regarded as black boxes: deep-learning techniques. CNNs only recently entered the field, but with undeniable success and superiority over previous approaches.

[Petrillo et al. \(2017\)](#) trained SL classifiers that were applied to the 255 square degrees of the KiDS survey, and their selection of 22 000 galaxies. They chose to consider only massive galaxies, luminous red galaxies (LRGs), which represent 80% of the lensing population (see e.g. [Möller et al. 2007](#)). Out of that sample, 761 candidates were found (amongst them two of the three known KiDS lenses). The algorithm was trained to recognise only fairly large Einstein radii features. The known lens that was missed by the CNN has an Einstein radius smaller than the training criterion, and thus could not be detected. With widening of the pre-selection criteria and smaller Einstein ring lenses, they claim that $\sim 2.4 \cdot 10^3$ lenses could be found. [Petrillo et al. \(2017\)](#) simulated a balanced training set: three millions non-lens examples and three million lens examples. 3.6% of the KiDS selected galaxy images were classified by the network as lenses, which were then inspected by humans. The architecture of the CNN is relatively simple: four convolution and two 1 024-neurons fully-connected layer leading to one output, the confidence of detection, and the binary cross-entropy cost function. Their main focus for future work is to reduce the false positives. Their aim at a FPR of 0.1% without decreasing the TPR significantly. Galaxies exhibiting ring-like structures were the largest contaminants.

[Jacobs et al. \(2017\)](#) applied CNNs to CFHTLS data, with a balanced training set of $\sim 1.2 \cdot 10^5$ simulated images. They trained four different networks to form a committee, which returned about 2 500 candidates out of $1.4 \cdot 10^6$ objects, 2 100 of which were false positives. Amid the detected objects, 117 were already known, 29 were confirmed new lenses and 266 potential lenses. They estimate a completeness, that is the fraction of the lenses that are in the predicted

³The Space Warps project ([Marshall et al. 2016](#)), <https://spacewarps.org/>

sample, of roughly a quarter, and a false positive rate of 0.1%. Note that this FPR is what to [Petrillo et al. \(2017\)](#) team is striving to achieve. They used two different architectures: two or three convolutional layers followed by two fully-connected layers (4 096 or 1 024 neurons respectively), with a softmax output layer. Ensemble learning was advantageous as it allowed to remove some false positives from the sample. This does represent a progress from the previous method. There is still the need for visual inspection, but the low FPR induces less images to vet.

The results quoted previously are for initially catalogued objects. A step forward would be to search all images, without extracting postage stamps of known objects, this increases the completeness and the probability of finding a unconventional lens. [Jacobs et al. \(2017\)](#) performed such a search, by extracting stamps with overlapping sliding window scheme. However, the blind search scheme proved less efficient and much more computationally expansive. The probability of finding unusual lenses is not high either because they were not included in the training set.

The even more recent technique of residual networks described in [He et al. \(2015\)](#) was proposed by [Lanusse et al. \(2017\)](#). An architecture of one convolutional layer and fifteen residual layer, each containing three convolutions, was applied to LSST simulated data. Almost perfect classification scores were obtained. The authors concluded that performances of their network were at least as good as [Petrillo et al. \(2017\)](#) when they compared on simulated data. They conclude that the realism of the simulation is not sufficient to discriminate between the two methods.

6.3 Deep convolutional networks for SL detection

The methods of the previous section build on the better understanding of galaxy-scale lensing, which translates into image simulations, and a tremendous effort in pattern recognition. We reviewed CNN-based methods, because they appeared in parallel, with the different groups working almost independently from each other. In this section, we will discuss a similar approach developed at Lastro and share lessons learned.

6.3.1 Strong lensing finding challenge

The *Euclid* Strong Lensing working group, in collaboration with the Bologna lens factory⁴, wary of the future data volumes, launched the Galaxy-Galaxy Strong Lensing challenge (GGSLC) (Metcalf et al, in prep.). This challenge contains two types of data: space-borne *Euclid*-like and ground-based KiDS-like, thus two categories. Both branches images are 101 pixels on a side. The space branch images are in one band with a 0.1 arcsec and a 0.18 arcsec FWHM Gaussian, essentially VIS-like. Ground images are in four bands (SDSS *u*, *g*, *r* and *i*) with a 0.2 arcsec pixel size, include masked regions to simulate artefacts and show a lower S/N range.

⁴<https://bolognalensfactory.wordpress.com/>

The size of the provided training sets are only 20 000 images. The training sets were available for the entire duration of the challenge, between November 2016 and February 2017. It was a blind challenge: methods were evaluated on the 100 000-strong test set. Here again the sets were balanced between the two classes. A team had twenty-four hours to process the test set and submit the predicted catalogue. This catalogue was not binary, but contained confidence values $p \in [0, 1]$ for each image. An object with a high confidence value was interpreted to be a lens. Submitting a confidence value rather than the classification allows to build ROC curves, and compare methods marginalising over the choice of non-lens/lens threshold. These community-wide challenges are viewed by many as accelerators for the development of methodologies and might phase out obsolete methods. Thirteen entries out of the total twenty-four for both branches were based on deep learning. Four others were based on a combination of human-designed features and machine learning. Seven entries were part of earlier classes of methods, including visual inspection. Those older methods consistently ranked lower.

Lastro entered the competition with CNN-based methods. The team was composed by Marco Geiger, Lastro master student who implemented the method and ran the pipeline. He was followed by PhD student Christoph Schäfer and Lastro director Jean-Paul Kneib for the lensing side. The author of this manuscript advised on machine-learning aspects.

6.3.2 CNN architectures

Four different architectures were applied to GGSCL: a conventional CNN architecture that we will define as our baseline model, a residual network, and two others tries to implement more invariance in the networks.

Baseline architecture, as shown in Fig. 6.1, is modelled on typical ImageNet architectures. The baseline architecture is comprised of eight convolutional layers, organised into three convolution blocks, two stand-alone layers, and three fully-connected (fc) layers as the last stage. There is a total of eleven layers. With the exception of the initial one, every convolution layer uses 3×3 convolution kernels for efficiency reasons. The first convolution layer uses a 4×4 kernel to yield an even number of pixels for easier manipulation. At each convolution block, the number of features were doubled, resulting in 256 features in the last block. The number of neurons per fc layer is kept to a reasonable number, hence a relatively low number of model parameters. Comparing to the other techniques studied in Sect. 6.2.2, our architecture contains more convolutional layers and more fc layers. This simple baseline architecture achieved first place, out of fourteen entries, in the space component of the GGSCL challenge.

For each layer, a modified version of the ReLU activation function is used. This choice is motivated by the sparsity of its activation, which is beneficial in (at least natural) image

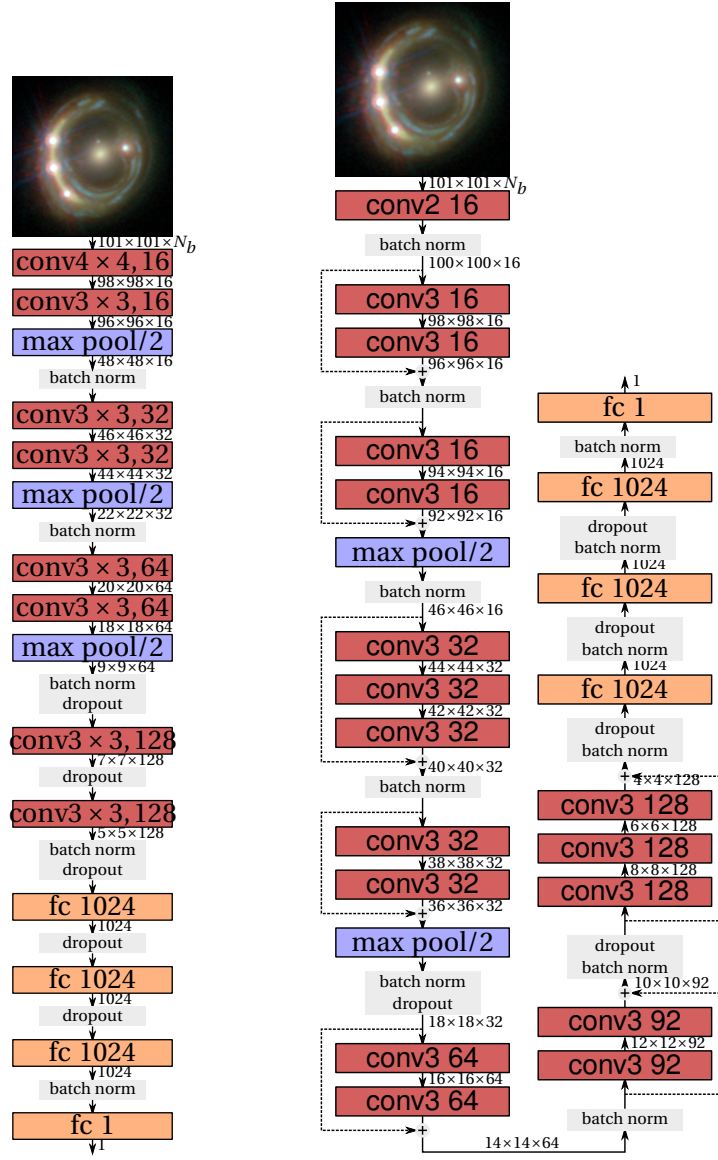


Figure 6.1 – Visualisation of the baseline and residual architecture for the CNN lensfinder. The convolution blocks (red) indicate the size of the kernel and the number of features. The fully-connected blocks (yellow) indicate the number of features. The arrows indicate the flow of the data and between the blocks we show the dimensionality of the input ($N_{\text{pixel}} \times N_{\text{pixel}} \times N_{\text{features}}$). The initial layer has either one or four features, depending on the category of the data (space and ground, respectively). Batch normalisation and dropout layers are indicated as grey blocs. (Left.) Baseline architecture. (Right.) Residual architecture.

classification (Glorot et al. 2011). The activation is given by

$$\sigma(z) = \frac{1}{\sqrt{\pi-1}} \left(\sqrt{2\pi} \max(0, z) - 1 \right). \quad (6.1)$$

Inputs of the networks have dimension of $101 \times 101 \times N_b$, where N_b is the number of bands

($N_b = 1$ for space and $N_b = 4$ for ground branches). The wavelength-dependent information (in the third dimension) is handled naturally by extending the kernel dimension from two to three (spatial to spatial plus wavelengths).

Residual architectures were also studied for our lens finder. A common way for improving CNN is to increase the depth, that is the number of convolutional layers a network has. When increasing the depth of CNNs, the unstable gradient problem (Sect. 4.3.3) arises and can become important. At some point in the training process, the accuracy starts to saturate and degrade, generating an upper limit to the possible depth of CNNs. He et al. (2015) introduced residual learning that mitigates the problem. The Lanusse et al. (2017) team, discussed in the previous section, came out first in the ground branch of the challenge. Our residual architecture is shown in Fig. 6.1. It is 20-layer deep with five two-convolution residual blocks, two three-convolution residual blocks and three fully-connected layers with 1 024 features. The two-step residual block is made of two convolutions and one short-cut pathway. The number of features is kept constant. The three-step residual block is composed of three convolutions and one short-cut followed by a convolution layer, doubling the number of features.

The idea behind the two remaining architectures is to exploit the invariant features of gravitational lenses. CNNs are, by design, already invariant to translation but not to rotation, scaling and flipping. The data augmentation operations (see later paragraph on training) allowed to teach some symmetries by including them in the training set, but it does not increase the invariance of the models. We present two techniques which aim at increasing model invariance.

The first of these models, dubbed **views** creates a committee of networks with two members trained to accomplish the same task, detecting lenses, on different scales. The committee architecture is presented in Fig. 6.2. A first network spans at the whole image detecting big lenses spanning the whole image. A second network explores the central part of the image. By combining the prediction of the two networks, smaller lenses should be detected while not neglecting the detection of the larger lenses. In other words, the first network takes as input the whole image, like the baseline model, while the second only accepts a smaller 45×45 pixels stamp. The network treating the smaller stamp uses only 5×5 kernels and less features at each layer. This reduces the number of model parameters, as the number of input features is drastically reduced from $101 \times 101 \sim 10^4$ to $45 \times 45 \sim 2 \cdot 10^3$.

The **invariance architecture** adds additional invariant properties into the model. While relatively untested, this has been used with success (Dieleman et al. 2015) for a galaxy morphology classifier on Galaxy Zoo⁵ data (Lintott et al. 2008). The invariant architecture takes advantage of the dihedral⁶ symmetries present in the lens-finding problem. At the level of the input layer, eight operations of the same convolution kernel, transformed by a different transformation

⁵<https://www.galaxyzoo.org/>

⁶Rotations by 90° , 180° , 270° and four reflections.

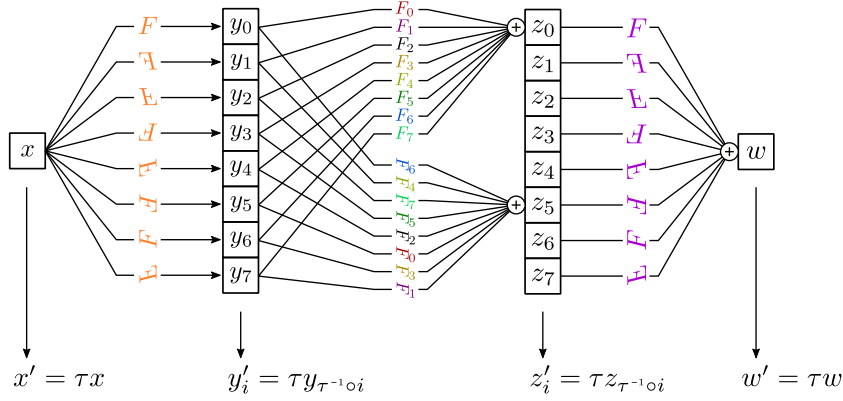


Figure 6.3 – Dihedral equivariant architecture. Kernels with identical colours but different orientations are identical kernels to which a different dihedral operation has been applied. (Phase 1.) Separation into eight channels, one for every input channel and member of the dihedral group. (Phase 2.) Convolution of the eight channels with eight separate kernels. (Phase 3.) The eight channels are added together giving a dihedral invariant result.

a dihedral invariant quantity. This is illustrated in Fig. 6.3. Our invariant architecture is shown in Fig. 6.2 and follows the same fundamental scheme as the baseline architecture. Since using eight channels increases computation time and makes the model more prone to overfitting, the number of features is divided by four. The invariance has been tested by checking that rotated and flipped versions of the same image are attributed the same score by the classifier.

