Model based forecasting for demand response strategies

Thèse N° 9415

Lorenzo NESPOLI

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

"Our world is increasingly complex, often chaotic, and always fast-flowing.
This makes forecasting something between tremendously difficult
and actually impossible, with a strong shift toward the latter
as timescales get longer."
— Andrew McAfee

"There is no way that we can predict the weather six months ahead
beyond giving the seasonal average"
— Stephen Hawking

To my parents, my brother, and my whole family…

Acknowledgements

Acknowledgements

Finally I would like to thanks my parents for their unconditional support they gave me, Miriam for her endless caring, both my brothers Federico and Tenchi, my old date friends Fabrizio, Simone, Filippo, Alessandro and all the other friends who make me feel home every time I come back, and Giulio.

Lugano, 11 March 2019   Lorenzo Nespoli.

Abstract

The incremental deployment of decentralized renewable energy sources in the distribution grid is triggering a paradigm change for the power sector. This shift from a centralized structure with big power plants to a decentralized scenario of distributed energy resources, such as solar and wind, calls for a more active management of the distribution grid. Conventional distribution grids were passive systems, in which the power was flowing unidirectionally from upstream to downstream. Nowadays, and increasingly in the future, the penetration of distributed generation (DG), with its stochastic nature and lack of controllability, represents a major challenge for the stability of the network, especially at the distribution level. In particular, the power flow reversals produced by DG cause voltage excursions, which must be compensated. This poses an obstacle to the energy transition towards a more sustainable energy mix, which can however be mitigated by using a more active approach towards the control of the distribution networks. Demand side management (DSM) offers a possible solution to the problem, allowing to actively control the balance between generation, consumption and storage, close to the local point of generation. An active energy management implies not only the capability to react promptly in case of disturbances, but also the ability to anticipate future events and take control actions accordingly. This is usually achieved through model predictive control (MPC), which requires a prediction of the future disturbances (in this case, produced and consumed energy) acting on the system.

This thesis treats challenges of distributed DSM, with a particular focus on the case of a high penetration of PV power plants. The first subject of the thesis is the evaluation of the performance of models for forecasting and control with low computational requirements, of distributed electrical batteries. The proposed methods are compared by means of closed loop deterministic and stochastic MPC performance. The second subject of the thesis is the development of model based forecasting for PV power plants, and methods to estimate these models without the use of dedicated sensors. The third subject of the thesis concerns strategies for increasing forecasting accuracy when dealing with multiple signals linked by hierarchical relations. Hierarchical forecasting methods are introduced and a distributed algorithm for reconciling base forecasters is presented. At the same time, a new methodology for generating aggregate consistent probabilistic forecasts is proposed. This method can be applied to distributed stochastic DSM, in the presence of high penetration of rooftop installed PV systems. In this case, the forecasts' errors become mutually dependent, raising difficulties in the control problem due to the nontrivial summation of dependent random variables. The benefits of considering dependent forecasting errors over considering them as

Abstract

independent and uncorrelated, are investigated. The last part of the thesis concerns models for distributed energy markets, relying on hierarchical aggregators. To be effective, DSM requires a considerable amount of flexible load and storage to be controllable. This generates the need to be able to pool and coordinate several units, in order to reach a critical mass. In a real case scenario, flexible units will have different owners, who will have different and possibly conflicting interests. In order to recruit as much flexibility as possible, it is therefore important to design incentive mechanisms that guarantee participants to benefit from their participation, while at the same time achieving a technical optimum. Since I didn't want to include any users' active decision nor belief in the formation of the market equilibrium price, or to model users' willingness to pay and their marginal utilities, I chose not to consider auction mechanisms. I instead used the well known concept of Lagrangian relaxation from optimization theory, and the interpretation of Lagrangian multipliers as marginal prices, as a way to automatically identify prices for the grid's constraints. I further avoided to model users' utility as intended in standard auction and game theory, and replaced it with costs and users' constraints sets. In fact, the latter can be interpreted as a binary and non differentiable utility function, and prevent us from making any assumption on users' marginal satisfaction with respect to consumed energy. The focus is on two approaches, the first one relying on distributed control theory, while the second one generates from non-cooperative games. I show that for the class of problems of our interest, the so called sharing problems, the two formulations are deeply linked, and can thus be treated using the same decomposition techniques.