Training and implementation

We chose to write the code using the TensorFlow library (Abadi et al. 2015). In all four implementations, we used the binary cross-entropy cost function to drive training. 20 000 examples for training could lead to overfitting as the ratio between the number of examples to model parameter is low. We thus performed data augmentation on the training set. Rotations by 90° and 180° were performed as well as reflections leading to a multiplication of the set by a factor of eight. Rotations different than a multiple of 90° were not performed to avoid having to interpolate in pixel space. The original training set was divided into an effective training set (of 17 000 images) and a validation set (3 000 images). The first set, the effective training set, was augmented and used by the models to learn from, while the validation set was used to check the training performances on an independent set. The performance was monitored every 1 000 steps by evaluating predictions made on the validation set. At each training step, we randomly selected batches of thirty images (with a balanced set) and run the learning procedure for ~ 300 epoches using the ADAM minimization algorithm (Kingma and Ba 2014). The models were trained on a GeForce GTX 1060 Graphic card. The training time took ~ 1 hour per 100 epochs for the baseline model and 2 hours/100 epochs for the residual model. The final prediction of the classification for the challenge on the 100 000 test images took ~ 20 minutes.

6.3.3 Performance of the models and outlooks on future development

A detailed study of results is presented in [Schaefer et al. \(2017\)](#). The two invariant and the residual models were not completed on time to be considered in the GGSLC rankings. We entered the baseline architecture only. The simplest, baseline, architecture proved to be the best classifier in the space branch, and achieved third place in the ground branch. While we obtained a better ranking on the space branch, the performance is actually worse: the AUC metric is of the order of 93% for space data and 97% for ground data. The AUC metric shows an excellent performance ($> 99\%$) on both branches during training. There is a discrepancy between the AUC values evaluated on the validation and test set of two to four percentage points, which can be explained by the small size of the training and validation sets. The better performance for KiDS-like data is likely due to more information in the form of the four bands, instead of the single VIS-like band for space data.

We evaluated the performance of the two invariant and the residual models after the completion of the competition. We ran five instances of each architecture. The four architectures return the same AUC values within standard deviations. There are some indications that invariant and residual methods could outperform the baseline method, especially on the space branch (see Fig. 6.4). This first stroke us as surprising as deeper networks, like the residual one, were expected to perform better than shallower models. We realised that most likely explanation was that the simulations were too simple to confidently distinguish between the architectures. The simulations did not include spiral galaxies or other ring-like objects acting as lens finder traps. Simple methods were sufficient to classify correctly the data of the GGSLC challenge.

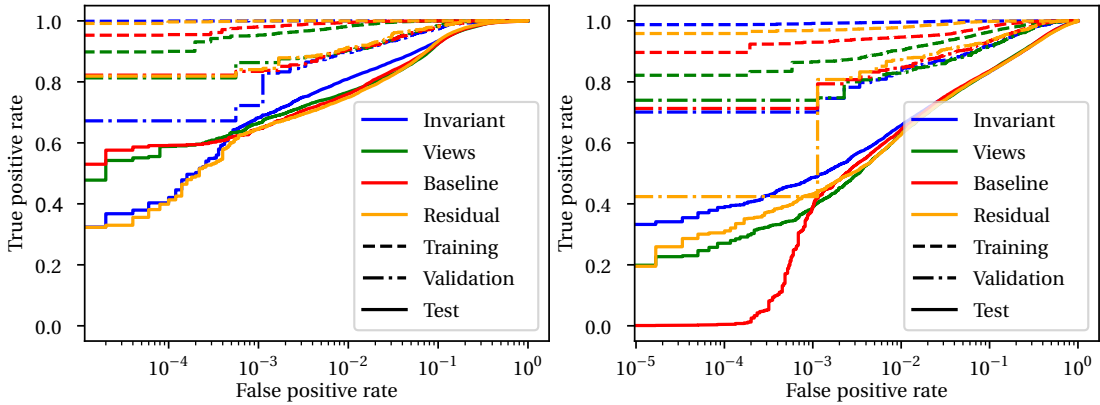


Figure 6.4 – Logarithmic ROC curves of the different architectures. Training (dotted line), validation (half dotted line) and test (solid line) score of all four architectures. Data comes from the best of five runs in terms of validation set score. (*Left.*) Ground KiDS-like data. (*Right.*) Space VIS-like data.

We note that the GGS LC challenge provided the already cut and masked image stamps. The methods we designed are thus catalogue-based techniques. The performances obtained in this section confirm the conclusions of other CNNs methods: deep learning is a promising option for galaxy-scale lenses detections. The community has now clearly demonstrated the validity of the approach, and should focus its resources on solving the major issue: the realism of the training set. Simulations are the limiting factor in the development of the methods, and the community should stride towards producing realistic images with many lensing configurations and include the many traps of real data, like ring galaxies. Improving the training set will reduce the number of false positives and enable the detection of a wide range of lens morphology. Future developments of the automated searches will have to study whether a blind search (opposed to catalogue-based) is advantageous from a point of view of (i) finding more unusual lenses and (ii) computing time.

7 Conclusions and outlooks

This study aims at providing tools for a more accurate weak lensing reduction pipeline. As we have seen, there are several major challenges that need to be solved to unlock the full potential of future surveys. The two main challenges are (i) a good reconstruction of the PSF and (ii) an unbiased shape measurement scheme. The importance of these aspects translates into stringent performance requirements imposed by, e.g., the space-borne mission *Euclid*. The weak lensing field appeared in the early years of the 21st century. At its dawn, weak lensing was limited by statistical uncertainties. With the current and future surveys, the field has moved to a systematics-dominated era. PSF correction and other shape measurement biases have however always been at the centre of attention. Despite deploying many resources to tackle these issues, state-of-the-art techniques do not consistently meet the requirements on full complexity simulations. We propose paths forward to both conundrums by introducing machine-learning techniques.

To properly correct for the PSF effects, their profile must be reconstructed to a high degree of fidelity. PSF models can be built from first principles and a deep knowledge of the architecture of the telescope. This approach however does not incorporate all of the key aspects of a space-borne telescope, like the effect of the launch on the optics, or of ageing. Complementing physical models by the measurement of the effective system PSF on observed point-like sources is a necessity. Stars are excellent approximation for point sources. Many objects however resemble stars, like small galaxies or unresolved multiple stars. The stellar candidates must be properly vetted to quench bias sources, and classified by spectral type in order to account for shape measurement colour biases stemming from broad observation filters.

The first contribution of the thesis is to understand the PSF determination biases due to multiple stars, and the development of vetting techniques to mitigate these biases. We have shown that unresolved multiple stars are a non-negligible source of bias in the *Euclid* framework, because of the stringent error budget and well controlled other sources of noise. This bias arises from the depth and resolution of *Euclid* images. We propose a multiple-star flagging scheme based on a statistical analysis of stellar profile parameters that could remove at least half of the most performance-degrading objects.

The second contribution is in the field of stellar classification: attributing spectral type without making use of multiple bands information. While this proposed method does not have the same performance as traditional methods, it allows to determine types to half a MK spectral class. Analysis of stellar profiles takes advantage of the simple relationship between the wavelength of the signal and the image resolution. This technique could be used in a *Euclid*-like pipeline to quickly determine the colour of the object, without waiting for the multi-band analysis. It can also be used as point-like objects vetting scheme by separating stars from galaxies. We have shown that the approach is also working on real HST data by reconstructing colour-magnitude diagrams.

The third contribution, and the last in the field of PSFs, is the reconstruction of the PSF itself based on machine-learning tools. Similarly to previous attempts at PSF reconstruction, we find a new representation of the profiles. This representation must be sparser than the original image, and provide a smooth coefficient field. These conditions allow to interpolate the PSF coefficients at any coordinate in the image. The task of finding the new base is left to unsupervised algorithm: auto-encoders. An advantage of this approach is that the PSF spatial, spectral, and temporal variations can be easily incorporated. The current approach is purely data-driven, but has demonstrated at different level of simulation complexity, including on stellar images that show wavelength-dependence, that it can reach the *Euclid* requirements.

The fourth contribution consists in an accurate shear measurement scheme. As demonstrated in this manuscript, measuring unbiased shear is hard. The approach does not propose a novel measurement method *per se*, but rather an calibration scheme using committees of ANNs. The technique, MegaLUT, can in principle accept inputs from any shape measurement method, but we chose to apply it in particular to adaptive second order moments. A decisive step forward in the development of the method was the change of the cost function from the classical mean-square error to the mean-square bias. This latter cost function essentially minimises the error marginalising over different realisations of the same galaxy image. This helps beat down the feature noise. An important result of the study is that the training and test distributions of parameters do not need to be similar. We demonstrated that training on simple parameter distributions yields the same performance as when training and testing on the same distributions. The scheme is tested on different simulated datasets, including on the GREAT3 challenge, and yields good performances. *Euclid*-like requirements are met for the datasets tested, including the case of a spatially-dependent PSF. As for the PSF reconstruction scheme, effects can be seamlessly added to the method by including them to the simulation, or by introducing relevant input features to the networks.

As we demonstrated in this thesis, individual modules of the weak lensing pipeline can reach the stringent constraints on datasets that are simpler than real *Euclid* data. Efforts to continue the developments of these algorithms and other modules must be sustained in the future. Consolidating data-driven PSFs with physical information may guarantee reaching the PSF requirements in the real *Euclid* data. On the shape measurement front, the results show real potential, but the method still need to incorporate complicated effects, like profile blending,

image artefacts, wavelength-dependency and multi-epoch analysis. Including these effects will lengthen the simulation and learning times, as the volume of data needed to correct the induced biases will be large.

We note that shape measurement is not the only hard problem in weak lensing analysis. Determining accurate redshifts is key to producing accurate estimation of cosmological parameters. As spectroscopic redshift determination necessitate large resources, most of the redshifts will be determined photometrically. Producing accurate photometric redshifts is at least as difficult as weak lensing (e.g. [Mandelbaum 2017](#), and reference therein). Important strides are made in this field using, again, machine-learning techniques. While beyond the scope of this thesis, photo-z determination is an important parallel branch in the cosmological parameter determination pipeline.

The advent of machine-learning techniques, and neural networks in particular, prompted new research in fields where the burden of long standing problems (like accurate shape measurements) was curbing the scientific enthusiasm. The novel methods that are emerging are stirring communities and start producing scientific results (e.g. good strong lensing finders, the fifth contribution of this thesis).

The commissioning date of Stage IV experiments, like LSST or *Euclid*, is fast approaching. This implies that the imperative to integrate the different modules into a consistent pipeline is increasing. To do so, the maturity of the algorithms, in both scientific and coding terms should be ramped up. Putting efforts into creating a reduction pipeline would also allow end-to-end error analysis. This would prompt studies to understand the effects of low-level biases (such as, e.g., the PSF ellipticity or binary bias) on the inferred cosmological parameters and their uncertainties. In turns, this enables deciding on priorities and devoting resources to the most important developments.

Bibliography

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z. et al (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- Abbott, T., Abdalla, F.B., Allam, S., Amara, A., Annis, J. et al (2016). Cosmology from cosmic shear with dark energy survey science verification data. *Phys. Rev. D*, 94:022001.
- Abdalla, F.B., Banerji, M., Lahav, O. and Rashkov, V. (2011). A comparison of six photometric redshift methods applied to 1.5 million luminous red galaxies. *MNRAS*, 417:1891–1903.
- Abt, H.A. and Levy, S.G. (1976). Multiplicity among solar-type stars. *ApJS*, 30:273–306.
- Albrecht, A., Bernstein, G., Cahn, R., Freedman, W.L., Hewitt, J. et al (2006). Report of the Dark Energy Task Force. *ArXiv Astrophysics e-prints*, astro-ph/0609591.
- Alcock, C., Allsman, R.A., Alves, D., Ansari, R., Aubourg, É. et al (1998). EROS and MACHO Combined Limits on Planetary-Mass Dark Matter in the Galactic Halo. *ApJ*, 499:L9–L12.
- Alcock, C., Allsman, R.A., Alves, D.R., Axelrod, T.S., Becker, A.C. et al (2000). The macho project: Microlensing results from 5.7 years of large magellanic cloud observations. *The Astrophysical Journal*, 542(1):281.
- Allen, J. (1972). *Apollo 15 Preliminary Science Report, Summary of Scientific Results*, volume SP-289. NASA.
- Alsing, J., Heavens, A. and Jaffe, A.H. (2017). Cosmological parameters, shear maps and power spectra from cftlens using bayesian hierarchical inference. *Monthly Notices of the Royal Astronomical Society*, 466(3):3272.
- Alsing, J., Kirk, D., Heavens, A. and Jaffe, A.H. (2015). Weak lensing with sizes, magnitudes and shapes. *Monthly Notices of the Royal Astronomical Society*, 452(2):1202.
- Amendola, L., Appleby, S., Avgoustidis, A., Bacon, D., Baker, T. et al (2016). Cosmology and Fundamental Physics with the Euclid Satellite. *ArXiv e-prints*, 1606.00180.
- Amendola, L., Frieman, J.A. and Waga, I. (1999). Weak gravitational lensing by voids. *Monthly Notices of the Royal Astronomical Society*, 309(2):465.
- An, J.H., Evans, N.W., Kerins, E., Baillon, P., Novati, S.C. et al (2004). The anomaly in the candidate microlensing event pa-99-n2. *The Astrophysical Journal*, 601(2):845.
- Anderson, J., King, I.R., Richer, H.B., Fahlman, G.G., Hansen, B.M.S. et al (2008). Deep advanced camera for surveys imaging in the globular cluster ngc 6397: Reduction methods. *The Astronomical Journal*, 135(6):2114.
- Andrae, R., Melchior, P. and Jahnke, K. (2011). Quantifying galaxy shapes: sérsiclets and beyond. *MNRAS*, 417:2465–2477.
- Arnouts, S., Cristiani, S., Moscardini, L., Matarrese, S., Lucchin, F., Fontana, A. and Giallongo, E. (1999). Measuring and modelling the redshift evolution of clustering: the hubble deep field north. *Monthly*

Bibliography

- Notices of the Royal Astronomical Society*, 310(2):540.
- Bacon, D.J., Goldberg, D.M., Rowe, B.T.P. and Taylor, A.N. (2006). Weak gravitational flexion. *Monthly Notices of the Royal Astronomical Society*, 365(2):414.
- Bacon, D.J., Refregier, A.R. and Ellis, R.S. (2000). Detection of weak gravitational lensing by large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 318(2):625.
- Barbera, F.L., de Carvalho, R.R., Gal, R.R., Busarello, G., Merluzzi, P., Capaccioli, M. and Djorgovski, S.G. (2005). Color gradients in early-type galaxies: Dependence on environment and redshift. *The Astrophysical Journal Letters*, 626(1):L19.
- Bartelmann, M. and Maturi, M. (2016). Weak gravitational lensing. *ArXiv e-prints*, 1612.06535.
- Bartelmann, M., Narayan, R., Seitz, S. and Schneider, P. (1996). Maximum-likelihood Cluster Reconstruction. *ApJ*, 464:L115.
- Bartelmann, M. and Schneider, P. (1992). A large-scale structure model for gravitational lensing. *A&A*, 259:413–422.
- Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *Physics Reports*, 340(4–5):291 – 472.
- Bartelmann, M., Viola, M., Melchior, P. and Schäfer, B. M. (2012). Calibration biases in measurements of weak lensing. *A&A*, 547:A98.
- Bate, M.R. (2009). Stellar, brown dwarf and multiple star properties from hydrodynamical simulations of star cluster formation. *MNRAS*, 392:590–616.
- Bate, M.R. (2012). Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation. *MNRAS*, 419:3115–3146.
- Beck, R., Lin, C.A., Ishida, E.E.O., Gieseke, F., de Souza, R.S., Costa-Duarte, M.V., Hattab, M.W. and Krone-Martins, A. (2017). On the realistic validation of photometric redshifts. *Monthly Notices of the Royal Astronomical Society*, 468(4):4323.
- Benabed, K. and Van Waerbeke, L. (2004). Constraining dark energy evolution with gravitational lensing by large scale structures. *Phys. Rev. D*, 70:123515.
- Bengio, Y. (2009). Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127.
- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *ArXiv e-prints*, 1206.5533.
- Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H. (2006). Greedy layer-wise training of deep networks. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, NIPS’06, pages 153–160, Cambridge, MA, USA. MIT Press.
- Bergé, J., Gamper, L., Réfrégier, A. and Amara, A. (2013). An ultra fast image generator (ufig) for wide-field astronomy. *Astronomy and Computing*, 1:23 – 32.
- Bernard, D. and LeClair, A. (2013). Scrutinizing the cosmological constant problem and a possible resolution. *Phys. Rev. D*, 87:063010.
- Bernhard, G. and Schneider, P. (1999). A simultaneous maximum likelihood approach for galaxy-galaxy lensing and cluster lens reconstruction. *Monthly Notices of the Royal Astronomical Society*, 302(1):118.
- Bernstein, G. and Jain, B. (2004). Dark energy constraints from weak-lensing cross-correlation cosmography. *The Astrophysical Journal*, 600(1):17.
- Bernstein, G.M. (2010). Shape measurement biases from underfitting and ellipticity gradients. *MNRAS*, 406:2793–2804.
- Bernstein, G.M. and Armstrong, R. (2014). Bayesian lensing shear measurement. *Monthly Notices of the*

- Royal Astronomical Society*, 438(2):1880.
- Bernstein, G.M., Armstrong, R., Krawiec, C. and March, M.C. (2016). An accurate and practical method for inference of weak gravitational lensing from galaxy images. *Monthly Notices of the Royal Astronomical Society*, 459(4):4467.
- Bernstein, G.M. and Jarvis, M. (2002). Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing. *AJ*, 123:583–618.
- Bertin, E. and Arnouts, S. (1996). Sextractor: Software for source extraction. *Astron. Astrophys. Suppl. Ser.*, 117(2):393–404.
- Bertone, G. and Hooper, D. (2016). A History of Dark Matter. *ArXiv e-prints*, 1605.04909.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Number 1. Springer-Verlag New York.
- Blandford, R.D. and Narayan, R. (1992). Cosmological applications of gravitational lensing. *ARA&A*, 30:311–358.
- Blandford, R.D., Saust, A.B., Brainerd, T.G. and Villumsen, J.V. (1991). The distortion of distant galaxy images by large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 251(4):600.
- Bolton, A.S., Burles, S., Koopmans, L.V.E., Treu, T., Gavazzi, R., Moustakas, L.A., Wayth, R. and Schlegel, D.J. (2008). The sloan lens acs survey. v. the full acs strong-lens sample. *The Astrophysical Journal*, 682(2):964.
- Bolzonella, M., Miralles, J.M. and Pelló, R. (2000). Photometric redshifts based on standard SED fitting procedures. *A&A*, 363:476–492.
- Bond, I.A., Bennett, D.P., Sumi, T., Udalski, A., Suzuki, D. et al (2017). The lowest mass ratio planetary microlens: Ogle 2016–blg–1195lb. *Monthly Notices of the Royal Astronomical Society*, 469(2):2434.
- Bonnet, H. and Mellier, Y. (1995). Statistical analysis of weak gravitational shear in the extended periphery of rich galaxy clusters. *A&A*, 303:331.
- Bonnett, C. (2015). Using neural networks to estimate redshift distributions. an application to cfhtlens. *Monthly Notices of the Royal Astronomical Society*, 449(1):1043.
- Bonvin, V., Courbin, F., Suyu, S.H., Marshall, P.J., Rusu, C.E. et al (2017). H0licow – v. new cosmograil time delays of he 0435-1223: H0 to 3.8 per cent precision from strong lensing in a flat Λ cdm model. *Monthly Notices of the Royal Astronomical Society*, 465(4):4914.
- Bonvin, V., Tewes, M., Courbin, F., Kuntzer, T., Sluse, D. and Meylan, G. (2016). Cosmograil: the cosmological monitoring of gravitational lenses - xv. assessing the achievability and precision of time-delay measurements. *A&A*, 585:A88.
- Brainerd, T.G., Blandford, R.D. and Smail, I. (1996). Weak Gravitational Lensing by Galaxies. *ApJ*, 466:623.
- Braut, F. and Gavazzi, R. (2015). Extensive light profile fitting of galaxy-scale strong lenses - towards an automated lens detection method. *A&A*, 577:A85.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Bridle, S., Balan, S.T., Bethge, M., Gentile, M., Harmeling, S. et al (2010). Results of the great08 challenge: an image analysis competition for cosmological lensing. *Monthly Notices of the Royal Astronomical Society*, 405(3):2044.
- Bridle, S., Shawe-Taylor, J., Amara, A., Applegate, D., Balan, S.T. et al (2009). Handbook for the great08 challenge: An image analysis competition for cosmological lensing. *Ann. Appl. Stat.*, 3(1):6–37.
- Brown, M., Bacon, D., Camera, S., Harrison, I., Joachimi, B. et al (2015). Weak gravitational lensing with the Square Kilometre Array. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 23.

Bibliography

- Browne, I.W.A., Wilkinson, P.N., Jackson, N.J.F., Myers, S.T., Fassnacht, C.D. et al (2003). The cosmic lens all-sky survey - ii. gravitational lens candidate selection and follow-up. *Monthly Notices of the Royal Astronomical Society*, 341(1):13.
- Bruderer, C., Chang, C., Refregier, A., Amara, A., Bergé, J. and Gamper, L. (2016). Calibrated ultra fast image simulations for the dark energy survey. *The Astrophysical Journal*, 817(1):25.
- Bruna, J., Mallat, S., Bacry, E. and Muzy, J.F. (2015). Intermittent process analysis with scattering moments. *Ann. Statist.*, 43(1):323–351.
- Bruna, J., Szlam, A. and LeCun, Y. (2013). Learning Stable Group Invariant Representations with Convolutional Networks. *ArXiv e-prints*, 1301.3537.
- Bull, P., Akrami, Y., Adamek, J., Baker, T., Bellini, E. et al (2016). Beyond Λ cdm: Problems, solutions, and the road ahead. *Physics of the Dark Universe*, 12:56 – 99.
- Burke, W.L. (1981). Multiple Gravitational Imaging by Distributed Masses. *ApJ*, 244:L1.
- Cacciato, M., Van Den Bosch, F.C., More, S., Li, R., Mo, H.J. and Yang, X. (2009). Galaxy clustering and galaxy–galaxy lensing: a promising union to constrain cosmological parameters. *Monthly Notices of the Royal Astronomical Society*, 394(2):929–946.
- Cain, B., Schechter, P.L. and Bautz, M.W. (2011). Measuring Gravitational Lensing Flexion in A1689 Using an Analytic Image Model. *ApJ*, 736:43.
- Calchi Novati, S., Mirzoyan, S., Jetzer, P. and Scarpetta, G. (2013). Microlensing towards the smc: a new analysis of ogle and eros results. *Monthly Notices of the Royal Astronomical Society*, 435(2):1582.
- Campbell, W.W. and Trumpler, R. (1923). Observations on the deflection of light in passing through the sun's gravitational field, made during the total solar eclipse of september 21, 1923. *Publications of the Astronomical Society of the Pacific*, 35(205):158.
- Casaponsa, B., Heavens, A.F., Kitching, T.D., Miller, L., Barreiro, R.B. and Martínez-González, E. (2013). Size magnification as a complement to cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 430(4):2844.
- Cavuoti, S., Brescia, M., Tortora, C., Longo, G., Napolitano, N.R. et al (2015). Machine-learning-based photometric redshifts for galaxies of the eso kilo-degree survey data release 2. *Monthly Notices of the Royal Astronomical Society*, 452(3):3100.
- Cavuoti, S., Tortora, C., Brescia, M., Longo, G., Radovich, M., Napolitano, N.R., Amaro, V. and Vellucci, C. (2017). Cooperative photometric redshift estimation. *ArXiv e-prints*, 1701.08120.
- Chalela, M., Gonzalez, E.J., Garcia Lambas, D. and Foëx, G. (2017). Compact group analysis using weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 467(2):1819.
- Chang, C., Jarvis, M., Jain, B., Kahn, S.M., Kirkby, D. et al (2013). The effective number density of galaxies for weak lensing measurements in the lsst project. *Monthly Notices of the Royal Astronomical Society*, 434(3):2121.
- Chantavat, T., Sawangwit, U. and Wandelt, B.D. (2017). Void profile from planck lensing potential map. *The Astrophysical Journal*, 836(2):156.
- Chouard, T. (2016). The go files: Ai computer wraps up 4-1 victory against human champion. *Nature*.
- Chwolson, O. (1924). Über eine mögliche form fiktiver doppelsterne. *Astron. Nachr.*, 221:329–330.
- Clampitt, J. and Jain, B. (2015). Lensing measurements of the mass distribution in sdss voids. *Monthly Notices of the Royal Astronomical Society*, 454(4):3357.
- Clampitt, J., Miyatake, H., Jain, B. and Takada, M. (2016). Detection of stacked filament lensing between sdss luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 457(3):2391.
- Clampitt, J., Sánchez, C., Kwan, J., Krause, E., MacCrann, N. et al (2017). Galaxy–galaxy lensing in the

- dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 465(4):4204.
- Clevert, D.A., Unterthiner, T. and Hochreiter, S. (2015). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *ArXiv e-prints*, 1511.07289.
- Clowe, D., Bradač, M., Gonzalez, A.H., Markevitch, M., Randall, S.W., Jones, C. and Zaritsky, D. (2006). A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters*, 648(2):L109.
- Clowe, D., Gonzalez, A. and Markevitch, M. (2004). Weak-lensing mass reconstruction of the interacting cluster 1e 0657–558: Direct evidence for the existence of dark matter. *The Astrophysical Journal*, 604(2):596.
- Clowe, D., Luppino, G.A., Kaiser, N. and Gioia, I.M. (2000). Weak lensing by high-redshift clusters of galaxies. i. cluster mass reconstruction. *The Astrophysical Journal*, 539(2):540.
- Cole, S., Percival, W.J., Peacock, J.A., Norberg, P., Baugh, C.M. et al (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362:505–534.
- Colless, M., Peterson, B.A., Jackson, C., Peacock, J.A., Cole, S. et al (2003). The 2dF Galaxy Redshift Survey: Final Data Release. *ArXiv Astrophysics e-prints*, astro-ph/0306581.
- Collett, T.E. (2015). The population of galaxy–galaxy strong lenses in forthcoming optical imaging surveys. *The Astrophysical Journal*, 811(1):20.
- Collister, A.A. and Lahav, O. (2004). ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, 116:345–351.
- Coupon, J., Broadhurst, T. and Umetsu, K. (2013). Cluster Lensing Profiles Derived from a Redshift Enhancement of Magnified BOSS-survey Galaxies. *ApJ*, 772:65.
- Crittenden, R.G., Natarajan, P., Pen, U.L. and Theuns, T. (2002). Discriminating weak lensing from intrinsic spin correlations using the curl-gradient decomposition. *The Astrophysical Journal*, 568(1):20.
- Cropper, M., Hoekstra, H., Kitching, T., Massey, R., Amiaux, J. et al (2013). Defining a weak lensing experiment in space. *Monthly Notices of the Royal Astronomical Society*, 431(4):3103.
- Cropper, M., Pottinger, S., Niemi, S., Azzollini, R., Denniston, J. et al (2016). Vis: the visible imager for euclid.
- Cummings, J.D., Geisler, D., Villanova, S. and Carraro, G. (2014). Uncovering multiple populations with washington photometry. i. the globular cluster ngc 1851. *The Astronomical Journal*, 148(2):27.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.
- Cypriano, E.S., Amara, A., Voigt, L.M., Bridle, S.L., Abdalla, F.B., Réfrégier, A., Seiffert, M. and Rhodes, J. (2010). Cosmic shear requirements on the wavelength dependence of telescope point spread functions. *Monthly Notices of the Royal Astronomical Society*, 405(1):494.
- Dark Energy Collaboration, Abbott, T., Abdalla, F.B., Aleksić, J., Allam, S. et al (2016). The dark energy survey: more than dark energy – an overview. *Monthly Notices of the Royal Astronomical Society*, 460(2):1270.
- Dawson, W.A., Schneider, M.D., Tyson, J.A. and Jee, M.J. (2016). The ellipticity distribution of ambiguously blended objects. *The Astrophysical Journal*, 816(1):11.
- de Bruijne, J. (2012). Science performance of gaia, esa's space-astrometry mission. *Astrophysics and Space Science*, 341(1):31–41.
- de Laplace, P.S. (1798). *Exposition du système du monde*. 2nd edition edition.
- Debono, I. and Smoot, G.F. (2016). General relativity and cosmology: Unsolved questions and future