Sinossi

La progressiva adozione di fonti energetiche rinnovabili decentralizzate nella rete di distribuzione sta innescando un cambiamento di paradigma per il settore energetico. Il passaggio da una struttura centralizzata a uno scenario di produzione di energia elettrica tramite fonti distribuite, come il solare e l'eolico, richiede una gestione più attiva della rete di distribuzione. Le reti di distribuzione erano considerate sistemi passivi, in cui la potenza fluisce unidirezionalmente dalla fonte di produzione al consumatore. Oggigiorno, e sempre più in futuro, la penetrazione della generazione distribuita (GD), con la sua natura stocastica e la mancanza di controllabilità, rappresenta una grande sfida per la stabilità della rete, soprattutto a livello della rete di distribuzione. In particolare, le inversioni del flusso di potenza prodotte da GD provocano escursioni di tensione, che devono essere compensate. Ciò rappresenta un ostacolo alla transizione energetica verso un mix energetico più sostenibile, che può tuttavia essere mitigato utilizzando un approccio più attivo verso il controllo delle reti di distribuzione. Il Demand Side Management (DSM) offre una possibile soluzione al problema, consentendo di controllare attivamente l'equilibrio tra generazione, consumo e stoccaggio. Una gestione energetica attiva implica non solo la capacità di reagire prontamente in caso di disturbi, ma anche la capacità di anticipare eventi futuri e intraprendere azioni di controllo di conseguenza. Questo è solitamente ottenuto attraverso modelli di controllo predittivo, che richiedono una previsione dei disturbi futuri che agiscono sul sistema. Questa tesi tratta le sfide del DSM distribuito, con particolare attenzione al caso di un' alta penetrazione di impianti fotovoltaici. Il primo argomento della tesi è la valutazione delle performance di modelli con requisiti computazionali limitati per il controllo di batterie elettriche distribuite. I metodi proposti sono confrontati per mezzo delle loro prestazioni a ciclo chiuso, sia per un controllo deterministico che stocasticho. Il secondo argomento della tesi è lo sviluppo di metodi per la previsione della potenza generata da impianti fotovoltaici, basati su modelli fisici, e la stima degli stessi senza l'uso di sensori dedicati. Il terzo argomento indagato nella tesi sono strategie per aumentare l'accuratezza della predizione della potenza e consumo elettrico in vari punti della rete di distribuzione, avendo a disposizione dati storici collegati da relazioni gerarchiche. In particolare, è presentato un metodo per calcolare in modo distribuito diverse techniche di riconciliazione per predittori gerarchci. In concomitanza, è presentata una nuova metodologia per l'aggregazione consistente delle distribuzioni di probabilità descriventi previsioni della generazione di potenza. Questo metodo può essere applicato al DSM stocastico distribuito, in presenza di elevata penetrazione di impianti fotovoltaici. In questo caso infatti, gli errori delle previsioni diventano mutualmente dipendenti, inducendo difficoltà nel problema di controllo,