Bibliography

- directions. *Universe*, 2(4).
- Del Popolo, A. and Le Delliou, M. (2017). Small Scale Problems of the Λ CDM Model: A Short Review. *Galaxies*, 5:17.
- Dell’Antonio, I.P. and Tyson, J.A. (1996). Galaxy dark matter: Galaxy-galaxy lensing in the hubble deep field. *The Astrophysical Journal Letters*, 473(1):L17.
- Demetroullas, C. and Brown, M.L. (2016). Cross-correlation cosmic shear with the sdss and vla first surveys. *Monthly Notices of the Royal Astronomical Society*, 456(3):3100.
- D’Eugenio, F., Houghton, R.C.W., Davies, R.L. and Dalla Bonta, E. (2015). On the distribution of galaxy ellipticity in clusters. *Monthly Notices of the Royal Astronomical Society*, 451(1):827.
- Dieleman, S., Willett, K.W. and Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459.
- Dietrich, J.P. and Hartlap, J. (2010). Cosmology with the shear-peak statistics. *Monthly Notices of the Royal Astronomical Society*, 402(2):1049.
- Dietrich, J.P., Werner, N., Clowe, D., Finoguenov, A., Kitching, T., Miller, L. and Simionescu, A. (2012). A filament of dark matter between two clusters of galaxies. *Nature*, 487:202–204.
- Dobler, G., Fassnacht, C.D., Treu, T., Marshall, P., Liao, K., Hojjati, A., Linder, E. and Rumbaugh, N. (2015). Strong lens time delay challenge. i. experimental design. *The Astrophysical Journal*, 799(2):168.
- Doersch, C. (2016). Tutorial on Variational Autoencoders. *ArXiv e-prints*, 1606.05908.
- Dubath, P., Apostolakis, N., Bonchi, A., Belikov, A., Brescia, M. et al (2016). The euclid data processing challenges. *Proceedings of the International Astronomical Union*, 12(S325):73–82.
- Duchêne, G. and Kraus, A. (2013). Stellar Multiplicity. *ARA&A*, 51:269–310.
- Duquennoy, A., Mayor, M. and Halbwachs, J.L. (1991). Multiplicity among solar type stars in the solar neighbourhood. I - CORAVEL radial velocity observations of 291 stars. *A&AS*, 88:281–324.
- Duvet, L., Gondoin, P. and Saavedra Criado, G. (2015). Euclid science requirements document. Technical Report issue 4, revision 0, ESA. DEM-SA-Dc-00001.
- Dyson, F.W., Eddington, A.S. and Davidson, C. (1920). A determination of the deflection of light by the sun’s gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 220(571-581):291–333.
- Ebbels, T., Kneib, J.P. and Ellis, R. (1999). Weak Galaxy-Galaxy Lensing in HST DATA. In Sato, K., editor, *Cosmological Parameters and the Evolution of the Universe*, volume 183 of *IAU Symposium*, page 247.
- Eddington (1919). Discussion on the theory of relativity. *Monthly Notices of the Royal Astronomical Society*, 80(2):96.
- Efstathiou, G. and Lemos, P. (2017). Problems with KiDS. *ArXiv e-prints*, 1707.00483.
- Eifler, T., Schneider, P. and Krause, E. (2010). Measuring cosmic shear with the ring statistics. *A&A*, 510:A7.
- Eigenbrod, A., Courbin, F., Meylan, G., Agol, E., Anguita, T., Schmidt, R. W. and Wambsganss, J. (2008). Microlensing variability in the gravitationally lensed quasar qso 2237+0305 \equiv the einstein cross - ii. energy profile of the accretion disk. *A&A*, 490(3):933–943.
- Einstein, A. (1915). Zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 778-786.
- Einstein, A. (1936). Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84(2188):506–507.
- Eisenstein, D. (2005). Dark energy and cosmic sound. *New Astronomy Reviews*, 49(7):360 – 365.

- Wide-Field Imaging from Space.
- Eisenstein, D.J., Zehavi, I., Hogg, D.W., Scoccimarro, R., Blanton, M.R. et al (2005). Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560.
- El-Badry, K., Rix, H.W., Ting, Y.S., Weisz, D.R., Bergemann, M., Cargile, P., Conroy, C. and Eilers, A.C. (2017). Signatures of unresolved binaries in stellar spectra: implications for spectral fitting. *Monthly Notices of the Royal Astronomical Society*, page stx2758.
- Ellis, R.S. (2010). Gravitational lensing: a unique probe of dark matter and dark energy. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1914):967–987.
- Epps, S.D. and Hudson, M.J. (2017). The weak-lensing masses of filaments between luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 468(3):2605.
- Erben, T., Van Waerbeke, L., Bertin, E., Mellier, Y. and Schneider, P. (2001). How accurately can we measure weak gravitational shear? *A&A*, 366(3):717–735.
- Eriksen, M. and Hoekstra, H. (2017). Implications of a wavelength dependent PSF for weak lensing measurements. *ArXiv e-prints*, 1707.04334.
- Eyer, L., Rimoldini, L., Holl, B., North, P., Zucker, S. et al (2015). The Gaia Mission, Binary Stars and Exoplanets. In Rucinski, S.M., Torres, G. and Zejda, M., editors, *Living Together: Planets, Host Stars and Binaries*, volume 496 of *Astronomical Society of the Pacific Conference Series*, page 121.
- Fahlman, G., Kaiser, N., Squires, G. and Woods, D. (1994). Dark matter in MS 1224 from distortion of background galaxies. *ApJ*, 437:56–62.
- Falco, E.E., Kochanek, C.S. and Muñoz, J.A. (1998). Limits on cosmological models from radio-selected gravitational lenses. *The Astrophysical Journal*, 494(1):47.
- Fan, Z., Shan, H. and Liu, J. (2010). Noisy Weak-lensing Convergence Peak Statistics Near Clusters of Galaxies and Beyond. *ApJ*, 719:1408–1420.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.
- Fenech Conti, I., Herbonnet, R., Hoekstra, H., Merten, J., Miller, L. and Viola, M. (2017). Calibration of weak-lensing shear in the kilo-degree survey. *Monthly Notices of the Royal Astronomical Society*, 467(2):1627.
- Fernie, J.D. (1970). The historical quest for the nature of the spiral nebulae. *Publications of the Astronomical Society of the Pacific*, 82(490):1189.
- Ferrara, A. (2017). Einstein and cosmology. *Lettera Matematica*, 5(1):33–37.
- Fischer, P., McKay, T.A., Sheldon, E., Connolly, A., Stebbins, A. et al (2000). Weak lensing with sloan digital sky survey commissioning data: The galaxy-mass correlation function to 1 h-1 mpc. *The Astronomical Journal*, 120(3):1198.
- Freedman, W.L. (2017). Cosmology at at Crossroads: Tension with the Hubble Constant. *Nat. Astron.*, 1:0169.
- Freese, K. (2017). Status of dark matter in the universe. *International Journal of Modern Physics D*, 26(06):1730012.
- Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10:377–386.
- Frieman, J.A., Turner, M.S. and Huterer, D. (2008). Dark Energy and the Accelerating Universe. *ARA&A*, 46:385–432.
- Fu, L. and Kilbinger, M. (2010). A new cosmic shear function: optimized E-/B-mode decomposition on a finite interval. *MNRAS*, 401:1264–1274.

Bibliography

- Fu, L., Kilbinger, M., Erben, T., Heymans, C., Hildebrandt, H. et al (2014). CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *MNRAS*, 441:2725–2743.
- Gaia Collaboration (2016). The gaia mission. *A&A*, 595:A1.
- Gaudi, B.S. (2012). Microlensing surveys for exoplanets. *Annual Review of Astronomy and Astrophysics*, 50(1):411–453.
- Gavazzi, R., Marshall, P.J., Treu, T. and Sonnenfeld, A. (2014). Ringfinder: Automated detection of galaxy-scale gravitational lenses in ground-based multi-filter imaging data. *The Astrophysical Journal*, 785(2):144.
- Gentile, M., Courbin, F. and Meylan, G. (2012). A simple fitting method (gfit) for galaxy shape measurement in weak lensing surveys. *ArXiv e-prints*, 1211.4847.
- Gibbs, S. (2015). Musk, wozniak and hawking urge ban on warfare ai and autonomous weapons. *The Guardian*.
- Gibbs, S. (2017). Elon musk leads 116 experts calling for outright ban on killer robots. *The Guardian*.
- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Teh, Y.W. and Titterton, M., editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy. PMLR.
- Glorot, X., Bordes, A. and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323.
- Goldberg, D.M. and Natarajan, P. (2002). The Galaxy Octopole Moment as a Probe of Weak-Lensing Shear Fields. *ApJ*, 564:65–72.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Gruen, D., Seitz, S., Koppenhoefer, J. and Riffeser, A. (2010). Bias-free shear estimation using artificial neural networks. *The Astrophysical Journal*, 720(1):639.
- Habets, G.M.H.J. and Heintze, J.R.W. (1981). Empirical bolometric corrections for the main-sequence. *A&AS*, 46:193–237.
- Hall, A. and Taylor, A. (2017). Cosmic shear measurement with maximum likelihood and maximum a posteriori inference. *Monthly Notices of the Royal Astronomical Society*, 468(1):346.
- Hall, P.J. (2005). *The Square Kilometre Array: An Engineering Perspective*.
- Hamana, T. and Mellier, Y. (2001). Numerical study of the statistical properties of the lensing excursion angles. *Monthly Notices of the Royal Astronomical Society*, 327(1):169.
- Hansen, B.M.S., Anderson, J., Brewer, J., Dotter, A., Fahlman, G.G. et al (2007). The white dwarf cooling sequence of ngc 6397. *The Astrophysical Journal*, 671(1):380.
- Harnois-Déraps, J. and van Waerbeke, L. (2015). Simulations of weak gravitational lensing - II. Including finite support effects in cosmic shear covariance matrices. *MNRAS*, 450:2857–2873.
- Harrison, I., Camera, S., Zuntz, J. and Brown, M.L. (2016). Ska weak lensing – i. cosmological forecasts and the power of radio-optical cross-correlations. *Monthly Notices of the Royal Astronomical Society*, 463(4):3674.
- Hartlap, J., Hilbert, S., Schneider, P. and Hildebrandt, H. (2011). A bias in cosmic shear from galaxy selection: results from ray-tracing simulations. *A&A*, 528:A51.
- Harvey, D., Massey, R., Kitching, T., Taylor, A. and Tittley, E. (2015). The nongravitational interactions of dark matter in colliding galaxy clusters. *Science*, 347(6229):1462–1465.

- Hawken, A.J. and Bridle, S.L. (2009). Gravitational flexion by elliptical dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 400(3):1132.
- Hazard, C., Mackey, M.B. and Shimmins, A.J. (1963). Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. *Nature*, 197:1037–1039.
- He, H. and Garcia, E.A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.
- He, K., Zhang, X., Ren, S. and Sun, J. (2015). Deep Residual Learning for Image Recognition. *ArXiv e-prints*, 1512.03385.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Identity Mappings in Deep Residual Networks. *ArXiv e-prints*, 1603.05027.
- Herbonnet, R., Buddendiek, A. and Kuijken, K. (2017). Shear nulling after PSF Gaussianisation: Moment-based weak lensing measurements with subpercent noise bias. *A&A*, 599:A73.
- Hettercheidt, M., Simon, P., Schirmer, M., Hildebrandt, H., Schrabbach, T., Erben, T. and Schneider, P. (2007). Gabods: The garching-bonn deep survey - vii. cosmic shear analysis. *A&A*, 468(3):859–876.
- Hewitt, J.N., Turner, E.L., Schneider, D.P., Burke, B.F. and Langston, G.I. (1988). Unusual radio source MG1131+0456 - A possible Einstein ring. *Nature*, 333:537–540.
- Heymans, C., Bell, E.F., Rix, H.W., Barden, M., Borch, A. et al (2006). A weak lensing estimate from gems of the virial to stellar mass ratio in massive galaxies to $z \sim 0.8$. *Monthly Notices of the Royal Astronomical Society: Letters*, 371(1):L60.
- Heymans, C., Van Waerbeke, L., Bacon, D., Berge, J., Bernstein, G. et al (2006). The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations. *MNRAS*, 368:1323–1339.
- High, F.W., Rhodes, J., Massey, R. and Ellis, R. (2007). Pixelation Effects in Weak Lensing. *PASP*, 119:1295–1307.
- Higuchi, Y., Oguri, M. and Hamana, T. (2013). Measuring the mass distribution of voids with stacked weak lensing. *Monthly Notices of the Royal Astronomical Society*, 432(2):1021.
- Higuchi, Y., Oguri, M., Tanaka, M. and Sakurai, J. (2015). Detection of a filament connected to CL0016 with weak gravitational lensing. *ArXiv e-prints*, 1503.06373.
- Hilbert, S., Marian, L., Smith, R.E. and Desjacques, V. (2012). Measuring primordial non-Gaussianity with weak lensing surveys. *MNRAS*, 426:2870–2888.
- Hildebrandt, H., Arnouts, S., Capak, P., Moustakas, L.A., Wolf, C. et al (2010). PHAT: PHoto-z Accuracy Testing. *A&A*, 523:A31.
- Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K. et al (2017). KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS*, 465:1454–1498.
- Hinton, G.E. (1986). Learning distributed representations of concepts. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 1–12.
- Hinton, G.E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14:1771–1800.
- Hinton, G.E. and Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv e-prints*, 1207.0580.
- Hirata, C. and Seljak, U. (2003). Shear calibration biases in weak-lensing surveys. *MNRAS*, 343:459–480.
- Hirata, C.M. and Seljak, U.c.v. (2004). Intrinsic alignment-lensing interference as a contaminant of

Bibliography

- cosmic shear. *Phys. Rev. D*, 70:063526.
- Hobson, M.P., Efstathiou, G.P. and Lasenby, A.N. (2006). *General Relativity*.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8):1735–1780.
- Hoekstra, H. (2004). The effect of imperfect models of point spread function anisotropy on cosmic shear measurements. *Monthly Notices of the Royal Astronomical Society*, 347(4):1337.
- Hoekstra, H., Franx, M. and Kuijken, K. (2000). Hubble Space Telescope Weak-Lensing Study of the $z=0.83$ Cluster MS 1054-03. *ApJ*, 532:88–108.
- Hoekstra, H., Franx, M., Kuijken, K. and Squires, G. (1998). Weak Lensing Analysis of CL 1358+62 Using Hubble Space Telescope Observations. *ApJ*, 504:636–660.
- Hoekstra, H., Herbonnet, R., Muzzin, A., Babul, A., Mahdavi, A., Viola, M. and Cacciato, M. (2015). The canadian cluster comparison project: detailed study of systematics and updated weak lensing masses. *Monthly Notices of the Royal Astronomical Society*, 449(1):685.
- Hoekstra, H. and Jain, B. (2008). Weak Gravitational Lensing and Its Cosmological Applications. *Annual Review of Nuclear and Particle Science*, 58:99–123.
- Hoekstra, H., Viola, M. and Herbonnet, R. (2017). A study of the sensitivity of shape measurements to the input parameters of weak-lensing image simulations. *Monthly Notices of the Royal Astronomical Society*, 468(3):3295–3311.
- Hoekstra, H., Yee, H.K.C. and Gladders, M.D. (2004). Properties of galaxy dark matter halos from weak lensing. *The Astrophysical Journal*, 606(1):67.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257.
- Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Hu, W. (1999). Power spectrum tomography with weak lensing. *The Astrophysical Journal Letters*, 522(1):L21.
- Hu, W., Eisenstein, D.J., Tegmark, M. and White, M. (1998). Observationally determining the properties of dark matter. *Phys. Rev. D*, 59:023512.
- Hubble, E. (1929a). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.
- Hubble, E. (1934). The Distribution of Extra-Galactic Nebulae. *ApJ*, 79:8.
- Hubble, E.P. (1929b). A Clue to the Structure of the Universe. *Leaflet of the Astronomical Society of the Pacific*, 1:93.
- Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E. and Perley, R. (1985). 2237 + 0305: A new and unusual gravitational lens. *AJ*, 90:691–696.
- Huff, E. and Mandelbaum, R. (2017). Metacalibration: Direct Self-Calibration of Biases in Shear Measurement. *ArXiv e-prints*, 1702.02600.
- Huterer, D. (2002). Weak lensing and dark energy. *Phys. Rev. D*, 65:063001.
- Huterer, D. (2010). Weak lensing, dark matter and dark energy. *General Relativity and Gravitation*, 42(9):2177–2195.
- Huterer, D., Kirkby, D., Bean, R., Connolly, A., Dawson, K. et al (2015). Growth of cosmic structure: Probing dark energy beyond expansion. *Astroparticle Physics*, 63:23 – 41. Dark Energy and CMB.
- Huterer, D., Takada, M., Bernstein, G. and Jain, B. (2006). Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration. *Monthly Notices of the Royal Astronomical Society*, 366(1):101.

- Ingrosso, G., Novati, S.C., De Paolis, F., Jetzer, P., Nucita, A.A. and Zakharov, A.F. (2009). Pixel lensing as a way to detect extrasolar planets in m31. *Monthly Notices of the Royal Astronomical Society*, 399(1):219.
- Irwin, J. and Shmakova, M. (2006). Observation of small-scale structure using sextupole lensing. *The Astrophysical Journal*, 645(1):17.
- Ivezić, Z. (2014). *Statistics, data mining, and machine learning in astronomy : a practical Python guide for the analysis of survey data*. Princeton series in modern observational astronomy. Princeton University Press, Princeton, N.J.
- Jacobs, C., Glazebrook, K., Collett, T., More, A. and McCarthy, C. (2017). Finding strong lenses in cftls using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 471(1):167–181.
- Jain, B. and Seljak, U. (1997). Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes. *ApJ*, 484:560–573.
- Jain, B., Seljak, U. and White, S. (2000). Ray-tracing simulations of weak lensing by large-scale structure. *The Astrophysical Journal*, 530(2):547.
- Jain, B., Seljak, U. and White, S. (2000). Ray-tracing Simulations of Weak Lensing by Large-Scale Structure. *ApJ*, 530:547–577.
- Jain, B. and Taylor, A. (2003). Cross-correlation tomography: Measuring dark energy evolution with weak lensing. *Phys. Rev. Lett.*, 91:141302.
- Jarvis, M., Bernstein, G. and Jain, B. (2004). The skewness of the aperture mass statistic. *MNRAS*, 352:338–352.
- Jarvis, M., Schechter, P. and Jain, B. (2008). Telescope Optics and Weak Lensing: PSF Patterns due to Low Order Aberrations. *ArXiv e-prints*, 0810.0027.
- Jee, M.J., Tyson, J.A., Hilbert, S., Schneider, M.D., Schmidt, S. and Wittman, D. (2016). Cosmic shear results from the deep lens survey. ii. full cosmological parameter constraints from tomography. *The Astrophysical Journal*, 824(2):77.
- Joachimi, B., Cacciato, M., Kitching, T.D., Leonard, A., Mandelbaum, R. et al (2015). Galaxy alignments: An overview. *Space Science Reviews*, 193(1):1–65.
- Joseph, R., Courbin, F., Metcalf, R. B., Giocoli, C., Hartley, P. et al (2014). A pca-based automated finder for galaxy-scale strong lenses. *A&A*, 566:A63.
- Joyce, A., Lombriser, L. and Schmidt, F. (2016). Dark Energy Versus Modified Gravity. *Annual Review of Nuclear and Particle Science*, 66:95–122.
- Kacprzak, T., Bridle, S., Rowe, B., Voigt, L., Zuntz, J., Hirsch, M. and MacCrann, N. (2014). Sérsic galaxy models in weak lensing shape measurement: model bias, noise bias and their interaction. *MNRAS*, 441:2528–2538.
- Kacprzak, T., Zuntz, J., Rowe, B., Bridle, S., Refregier, A., Amara, A., Voigt, L. and Hirsch, M. (2012). Measurement and calibration of noise bias in weak lensing galaxy shape estimation. *Monthly Notices of the Royal Astronomical Society*, 427(4):2711.
- Kahlhoefer, F., Schmidt-Hoberg, K., Frandsen, M.T. and Sarkar, S. (2014). Colliding clusters and dark matter self-interactions. *Monthly Notices of the Royal Astronomical Society*, 437(3):2865.
- Kaiser, N. (1986). Evolution and clustering of rich clusters. *Monthly Notices of the Royal Astronomical Society*, 222(2):323.
- Kaiser, N. (1992). Weak gravitational lensing of distant galaxies. *ApJ*, 388:272–286.
- Kaiser, N. (1998). Weak lensing and cosmology. *The Astrophysical Journal*, 498(1):26.

Bibliography

- Kaiser, N. (2000). A New Shear Estimator for Weak-Lensing Observations. *ApJ*, 537:555–577.
- Kaiser, N. and Squires, G. (1993). Mapping the dark matter with weak gravitational lensing. *ApJ*, 404:441–450.
- Kaiser, N., Squires, G. and Broadhurst, T. (1995). A Method for Weak Lensing Observations. *ApJ*, 449:460.
- Kaiser, N., Wilson, G. and Luppino, G.A. (2000). Large-Scale Cosmic Shear Measurements. *ArXiv Astrophysics e-prints*, astro-ph/0003338.
- Karhunen, J., Raiko, T. and Cho, K. (2015). Unsupervised deep learning: A short review. *Advances in Independent Component Analysis and Learning Machines*, 125.
- Kelly, P.L., Rodney, S.A., Treu, T., Foley, R.J., Brammer, G. et al (2015). Multiple images of a highly magnified supernova formed by an early-type cluster galaxy lens. *Science*, 347(6226):1123–1126.
- Kettula, K., Giodini, S., van Uitert, E., Hoekstra, H., Finoguenov, A. et al (2015). Cfhtlens: weak lensing calibrated scaling relations for low-mass clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 451(2):1460.
- Kiessling, A., Cacciato, M., Joachimi, B., Kirk, D., Kitching, T.D. et al (2015). Galaxy alignments: Theory, modelling & simulations. *Space Science Reviews*, 193(1):67–136.
- Kilbinger, M. (2015). Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901.
- Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J. et al (2013). Cfhtlens: combined probe cosmological model comparison using 2d weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 430(3):2200.
- Killedar, M., Lasky, P.D., Lewis, G.F. and Fluke, C.J. (2012). Gravitational lensing with three-dimensional ray tracing. *MNRAS*, 420:155–169.
- Kingma, D.P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, 1412.6980.
- Kirk, D., Brown, M.L., Hoekstra, H., Joachimi, B., Kitching, T.D. et al (2015). Galaxy alignments: Observations and impact on cosmology. *Space Science Reviews*, 193(1):139–211.
- Kirk, D., Rassat, A., Host, O. and Bridle, S. (2012). The cosmological impact of intrinsic alignment model choice for cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 424(3):1647.
- Kitching, T., Amara, A., Gill, M., Harmeling, S., Heymans, C. et al (2011). Gravitational lensing accuracy testing 2010 (great10) challenge handbook. *Ann. Appl. Stat.*, 5(3):2231–2263.
- Kitching, T.D., Balan, S.T., Bridle, S., Cantale, N., Courbin, F. et al (2012). Image analysis for cosmology: results from the GREAT10 Galaxy Challenge. *MNRAS*, 423:3163–3208.
- Kitching, T.D., Balan, S.T., Bridle, S., Cantale, N., Courbin, F. et al (2012). Image analysis for cosmology: results from the great10 galaxy challenge. *Monthly Notices of the Royal Astronomical Society*, 423(4):3163.
- Kitching, T.D., Heavens, A.F., Alsing, J., Erben, T., Heymans, C. et al (2014). 3d cosmic shear: cosmology from cfhtlens. *Monthly Notices of the Royal Astronomical Society*, 442(2):1326.
- Kitching, T.D., Miller, L., Heymans, C.E., Van Waerbeke, L. and Heavens, A.F. (2008). Bayesian galaxy shape measurement for weak lensing surveys – ii. application to simulations. *Monthly Notices of the Royal Astronomical Society*, 390(1):149.
- Kitching, T.D., Rowe, B., Gill, M., Heymans, C., Massey, R. et al (2013). Image analysis for cosmology: Results from the great10 star challenge. *The Astrophysical Journal Supplement Series*, 205(2):12.
- Kleinbaum, D. and Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. Springer-Verlag New York.
- Klimov, Y.G. (1963). The Deflection of Light Rays in the Gravitational Fields of Galaxies. *Soviet Physics Doklady*, 8:119.

- Kneib, J.P., Ellis, R.S., Smail, I., Couch, W.J. and Sharples, R.M. (1996). Hubble Space Telescope Observations of the Lensing Cluster Abell 2218. *ApJ*, 471:643.
- Koo, D.C. (1985). Optical multicolors - A poor person's Z machine for galaxies. *AJ*, 90:418–440.
- Kratochvil, J.M., Haiman, Z. and May, M. (2010). Probing cosmology with weak lensing peak counts. *Phys. Rev. D*, 81:043519.
- Krause, E., Chang, T.C., Doré, O. and Umetsu, K. (2013). The Weight of Emptiness: The Gravitational Lensing Signal of Stacked Voids. *ApJ*, 762:L20.
- Krist, J.E., Hook, R.N. and Stoehr, F. (2011). 20 years of Hubble Space Telescope optical modeling using Tiny Tim. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8127 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0.
- Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.
- Kuijken, K. (1999). Weak weak lensing: correcting weak shear measurements accurately for PSF anisotropy. *A&A*, 352:355–362.
- Kuijken, K. (2006). Shears from shapelets. *A&A*, 456:827–838.
- Kuijken, K., Heymans, C., Hildebrandt, H., Nakajima, R., Erben, T. et al (2015). Gravitational lensing analysis of the kilo-degree survey. *Monthly Notices of the Royal Astronomical Society*, 454(4):3500.
- Kuntzer, T. and Courbin, F. (2017). Detecting unresolved binary stars in euclid vis images. *A&A*, 606:A119.
- Kuntzer, T., Courbin, F. and Meylan, G. (2016a). Evaluating the effect of stellar multiplicity on the point spread function of space-based weak lensing surveys. *A&A*, 586:A74.
- Kuntzer, T., Tewes, M. and Courbin, F. (2016b). Stellar classification from single-band imaging using machine learning. *A&A*, 591:A54.
- Lambas, D.G., Maddox, S.J. and Loveday, J. (1992). On the true shapes of galaxies. *MNRAS*, 258:404–414.
- Lanusse, F., Ma, Q., Li, N., Collett, T.E., Li, C.L., Ravanbakhsh, S., Mandelbaum, R. and Poczos, B. (2017). CMU DeepLens: Deep Learning For Automatic Image-based Galaxy-Galaxy Strong Lens Finding. *Monthly Notices of the Royal Astronomical Society*.
- Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P. and Bottou, L. (2008). Exploring strategies for training deep neural networks. *Journal of Machine Learning Research*, page 2009.
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J., Brinchmann, J. et al (2011). Euclid Study Report. *Euclid Study Report*.
- Leauthaud, A., Finoguenov, A., Kneib, J.P., Taylor, J.E., Massey, R. et al (2010). A weak lensing study of x-ray groups in the cosmos survey: Form and evolution of the mass-luminosity relation. *The Astrophysical Journal*, 709(1):97.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lee, H., Ekanadham, C. and Ng, A.Y. (2008). Sparse deep belief net model for visual area v2. In *Advances in neural information processing systems*, pages 873–880.
- Lee, K.H., Kang, S.J., Kang, W.H. and Kim, N.S. (2016). Two-stage noise aware training using asymmetric deep denoising autoencoder. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5765–5769.
- Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*,

Bibliography

- 47:49–59.
- Leonard, A., Lanusse, F. and Starck, J.L. (2014). Glimpse: accurate 3d weak lensing reconstructions using sparsity. *Monthly Notices of the Royal Astronomical Society*, 440(2):1281.
- Lewis, A. (2009). Galaxy shear estimation from stacked images. *MNRAS*, 398:471–476.
- Lewis, A. and Challinor, A. (2006). Weak gravitational lensing of the CMB. *Phys. Rep.*, 429:1–65.
- Li, R., Mo, H.J., Fan, Z., Cacciato, M., Van Den Bosch, F.C., Yang, X. and More, S. (2009). Modelling galaxy–galaxy weak lensing with sloan digital sky survey groups. *Monthly Notices of the Royal Astronomical Society*, 394(2):1016.
- Libeskind, N.I., van de Weygaert, R., Cautun, M., Falck, B., Tempel, E. et al (2018). Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217.
- Liebes, S. (1964). Gravitational Lenses. *Physical Review*, 133:835–844.
- Lin, C.A. and Kilbinger, M. (2015). A new model to predict weak-lensing peak counts. I. Comparison with N-body simulations. *A&A*, 576:A24.
- Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S. et al (2008). Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey*. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189.
- Livermore, R.C., Jones, T.A., Richard, J., Bower, R.G., Swinbank, A.M. et al (2015). Resolved spectroscopy of gravitationally lensed galaxies: global dynamics and star-forming clumps on ~ 100 pc scales at $1 < z < 4$. *Monthly Notices of the Royal Astronomical Society*, 450(2):1812.
- López-Corredoira, M. (2017). Tests and problems of the standard model in cosmology. *Foundations of Physics*, 47(6):711–768.
- LoVerde, M. and Afshordi, N. (2008). Extended limber approximation. *Phys. Rev. D*, 78:123506.
- LSST Science Collaboration, Abell, P.A., Allison, J., Anderson, S.F., Andrew, J.R. et al (2009). LSST Science Book, Version 2.0. *ArXiv e-prints*, 0912.0201.
- Lu, T., Zhang, J., Dong, F., Li, Y., Liu, D., Fu, L., Li, G. and Fan, Z. (2017). Testing PSF Interpolation in Weak Lensing with Real Data. *AJ*, 153:197.
- Lundmark, K. (1930). Über die Bestimmung der Entfernungen, Dimensionen, Massen und Dichtigkeit für die nächstgelegenen anagalactischen Sternsysteme. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 125:1–13.
- Luppino, G.A. and Kaiser, N. (1997). Detection of Weak Lensing by a Cluster of Galaxies at $z = 0.83$. *ApJ*, 475:20–28.
- Lynds, R. and Petrosian, V. (1986). Giant Luminous Arcs in Galaxy Clusters. In *Bulletin of the American Astronomical Society*, volume 18 of BAAS, page 1014.
- Lyon, R.J., Stappers, B.W., Cooper, S., Brooke, J.M. and Knowles, J.D. (2016). Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, 459(1):1104–1123.
- Längkvist, M., Karlsson, L. and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11 – 24.
- Maeder, A. (2009). *Physics, Formation and Evolution of Rotating Stars*. Springer-Verlag Berlin Heidelberg.
- Majumdar, A. and Tripathi, A. (2017). Asymmetric stacked autoencoder. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 911–918.
- Makhzani, A. and Frey, B. (2013). k-Sparse Autoencoders. *ArXiv e-prints*, 1312.5663.
- Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal*

- Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2065).
- Mandelbaum, R. (2017). Weak lensing for precision cosmology. *ArXiv e-prints*, 1710.03235.
- Mandelbaum, R., Hirata, C.M., Leauthaud, A., Massey, R.J. and Rhodes, J. (2012). Precision simulation of ground-based lensing data using observations from space. *Monthly Notices of the Royal Astronomical Society*, 420(2):1518.
- Mandelbaum, R., Hirata, C.M., Seljak, U., Guzik, J., Padmanabhan, N. et al (2005). Systematic errors in weak lensing: application to sdss galaxy-galaxy weak lensing. *Monthly Notices of the Royal Astronomical Society*, 361(4):1287.
- Mandelbaum, R., Lanusse, F., Leauthaud, A., Armstrong, R., Simet, M. et al (2017). Weak lensing shear calibration with simulations of the HSC survey. *ArXiv e-prints*, 1710.00885.
- Mandelbaum, R., Rowe, B., Armstrong, R., Bard, D., Bertin, E. et al (2015). Great3 results – i. systematic errors in shear estimation and the impact of real galaxy morphology. *Monthly Notices of the Royal Astronomical Society*, 450(3):2963.
- Mandelbaum, R., Rowe, B., Bosch, J., Chang, C., Courbin, F. et al (2014). The third gravitational lensing accuracy testing (great3) challenge handbook. *The Astrophysical Journal Supplement Series*, 212(1):5.
- Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C.M. and Brinkmann, J. (2006). Galaxy halo masses and satellite fractions from galaxy–galaxy lensing in the sloan digital sky survey: stellar mass, luminosity, morphology and environment dependencies. *Monthly Notices of the Royal Astronomical Society*, 368(2):715.
- Mandelbaum, R., Slosar, A., Baldauf, T., Seljak, U., Hirata, C.M., Nakajima, R., Reyes, R. and Smith, R.E. (2013). Cosmological parameter constraints from galaxy–galaxy lensing and galaxy clustering with the sdss dr7. *Monthly Notices of the Royal Astronomical Society*, 432(2):1544.
- Mantz, A.B., Allen, S.W., Morris, R.G., von der Linden, A., Applegate, D.E. et al (2016). Weighing the giants – v. galaxy cluster scaling relations. *Monthly Notices of the Royal Astronomical Society*, 463(4):3582.
- Mao, S. and Paczynski, B. (1991). Gravitational microlensing by double stars and planetary systems. *ApJ*, 374:L37–L40.
- Maoli, R., Van Waerbeke, L., Mellier, Y., Schneider, P., Jain, B., Bernardeau, F., Erben, T. and Fort, B. (2001). Cosmic shear analysis in 50 uncorrelated vlt fields. implications for ω_0 , σ_8 . *A&A*, 368(3):766–775.
- Maoz, D. and Mannucci, F. (2012). Type-Ia Supernova Rates and the Progenitor Problem: A Review. *PASA*, 29:447–465.
- Marcy, G.W. and Benitz, K.J. (1989). A search for substellar companions to low-mass stars. *ApJ*, 344:441–453.
- Marshall, P.J., Verma, A., More, A., Davis, C.P., More, S. et al (2016). Space warps – i. crowdsourcing the discovery of gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 455(2):1171–1190.
- Masci, J., Meier, U., Cireşan, D. and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 52–59.
- Massey, R., Heymans, C., Bergé, J., Bernstein, G., Bridle, S. et al (2007a). The Shear Testing Programme 2: Factors affecting high-precision weak-lensing analyses. *MNRAS*, 376:13–38.
- Massey, R., Hoekstra, H., Kitching, T., Rhodes, J., Cropper, M. et al (2013). Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation). *Monthly Notices of the Royal Astronomical Society*, 429(1):661.
- Massey, R., Kitching, T. and Richard, J. (2010). The dark matter of gravitational lensing. *Reports on Progress in Physics*, 73(8):086901.

Bibliography

- Massey, R. and Refregier, A. (2005). Polar shapelets. *MNRAS*, 363:197–210.
- Massey, R., Rhodes, J., Ellis, R., Scoville, N., Leauthaud, A. et al (2007b). Dark matter maps reveal cosmic scaffolding. *Nature*, 445:286–290.
- Massey, R., Rhodes, J., Leauthaud, A., Capak, P., Ellis, R. et al (2007c). COSMOS: Three-dimensional Weak Lensing and the Growth of Structure. *ApJS*, 172:239–253.
- McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.
- Mediavilla, E., Jiménez-Vicente, J., Muñoz, J.A., Vives-Arias, H. and Calderón-Infante, J. (2017). Limits on the mass and abundance of primordial black holes from quasar gravitational microlensing. *The Astrophysical Journal Letters*, 836(2):L18.
- Melchior, P., Böhnert, A., Lombardi, M. and Bartelmann, M. (2010). Limitations on shapelet-based weak-lensing measurements. *A&A*, 510:A75.
- Melchior, P., Sutter, P.M., Sheldon, E.S., Krause, E. and Wandelt, B.D. (2014). First measurement of gravitational lensing by cosmic voids in SDSS. *MNRAS*, 440:2922–2927.
- Melchior, P. and Viola, M. (2012). Means of confusion: how pixel noise affects shear estimates for weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 424(4):2757.
- Melchior, P., Viola, M., Schäfer, B.M. and Bartelmann, M. (2011). Weak gravitational lensing with deimos. *Monthly Notices of the Royal Astronomical Society*, 412(3):1552.
- Melia, F., Wei, J.J. and Wu, X.F. (2015). A comparison of cosmological models using strong gravitational lensing galaxies. *The Astronomical Journal*, 149(1):2.
- Messick, C., Blackburn, K., Brady, P., Brockill, P., Cannon, K. et al (2017). Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95:042001.
- Metcalf, R.B. and Amara, A. (2012). Small-scale structures of dark matter and flux anomalies in quasar gravitational lenses. *MNRAS*, 419:3414–3425.
- Michell, J. (1784). On the means of discovering the distance, magnitude, &c. of the fixed stars, in consequence of the diminution of the velocity of their light, in case such a diminution should be found to take place in any of them, and such other data should be procured from observations, as would be farther necessary for that purpose. by the rev. john michell, b. d. f. r. s. in a letter to henry cavendish, esq. f. r. s. and a. s. *Philosophical Transactions of the Royal Society of London*, 74:35–57.
- Miller, L., Heymans, C., Kitching, T.D., van Waerbeke, L., Erben, T. et al (2013). Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. *MNRAS*, 429:2858–2880.
- Miller, L., Kitching, T.D., Heymans, C., Heavens, A.F. and van Waerbeke, L. (2007). Bayesian galaxy shape measurement for weak lensing surveys - I. Methodology and a fast-fitting algorithm. *MNRAS*, 382:315–324.
- Milone, A.P., Piotto, G., Bedin, L.R., King, I.R., Anderson, J. et al (2012). Multiple stellar populations in 47 Tucanae. *The Astrophysical Journal*, 744(1):58.
- Minsky, M.L. and Papert, S.A. (1988). *Perceptrons: Expanded Edition*. MIT Press, Cambridge, MA, USA.
- Miralda-Escude, J. (1991). The correlation function of galaxy ellipticities produced by gravitational lensing. *ApJ*, 380:1–8.
- Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J. and Tozzi, P. (1999). Dark matter substructure within galactic halos. *The Astrophysical Journal Letters*, 524(1):L19.
- More, S., Miyatake, H., Mandelbaum, R., Takada, M., Spergel, D.N., Brownstein, J.R. and Schneider, D.P. (2015). The weak lensing signal and the clustering of boss galaxies. ii. astrophysical and cosmological constraints. *The Astrophysical Journal*, 806(1):2.

- Motta, V., Mediavilla, E., Rojas, K., Falco, E.E., Jiménez-Vicente, J. and Muñoz, J.A. (2017). Probing the broad-line region and the accretion disk in the lensed quasars he 0435-1223, wfi 2033-4723, and he 2149-2745 using gravitational microlensing. *The Astrophysical Journal*, 835(2):132.
- Munshi, D., Valageas, P., van Waerbeke, L. and Heavens, A. (2008). Cosmology with weak lensing surveys. *Physics Reports*, 462(3):67 – 121.
- Möller, O., Kitzbichler, M. and Natarajan, P. (2007). Strong lensing statistics in large, $z \lesssim 0.2$, surveys: bias in the lens galaxy population. *Monthly Notices of the Royal Astronomical Society*, 379(3):1195–1208.
- Nair, V. and Hinton, G.E. (2009). 3d object recognition with deep belief nets. In *Advances in neural information processing systems*, pages 1339–1347.
- Nakajima, R. and Bernstein, G. (2007). Shear recovery accuracy in weak-lensing analysis with the elliptical gauss-laguerre method. *The Astronomical Journal*, 133(4):1763.
- Napolitano, N.R., Covone, G., Roy, N., Tortora, C., La Barbera, F. et al (2016). *Strong Lens Search in the ESO Public Survey KiDS*, pages 129–133. Springer International Publishing, Cham.
- Nardiello, D., Milone, A. P., Piotto, G., Marino, A. F., Bellini, A. and Cassisi, S. (2015). Observing multiple stellar populations with vlt/fors2. *A&A*, 573:A70.
- Natarajan, P., Kneib, J.P. and Smail, I. (2002). Evidence for tidal stripping of dark matter halos in massive cluster lenses. *The Astrophysical Journal Letters*, 580(1):L11.
- Navarro, J.F., Frenk, C.S. and White, S.D.M. (1997). A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 490(2):493.
- Negrello, M., Hopwood, R., De Zotti, G., Cooray, A., Verma, A. et al (2010). The Detection of a Population of Submillimeter-Bright, Strongly Lensed Galaxies. *Science*, 330:800.
- Ngan, W., Van Waerbeke, L., Mahdavi, A., Heymans, C. and Hoekstra, H. (2009). Sérsiclets – a matched filter extension of shapelets for weak lensing studies. *Monthly Notices of the Royal Astronomical Society*, 396(2):1211.
- Nielsen, M. (2015). *Neural Networks and Deep Learning*. Determination Press. <http://neuralnetworksanddeeplearning.com>.
- Niemi, S.M., Kitching, T.D. and Cropper, M. (2015). On weak lensing shape noise. *Monthly Notices of the Royal Astronomical Society*, 454(2):1221.
- Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2 edition.
- Okamoto, T. and Hu, W. (2003). Cosmic microwave background lensing reconstruction on the full sky. *Phys. Rev. D*, 67:083002.
- Okura, Y. and Futamase, T. (2009). A new method for measuring weak gravitational lensing shear using higher order spin-2 holics. *The Astrophysical Journal*, 699(1):143.
- Okura, Y. and Futamase, T. (2012). Elliptical-weighted holics for weak lensing shear measurement. ii. point-spread function correction and application to a370. *The Astrophysical Journal*, 748(2):112.
- Paczynski, B. (1986a). Gravitational microlensing at large optical depth. *ApJ*, 301:503–516.
- Paczynski, B. (1986b). Gravitational microlensing by the galactic halo. *ApJ*, 304:1–5.
- Paczynski, B. (1987). Giant luminous arcs discovered in two clusters of galaxies. *Nature*, 325:572–573.
- Paraficz, D., Courbin, F., Tramacere, A., Joseph, R., Metcalf, R. B. et al (2016). The pca lens-finder: application to cfhtls. *A&A*, 592:A75.
- Parker, L.C., Hoekstra, H., Hudson, M.J., van Waerbeke, L. and Mellier, Y. (2007). The masses and shapes of dark matter halos from galaxy-galaxy lensing in the cfht legacy survey. *The Astrophysical Journal*, 669(1):21.