dovute alla non trivialità della somma di distribuzioni di probabilità dipendenti. I vantaggi di considerare gli errori di previsione come dipendenti, sono mostrati nella tesi. L'ultima parte della tesi riguarda modelli per i mercati elettrici distribuiti, basati su aggregatori gerarchici. Per essere efficace, il DSM richiede il controllo di un gran numero di utenti. Questo genera la necessità di essere in grado di raggruppare e coordinare diversi consumatori e produttori nella rete elettica, al fine di raggiungere una massa critica. In un caso reale, le risorse di flessibilità avranno proprietari diversi, con interessi economici conflittuali. Al fine di reclutare la massima flessibilità possibile, è quindi importante progettare meccanismi di incentivazione che garantiscano ai partecipanti di beneficiare della loro partecipazione, ottenendo allo stesso tempo un optimum tecnico. Poiché non volevo delegare nessun ruolo attivo agli utenti nella determinazione del prezzo di equilibrio del mercato, né modellare la loro utilità marginale, ho scelto di non considerare meccanismi d'asta. Ho invece usato concetti di ottimizzazione e in particolare la teoria dei moltiplicatori di Lagrange per determinare il prezzo legato ai vincoli di rete in modo automatico. Ho inoltre evitato di modellare l'utilità degli agenti, come intesa nella teoria delle aste a dei giochi, sostituendola con costi e set di vincoli per gli utenti. Questi ultimi possono infatti essere considerati come funzioni di utilità binarie non differenziabili, e permettono di non fare nessun assunzione sulla soddisfazione marginale degli utenti nel consumare energia. In particolare, ho concentrato l'attenzione su due approcci diversi: il primo basato sulla teoria del controllo distribuito, il secondo basato sulla teoria dei giochi non cooperativi. Nella tesi viene dimostrato che per la classe di problemi di nostro interesse, i cosiddetti sharing problems, le due formulazioni sono profondamente collegate e possono quindi essere trattate usando le stesse tecniche di decomposizione.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

$ADMM$	alternating direction method of multipliers
AOI	angle of incidence
BRP	balance responsible party
$C + I$	consensus plus innovation
CC	chance constraint
$D - OPF$	distributed OPF
DER	distributed energy resources
DG	distributed generation
DHI	diffuse horizontal irradiance
DNI	direct normal irradiance
DR	Douglas-Rachford
DR	demand response
DSM	demand side management
DSO	distributed system operator
DW	Dantzig-Wolfe
ELM	extreme learning machine
FBF	Forward Backward Forward
FFT	fast Fourier transform
GHI	global horizontal irradiance
GNE	generalized Nash Equilibrium
HP	heat pump
HW	Holt-Winters
IAM	incidence angle modifier
iid	independent and identically distributed
ISO	independent system operator
ISO	independent system operator
KPI	key performance indicator
LV	low voltage
MAE	mean absolute error

Nomenclature

$MIMO$	multiple input multiple output
$MISO$	multiple input single output
MPC	model predictive control
MV	medium voltage
NE	Nash equilibrium
$nMAE$	normalized mean absolute error
$nRMSE$	normalized root mean squared error
NWP	numerical weather prediction
ODE	ordinary differential equation
OPF	optimal power flow
PF	power flow
pFB	preconditined Forward Backward
PV	photovoltaic
PV	photovoltaic
QRF	quantile regression forest
$RMSE$	root mean squared error
$RQRF$	recursive quantile regression forest
SB	scenario based
SCC	self consumption community
$SMPC$	stochastic MPC
$SOPF$	stochastic OPF
$TBSMPC$	tree based stochastic MPC
TSO	transmission system operator
UC	unit commitment
UN	uncontrolled loads
VCG	Vickrey-Clarke-Groves
VNE	variational Nash Equilibrium
WM	welfare maximization

Chapter 1 - Background and state of the art

$\langle x, y \rangle$	scalar product between x and y
ϵ	disturbance
$\hat{F}$	empirical cumulative density function
$\mathbf{prox}(\cdot)$	proximal operator
$\mathscr{F}$	$\sigma-$field
$\mathscr{M}$	model
$\mathscr{N}$	normal distribution
$\mathscr{X}$	constraint set
$\mathbb{E}$	expectation operator
$\mathbb{I}$	indicator function

$\mathbb{P}$	probability measure
μ	mean
Ω	sample space
Φ	Gaussian cumulative function
π	node probability
Σ	covariance matrix
σ	standard deviation
τ	rooted tree
D_{te}	testing dataset
D_{tr}	training dataset
F	cumulative density function
$f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$	operator/map/multi-valued function
F^{-1}	inverse cumulative density function (quantile)
l	loss function
n_b	number of bootstrapped dataset
n_{obs}	number of observations
n_s	number of scenarios
p	probability density function
q_α	quantile of level α
T	prediction horizon
t	time
t_{end}^h	ending time of hour h
t_{start}^h	starting time of hour h
U	uniform distribution
u_i	cost function of agent i
x	decision variable
x^*	optimal solution
x_{-i}	concatenation of all agents' actions but the ones from agent i: $[x_1^T, x_2^T, \ldots, x_{i-1}^T, x_{i+1}^T, \ldots, x_N^T]^T$