Bibliography

- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A. and Bridle, S.L. (2008). Point spread function calibration requirements for dark energy from cosmic shear. *A&A*, 484:67–77.
- Peacock, J.A. (2003). Large-scale structure and matter in the universe. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 361(1812):2479–2495.
- Peacock, J.A., Schneider, P., Efstathiou, G., Ellis, J.R., Leibundgut, B., Lilly, S.J. and Mellier, Y. (2006). ESA-ESO Working Group on “Fundamental Cosmology”. Technical report.
- Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. et al (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Peebles, P.J.E. and Yu, J.T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815.
- Peel, A., Lin, C.A., Lanusse, F., Leonard, A., Starck, J.L. and Kilbinger, M. (2017). Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *A&A*, 599:A79.
- Pen, U.L., Lu, T., van Waerbeke, L. and Mellier, Y. (2003). The three-dimensional power spectrum of dark and luminous matter from the VIRMOS-DESCART cosmic shear survey. *MNRAS*, 346:994–1008.
- Penzias, A.A. and Wilson, R.W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421.
- Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R.A., Nugent, P. et al (1999). Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565.
- Petrillo, C.E., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L.V.E. et al (2017). Finding strong gravitational lenses in the kilo degree survey with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 472(1):1129–1150.
- Pickles, A.J. (1998). A Stellar Spectral Flux Library: 1150-25000 Å. *PASP*, 110:863–878.
- Pires, S. and Amara, A. (2010). Weak lensing mass reconstruction: Flexion versus shear. *The Astrophysical Journal*, 723(2):1507.
- Pires, S., Starck, J.L., Amara, A., Teyssier, R., Réfrégier, A. and Fadili, J. (2009a). FASt STATistics for weak Lensing (FASTLens): fast method for weak lensing statistics and map making. *MNRAS*, 395:1265–1279.
- Pires, S., Starck, J.L. and Refregier, A. (2010). Light on dark matter with weak gravitational lensing. *IEEE Signal Processing Magazine*, 27(1):76–85.
- Pires, S., Starck, J.-L., Amara, A., Réfrégier, A. and Teyssier, R. (2009b). Cosmological model discrimination with weak lensing. *A&A*, 505(3):969–979.
- Planck Collaboration (2016a). Planck 2015 results - xiii. cosmological parameters. *A&A*, 594:A13.
- Planck Collaboration (2016b). Planck 2015 results - xv. gravitational lensing. *A&A*, 594:A15.
- Potter, D., Stadel, J. and Teyssier, R. (2017). Pkdgrav3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1):2.
- Raghavan, D., McAlister, H.A., Henry, T.J., Latham, D.W., Marcy, G.W. et al (2010). A survey of stellar families: Multiplicity of solar-type stars. *The Astrophysical Journal Supplement Series*, 190(1):1.
- Ranzato, M., Huang, F., Boureau, Y. and LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Proc. Computer Vision and Pattern Recognition Conference (CVPR’07)*. IEEE Press.
- Refregier, A. (2003). Shapelets – i. a method for image analysis. *Monthly Notices of the Royal Astronomical Society*, 338(1):35.

- Refregier, A. and Bacon, D. (2003). Shapelets – ii. a method for weak lensing measurements. *Monthly Notices of the Royal Astronomical Society*, 338(1):48.
- Refregier, A., Kacprzak, T., Amara, A., Bridle, S. and Rowe, B. (2012). Noise bias in weak lensing shape measurements. *Monthly Notices of the Royal Astronomical Society*, 425(3):1951.
- Refsdal, S. (1964). On the possibility of determining hubble’s parameter and the masses of galaxies from the gravitational lens effect. *Monthly Notices of the Royal Astronomical Society*, 128(4):307.
- Reipurth, B., Clarke, C.J., Boss, A.P., Goodwin, S.P., Rodríguez, L.F., Stassun, K.G., Tokovinin, A. and Zinnecker, H. (2014). Multiplicity in Early Stellar Evolution. *Protostars and Planets VI*, pages 267–290.
- Renn, J., Sauer, T. and Stachel, J. (1997). The origin of gravitational lensing: A postscript to einstein’s 1936 science paper. *Science*, 275(5297):184–186.
- Rexroth, M. (2015). Substructure in the frontier fields from weak lensing flexion. *Proceedings of the International Astronomical Union*, 11(A29B):795–796.
- Rexroth, M., Natarajan, P. and Kneib, J.P. (2016). A new method to break the mass-sheet degeneracy using aperture moments. *Monthly Notices of the Royal Astronomical Society*, 460(3):2505.
- Rhodes, J., Nichol, R.C., Aubourg, É., Bean, R., Boutigny, D. et al (2017). Scientific Synergy Between LSST and Euclid. *ArXiv e-prints*, 1710.08489.
- Richer, H.B., Dotter, A., Hurley, J., Anderson, J., King, I. et al (2008). Deep advanced camera for surveys imaging in the globular cluster ngc 6397: The cluster color-magnitude diagram and luminosity function. *The Astronomical Journal*, 135(6):2141.
- Riddle, R.L., Tokovinin, A., Mason, B.D., Hartkopf, W.I., Jr., L.C.R. et al (2015). A survey of the high order multiplicity of nearby solar-type binary stars with robo-ao. *The Astrophysical Journal*, 799(1):4.
- Riess, A.G., Filippenko, A.V., Challis, P., Clocchiatti, A., Diercks, A. et al (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038.
- Riess, A.G., Macri, L.M., Hoffmann, S.L., Scolnic, D., Casertano, S. et al (2016). A 2.4% Determination of the Local Value of the Hubble Constant. *ApJ*, 826:56.
- Rifai, S., Vincent, P., Muller, X., Glorot, X. and Bengio, Y. (2011). Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML’11, pages 833–840, USA. Omnipress.
- Rix, H.W., Barden, M., Beckwith, S.V.W., Bell, E.F., Borch, A. et al (2004). Gems: Galaxy evolution from morphologies and sed. *The Astrophysical Journal Supplement Series*, 152(2):163.
- Roberts, M.S. and Rots, A.H. (1973). Comparison of Rotation Curves of Different Galaxy Types. *A&A*, 26:483–485.
- Robin, A.C., Reylé, C., Derrière, S. and Picaud, S. (2003). A synthetic view on structure and evolution of the Milky Way. *A&A*, 409:523–540.
- Rodney, S.A., Patel, B., Scolnic, D., Foley, R.J., Molino, A. et al (2015). Illuminating a dark lens : A type ia supernova magnified by the frontier fields galaxy cluster abell 2744. *The Astrophysical Journal*, 811(1):70.
- Rodríguez, D.R., Duchêne, G., Tom, H., Kennedy, G.M., Matthews, B., Greaves, J. and Butner, H. (2015). Stellar multiplicity and debris discs: an unbiased sample. *Monthly Notices of the Royal Astronomical Society*, 449(3):3160–3170.
- Rodríguez, S., Padilla, N.D. and García Lambas, D. (2016). The effects of environment on the intrinsic shape of galaxies. *Monthly Notices of the Royal Astronomical Society*, 456(1):571.
- Rodríguez, S., Guez, S. and Padilla, N.D. (2013). The intrinsic shape of galaxies in sdss/galaxy zoo. *Monthly Notices of the Royal Astronomical Society*, 434(3):2153.

Bibliography

- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- Rosenblatt, F. (1962). Principles of neurodynamics.
- Rowe, B. (2010). Improving psf modelling for weak gravitational lensing using new methods in model selection. *Monthly Notices of the Royal Astronomical Society*, 404(1):350.
- Rowe, B., Bacon, D., Massey, R., Heymans, C., Häußler, B., Taylor, A., Rhodes, J. and Mellier, Y. (2013). Flexion measurement in simulations of hubble space telescope data. *Monthly Notices of the Royal Astronomical Society*, 435(1):822.
- Rowe, B., Jarvis, M., Mandelbaum, R., Bernstein, G., Bosch, J. et al (2015). Galsim: The modular galaxy image simulation toolkit. *Astronomy and Computing*, 10(0):121 – 150.
- Rubin, V.C. and Ford, Jr., W.K. (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.
- Sadeh, I., Abdalla, F.B. and Lahav, O. (2016). ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning. *PASP*, 128(10):104502.
- Sample, I. (2017). Computer says no: why making ais fair, accountable and transparent is crucial. *The Guardian*.
- Sarajedini, A., Bedin, L.R., Chaboyer, B., Dotter, A., Siegel, M. et al (2007). The acs survey of galactic globular clusters. i. overview and clusters without previous hubble space telescope photometry. *The Astronomical Journal*, 133(4):1658.
- Sato, M., Hamana, T., Takahashi, R., Takada, M., Yoshida, N., Matsubara, T. and Sugiyama, N. (2009). Simulations of Wide-Field Weak Lensing Surveys. I. Basic Statistics and Non-Gaussian Effects. *ApJ*, 701:945–954.
- Schaefer, C., Geiger, M., Kuntzer, T. and Kneib, J. (2017). Deep Convolutional Neural Networks as strong gravitational lens detectors. *ArXiv e-prints*, 1705.07132.
- Schechter, P.L. and Wambsganss, J. (2002). Quasar microlensing at high magnification and the role of dark matter: Enhanced fluctuations and suppressed saddle points. *The Astrophysical Journal*, 580(2):685.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].
- Schneider, P. (1996). Detection of (dark) matter concentrations via weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 283(3):837.
- Schneider, P. (2006). *Extragalactic Astronomy and Cosmology: An Introduction*. Springer.
- Schneider, P., Ehlers, J. and Falco, E. (1992). *Gravitational Lenses*. Springer Berlin Heidelberg.
- Schneider, P., Eifler, T. and Krause, E. (2010). COSEBIs: Extracting the full E-/B-mode information from cosmic shear correlation functions. *A&A*, 520:A116.
- Schneider, P. and Er, X. (2008). Weak lensing goes bananas: what flexion really measures. *A&A*, 485(2):363–376.
- Schneider, P., Kochanek, S. and Wambsganss, J. (2006). *Gravitational Lensing: Strong, Weak and Micro*, volume 33 of *Saas-Fee Advanced Courses*. Springer Berlin Heidelberg.
- Schneider, P. and Seitz, C. (1995). Steps towards nonlinear cluster inversion through gravitational distortions. 1: Basic considerations and circular clusters. *A&A*, 294:411–431.
- Schneider, P., van Waerbeke, L., Jain, B. and Kruse, G. (1998). A new measure for cosmic shear. *MNRAS*,

- 296:873–892.
- Schneider, P., van Waerbeke, L., Kilbinger, M. and Mellier, Y. (2002a). Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *A&A*, 396:1–19.
- Schneider, P., van Waerbeke, L. and Mellier, Y. (2002b). B-modes in cosmic shear from source redshift clustering. *A&A*, 389(3):729–741.
- Schwarzschild, K. (1916). Über das gravitationsfeld eines massenpunktes nach der einsteinschen theorie. *Sitzungsberichte der Koeniglich Preussischen Akademie der Wissenschaften*, pages pp. 189–196.
- Scoville, N., Aussel, H., Brusa, M., Capak, P., Carollo, C.M. et al (2007). The cosmic evolution survey (cosmos): Overview. *The Astrophysical Journal Supplement Series*, 172(1):1.
- Seitz, C. and Schneider, P. (1997). Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources. *A&A*, 318:687–699.
- Seljak, U. (1996). Gravitational Lensing Effect on Cosmic Microwave Background Anisotropies: A Power Spectrum Approach. *ApJ*, 463:1.
- Semboloni, E., Hoekstra, H., Huang, Z., Cardone, V.F., Cropper, M. et al (2013). On the shear estimation bias induced by the spatial variation of colour across galaxy profiles. *Monthly Notices of the Royal Astronomical Society*, 432(3):2385.
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M.P. and McCarthy, I.G. (2011). Quantifying the effect of baryon physics on weak lensing tomography. *MNRAS*, 417:2020–2035.
- Sérsic, J.L. (1963). Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletín de la Asociacion Argentina de Astronomia La Plata Argentina*, 6:41.
- Sharma, S., Stello, D., Buder, S., Kos, J., Bland-Hawthorn, J. et al (2018). The tess–hermes survey data release 1: high-resolution spectroscopy of the tess southern continuous viewing zone. *Monthly Notices of the Royal Astronomical Society*, 473(2):2004–2019.
- Sheldon, E.S. (2014). An implementation of Bayesian lensing shear measurement. *MNRAS*, 444:L25–L29.
- Sheldon, E.S. and Huff, E.M. (2017). Practical weak-lensing shear measurement with metacalibration. *The Astrophysical Journal*, 841(1):24.
- Sheldon, E.S., Johnston, D.E., Frieman, J.A., Scranton, R., McKay, T.A. et al (2004). The galaxy-mass correlation function measured from weak lensing in the sloan digital sky survey. *The Astronomical Journal*, 127(5):2544.
- Shirasaki, M., Hamana, T. and Yoshida, N. (2015). Probing cosmology with weak lensing selected clusters – i. halo approach and all-sky simulations. *Monthly Notices of the Royal Astronomical Society*, 453(3):3043.
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L. et al (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489.
- Simet, M., McClintock, T., Mandelbaum, R., Rozo, E., Rykoff, E., Sheldon, E. and Wechsler, R.H. (2017). Weak lensing measurement of the mass–richness relation of sdss redmapper clusters. *Monthly Notices of the Royal Astronomical Society*, 466(3):3103.
- Simon, Patrick (2013). Improving three-dimensional mass mapping with weak gravitational lensing using galaxy clustering. *A&A*, 560:A33.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- Sirianni, M., Jee, M.J., Benítez, N., Blakeslee, J.P., Martel, A.R. et al (2005). The photometric performance and calibration of the hubble space telescope advanced camera for surveys. *Publications of the Astronomical Society of the Pacific*, 117(836):1049.

Bibliography

- Slipher, V.M. (1913). The radial velocity of the Andromeda Nebula. *Lowell Observatory Bulletin*, 2:56–57.
- Slipher, V.M. (1915). Spectrographic Observations of Nebulae. *Popular Astronomy*, 23:21–24.
- Smail, I., Ivison, R.J. and Blain, A.W. (1997). A Deep Sub-millimeter Survey of Lensing Clusters: A New Window on Galaxy Formation and Evolution. *ApJ*, 490:L5–L8.
- Soucail, G., Fort, B., Mellier, Y. and Picat, J.P. (1987). A blue ring-like structure, in the center of the A 370 cluster of galaxies. *A&A*, 172:L14–L16.
- Soule, F., Courbin, F. and Unser, M. (2016). Back-propagating the light of field stars to probe telescope mirrors aberrations. In *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II*, volume 9912 of *Proc. SPIE*, page 991277.
- Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J. et al (2015). Wide-Field Infrared Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *ArXiv e-prints*, 1503.03757.
- Spergel, D.N., Bean, R., Doré, O., Nolte, M.R., Bennett, C.L. et al (2007). Three-year wilkinson microwave anisotropy probe (wmap) observations: Implications for cosmology. *The Astrophysical Journal Supplement Series*, 170(2):377.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958.
- Starck, J., Murtagh, F. and Fadili, J. (2010). *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press.
- Starck, J.L. and Murtagh, F. (2006). *Astronomical Image and Data Analysis*.
- Stark, D.P., Swinbank, A.M., Ellis, R.S., Dye, S., Smail, I.R. and Richard, J. (2008). The formation and assembly of a typical star-forming galaxy at redshift $z \sim 3$. *Nature*, 455:775–777.
- Stebbins, A. (1996). Weak Lensing On the Celestial Sphere. *ArXiv Astrophysics e-prints*, astro-ph/9609149.
- STScI Development Team (2013). pynphot: Synthetic photometry software package. Astrophysics Source Code Library.
- STScI Science Definition Team (2016). Enhancing stsci’s astronomical data science capabilities over the next five years. Technical report, STScI.
- Surdej, J., Magain, P., Swings, J.P., Borgeest, U., Courvoisier, T.J.L. et al (1987). A new case of gravitational lensing. *Nature*, 329:695.
- Sutter, P.M., Elahi, P., Falck, B., Onions, J., Hamaus, N., Knebe, A., Srisawat, C. and Schneider, A. (2014). The life and death of cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 445(2):1235.
- Sutter, P.M., Lavaux, G., Wandelt, B.D. and Weinberg, D.H. (2012). A first application of the alcock-paczynski test to stacked cosmic voids. *The Astrophysical Journal*, 761(2):187.
- Suyu, S.H., Bonvin, V., Courbin, F., Fassnacht, C.D., Rusu, C.E. et al (2017). Holicow – i. h0 lenses in cosmograil’s wellspring: program overview. *Monthly Notices of the Royal Astronomical Society*, 468(3):2590.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. et al (2014). Going Deeper with Convolutions. *ArXiv e-prints*, 1409.4842.
- Szepietowski, R.M., Bacon, D.J., Dietrich, J.P., Busha, M., Wechsler, R. and Melchior, P. (2014). Density mapping with weak lensing and phase information. *MNRAS*, 440:2191–2200.
- Sánchez, C., Carrasco Kind, M., Lin, H., Miquel, R., Abdalla, F.B. et al (2014). Photometric redshift analysis in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 445(2):1482.
- Takada, M. and Jain, B. (2004). Cosmological parameters from lensing power spectrum and bispectrum

- tomography. *MNRAS*, 348:897–915.
- Taylor, A.N. (2001). Imaging the 3-D cosmological mass distribution with weak gravitational lensing. *ArXiv Astrophysics e-prints*, astro-ph/0111605.
- Tewes, M., Cantale, N., Courbin, F., Kitching, T. and Meylan, G. (2012). A fast empirical method for galaxy shape measurements in weak lensing surveys. *A&A*, 544:A8.
- Tewes, M., Courbin, F. and Meylan, G. (2013). Cosmograil: the cosmological monitoring of gravitational lenses - xi. techniques for time delay measurement in presence of microlensing. *A&A*, 553:A120.
- Theia Collaboration (2017). Theia: Faint objects in motion or the new astrometry frontier. *ArXiv e-prints*, 1707.01348.
- Tramacere, A., Paraficz, D., Dubath, P., Kneib, J.P. and Courbin, F. (2016). Asterism: application of topometric clustering algorithms in automatic galaxy detection and classification. *Monthly Notices of the Royal Astronomical Society*, 463(3):2939–2957.
- Treu, T. (2010). Strong Lensing by Galaxies. *ARA&A*, 48:87–125.
- Treu, T., Brammer, G., Diego, J.M., Grillo, C., Kelly, P.L. et al (2016). "refsdal" meets popper: Comparing predictions of the re-appearance of the multiply imaged supernova behind macsj1149.5+2223. *The Astrophysical Journal*, 817(1):60.
- Treu, T. and Koopmans, L.V.E. (2002). The internal structure of the lens pg1115+080: breaking degeneracies in the value of the hubble constant. *Monthly Notices of the Royal Astronomical Society*, 337(2):L6.
- Treu, T. and Marshall, P.J. (2016). Time delay cosmography. *The Astronomy and Astrophysics Review*, 24(1):11.
- Troxel, M. and Ishak, M. (2015). The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Physics Reports*, 558:1 – 59. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology.
- Tyson, J.A., Valdes, F., Jarvis, J.F. and Mills, Jr., A.P. (1984). Galaxy mass distribution from gravitational light deflection. *ApJ*, 281:L59–L62.
- Tyson, J.A., Valdes, F. and Wenk, R.A. (1990). Detection of systematic gravitational lens galaxy image alignments - Mapping dark matter in galaxy clusters. *ApJ*, 349:L1–L4.
- Vale, C., Hoekstra, H., van Waerbeke, L. and White, M. (2004). Large-scale systematic signals in weak lensing surveys. *The Astrophysical Journal Letters*, 613(1):L1.
- Vale, C. and White, M. (2003). Simulating weak lensing by large-scale structure. *The Astrophysical Journal*, 592(2):699.
- van Daalen, M.P., Schaye, J., Booth, C.M. and Dalla Vecchia, C. (2011). The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 415(4):3649.
- Van De Weygaert, R. and Platen, E. (2011). Cosmic voids: Structure, dynamics and galaxies. *International Journal of Modern Physics: Conference Series*, 01:41–66.
- Van Waerbeke, L. and Mellier, Y. (2003). Gravitational Lensing by Large Scale Structures: A Review. *ArXiv Astrophysics e-prints*, astro-ph/0305089.
- Van Waerbeke, L., Mellier, Y., Erben, T., Cuillandre, J.C., Bernardeau, F. et al (2000). Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. *A&A*, 358:30–44.
- Van Waerbeke, L., Mellier, Y., Radovich, M., Bertin, E., Dantel-Fort, M. et al (2001). Cosmic shear statistics and cosmology. *A&A*, 374(3):757–769.

Bibliography

- Vegetti, S., Lagattuta, D.J., McKean, J.P., Auger, M.W., Fassnacht, C.D. and Koopmans, L.V.E. (2012). Gravitational detection of a low-mass dark satellite galaxy at cosmological distance. *Nature*, 481:341–343.
- Vegetti, S. and Vogelsberger, M. (2014). On the density profile of dark matter substructure in gravitational lens galaxies. *Monthly Notices of the Royal Astronomical Society*, 442(4):3598.
- Velander, M., van Uitert, E., Hoekstra, H., Coupon, J., Erben, T. et al (2014). Cfhtlens: the relation between galaxy dark matter haloes and baryons from weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 437(3):2111.
- Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA. ACM.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. and Manzagol, P.A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408.
- Viola, M. (2014). The probability distribution of ellipticity: implications for weak lensing measurement. *Proceedings of the International Astronomical Union*, 10(S306):94–98.
- Viola, M., Cacciato, M., Brouwer, M., Kuijken, K., Hoekstra, H. et al (2015). Dark matter halo properties of gama galaxy groups from 100 square degrees of kids weak lensing data. *Monthly Notices of the Royal Astronomical Society*, 452(4):3529.
- Viola, M., Kitching, T.D. and Joachimi, B. (2014). On the probability distributions of ellipticity. *Monthly Notices of the Royal Astronomical Society*, 439(2):1909.
- Viola, M., Melchior, P. and Bartelmann, M. (2011). Biases in, and corrections to, ksb shear measurements. *Monthly Notices of the Royal Astronomical Society*, 410(4):2156.
- Voigt, L.M. and Bridle, S.L. (2010). Limitations of model-fitting methods for lensing shear estimation. *MNRAS*, 404:458–467.
- Voigt, L.M., Bridle, S.L., Amara, A., Cropper, M., Kitching, T.D., Massey, R., Rhodes, J. and Schrabback, T. (2012). The impact of galaxy colour gradients on cosmic shear measurement. *Monthly Notices of the Royal Astronomical Society*, 421(2):1385.
- von Soldner, J. (1804). On the deflection of a light ray from its rectilinear motion. *Berliner Astronomisches Jahrbuch*, pages pp. 161–172.
- Waerbeke, L.V. (2000). Noise properties of gravitational lens mass reconstruction. *Monthly Notices of the Royal Astronomical Society*, 313(3):524.
- Walker, A.G. (1935). On Riemannian spaces with spherical symmetry about a line, and the conditions for isotropy in general relativity. *The Quarterly Journal of Mathematics*, 6:81–93.
- Walsh, D., Carswell, R.F. and Weymann, R.J. (1979). 0957 + 561 A, B - Twin quasistellar objects or gravitational lens. *Nature*, 279:381–384.
- Wang, B., Abdalla, E., Atrio-Barandela, F. and Pavón, D. (2016). Dark matter and dark energy interactions: theoretical challenges, cosmological implications and observational signatures. *Reports on Progress in Physics*, 79(9):096901.
- Weinberg, D.H., Bullock, J.S., Governato, F., Kuzio de Naray, R. and Peter, A.H.G. (2015). Cold dark matter: Controversies on small scales. *Proceedings of the National Academy of Science*, 112:12249–12255.
- Weinberg, D.H., Mortonson, M.J., Eisenstein, D.J., Hirata, C., Riess, A.G. and Rozo, E. (2013). Observational probes of cosmic acceleration. *Physics Reports*, 530(2):87 – 255. Observational Probes of Cosmic Acceleration.

- Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*.
- Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23.
- Welikala, N. and Kneib, J.P. (2012). Color Gradients in Galaxies Out to $z \sim 3$: Dependence on Galaxy Properties. *ArXiv e-prints*, 1202.0494.
- White, R.J. and Ghez, A.M. (2001). Observational Constraints on the Formation and Evolution of Binary Stars. *ApJ*, 556:265–295.
- Wilson, G., Kaiser, N., Luppino, G.A. and Cowie, L.L. (2001). Galaxy halo masses from galaxy-galaxy lensing. *The Astrophysical Journal*, 555(2):572.
- Wittman, D. (2002). Weak Lensing. In Courbin, F. and Minniti, D., editors, *Gravitational Lensing: An Astrophysical Tool*, volume 608 of *Lecture Notes in Physics*, Berlin Springer Verlag, page 55.
- Wittman, D.M., Tyson, J.A., Kirkman, D., Dell’Antonio, I. and Bernstein, G. (2000). Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. *Nature*, 405:143–148.
- Wong, K.C., Suyu, S.H., Auger, M.W., Bonvin, V., Courbin, F. et al (2017). H0licow – iv. lens mass model of he 0435–1223 and blind measurement of its time-delay distance for cosmology. *Monthly Notices of the Royal Astronomical Society*, 465(4):4895.
- Wu, K.K.S., Lahav, O. and Rees, M.J. (1999). The large-scale smoothness of the Universe. *Nature*, 397:225–230.
- Xu, D., Sluse, D., Gao, L., Wang, J., Frenk, C., Mao, S., Schneider, P. and Springel, V. (2015). How well can cold dark matter substructures account for the observed radio flux-ratio anomalies. *Monthly Notices of the Royal Astronomical Society*, 447(4):3189.
- Zehavi, I., Zheng, Z., Weinberg, D.H., Blanton, M.R., Bahcall, N.A. et al (2011). Galaxy clustering in the completed sdss redshift survey: The dependence on color and luminosity. *The Astrophysical Journal*, 736(1):59.
- Zhou, N. (2017). Volvo admits its self-driving cars are confused by kangaroos. *The Guardian*.
- Zuntz, J., Kacprzak, T., Voigt, L., Hirsch, M., Rowe, B. and Bridle, S. (2013). im3shape: a maximum likelihood galaxy shear measurement code for cosmic gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 434(2):1604.
- Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.
- Zwicky, F. (1937a). Nebulae as gravitational lenses. *Phys. Rev.*, 51:290–290.
- Zwicky, F. (1937b). On the probability of detecting nebulae which act as gravitational lenses. *Phys. Rev.*, 51:679–679.

Thibault Kuntzer

PhD student in data science for astrophysics



CONTACT INFORMATION

Rue de Fontaine-André 50
2000 Neuchâtel
++41 (0)79 763 31 83
thibault.kuntzer@ssvnn.net
<https://www.linkedin.com/in/thibault-kuntzer>

- ✓ Data analysis and machine learning
- ✓ Goal-oriented and hard working
- ✓ Team leader

LANGUAGES

French : Mother tongue
English : Certificate in Advanced English (C1). Lived for one year in London.
German : Good level of written and oral German (B2).

PROFESSIONAL EXPERIENCE

PhD Thesis in Physics – Ecole Polytechnique Fédérale de Lausanne

Laboratoire d'Astrophysique (LASTRO)

2013-2018

- PhD thesis in observational cosmology. Supervisor: Dr. F. Courbin.
- Weak gravitational lensing effect: implementation of new machine-learning techniques to measure properties of galaxies much faster and more accurately.
- Determination and reconstruction of the Point Spread Function (PSF) for the ESA Euclid space telescope mission: quantification and detection of adversarial effects for reconstruction, performance improvements in empirical reconstruction and interpolation using auto-encoders and neural networks.
- Designing a new classifier of stellar images by neural networks for space-based telescopes which uses only single-band images instead of multiple images.
- Responsible for operation and observations for 30 nights at the Swiss Euler ground-based telescope in La Silla (Chile).

Member of Instrument Working Group of ESA CHEOPS satellite

2013-Current

- Responsible for determination and analysis of the observable sky.
- Performance analysis based on the CHEOPS satellite science requirements.
- Optimisation of the planned orbit of CHEOPS to maximise science returns.

Teaching assistant in Physics & Space Technologies

2011-Current

Preparing problems, leading exercise sessions and examinations for courses in fundamental physics, introduction to astrophysics (bachelor level; also its online version), Space Mission Design & Operations (master level; also online) and tutoring students during their practical works and master project in astrophysics and machine learning.

Duvas Technologies Ltd., London, supervisor Dr S. Wilkins

2009-2010

- Development and testing of prototype NO & NO₂ sensing equipment.

COMPUTER SKILLS

Python, C/C++, Fortran, PHP, HTML, Matlab, LaTeX, git, unix, TensorFlow, notions of MySQL, Arduino, STK, R.

EDUCATION	Master of Sciences in Physics with Minor in Space Technologies	2011-2013
	EPFL, Swiss Space Center, University of Bern	
	<ul style="list-style-type: none"> • Thesis: "Simulation of Stray Light Contamination on CHEOPS Detector," Earth stray light contamination simulation and analysis according to mission profile. Determination of the observable sky. Supervisors: Dr. A. Fortier, Prof. W. Benz. • Summer school 2013 at Bauman Moscow State Technical University (Russia) "Space Development: Theory & Practice," Russia-USA-Switzerland exchange program. Multidisciplinary group works in the field of space mission design. 	
	Bachelor in Physics	2007-2010
MILITARY CAREER	ERASMUS student at Imperial College London.	2009-2010
	Was awarded the <i>Imperial College International Diploma in Physics</i> .	
	Ecole Polytechnique Fédérale de Lausanne.	2007-2009
	Officer of the Swiss Armed Forces	
	<ul style="list-style-type: none"> • Staff officer in the Space Cell. 	2017-Current
	<ul style="list-style-type: none"> • Staff officer at the battalion level in the Air Defence Corps. 	2014-2016
	<ul style="list-style-type: none"> • Platoon leader. 	2011-2014
CIVIL STATUS	Swiss national Born on 25 October 1988	
LEISURE	<ul style="list-style-type: none"> • Swimming Instructor (J+S). • Several certificates in rescue swimming (UK and Switzerland). • Member and competitor (since 2003) of the Neuchâtel Lifesaving Society (SSVNN). Occupied several committee and subcommittee posts. • Sailing. • Kayaking. 	