Chapter 2 - Multi-step-ahead forecasting for demand response applications

α	tilt angle of PV field
$\oplus$	direct sum operator
$\dagger$	pseudoinverse opeartor
$\dot{m}$	inlet mass flow
$\dot{Q}_{buo}$	heat power due to buoyancy effect
$\dot{Q}_{cond}$	heat power due to conduction
$\dot{Q}_h$	heat power due to heating element (electric resistance)
$\dot{Q}_{loss}$	heat power exchanged with the ambient
η_t	combined module and inverter efficiency
$\hat{P}$	forecasted power profile

Nomenclature

λ	Ridge regularization coefficient for the ELM
$\mathscr{D}_{k,j}$	set of descendants of node j at timestep k
$\mathscr{U}$	control variable constraint set
$\mathbb{E}_D$	expectation with respect to the dataset D
$\theta_{a,t}$	azimuth angle of the sun at time t
$\theta_{z,t}$	zenith angle of the sun at time t
$\otimes$	Kronecker product
Φ_x	enthalpy flow inside the heating pipe, at position x
$\sigma(\cdot)$	activation function for the ELM
τ^{μ_i,μ_j}	tree structure with μ_i nodes in the first step ahead and μ_j nodes in the last step ahead
θ	matrix of output weights for the ELM
A_d	discrete dynamics state matrix for the controlled system
B_d	discrete dynamics control matrix for the controlled system
C	heat capacity of the water layer
$c(\cdot)$	energy cost function
c_p	specific heat capacity of water
I	radiation on a given surface
I_b	direct radiation on a given surface
I_d	diffuse radiation on a given surface
I_g	ground reflected radiation on a given surface
$I_{i,t}$	oedinary differential equation
I_{STC}	reference irradiance under standard conditions
k	thermal conductivity of water
K_{re}^*	expectation of the ratio between the closed loop performance gap and the prescient controller's performance
K_{re}^*	expected ratio of the performance gap of the stochastic controller and the performance gap of the deterministic controller, w.r.t. the prescient controller
K_{re}^*	ratio between the expectation of the closed loop performance gap and the expectation of the prescient controller's performance
K_{re}^*	ratio between the expected performance gap of the stochastic controller and the expected performance gap of the deterministic controller, w.r.t. the prescient controller
K_{i,j,n_a}^*	controller closed loop performance for a given E-rate, capacity and number of aggregation steps, normalized with the performance of the uniform MPC formulation
M_x	block diagonal matrix for the reduction of the regressors dataset
M_y	block diagonal matrix for the reduction of the target dataset
n_a	number of aggregated steps for the nonuniform stepsize MPC formulation
n_k	number of original steps which are used for the k_{th} aggregation
n_x	number of regressors
O	output matrix of the first layer of the ELM

p_b	energy buying price
p_s	energy selling price
S	summation matrix for the battery operations
$S(\cdot)$	quantile skill score
T_a	ambient temperature
T_{cell}	cell temperature of a given PV module
u_{amb}	equivalent thermal loss coefficient between the tank and the ambient
W	matrix of inner layer wights for the ELM
w_k	objective function re-weighting coefficient for the k_{th} aggregation
x_{start}	initial state of the controlled system

Chapter 3 - PV modeling for power forecasting

ϵ_{cl}	ratio between the estimated GHI and extraterrestrial irradiance in the reference period
$\hat{\omega}$	identified PV model coefficients
$\hat{\omega}_{bl}$	PV model coefficients, blindly identified from composite power flow
E_t	extra-terrestrial irradiance at time t
P_l	power consumption of the loads
P_m	composite power measurement
s_t	clear sky indicator at time t

Chapter 4 - Hierarchical forecasts for distributed control algorithms

$\hat{P}_{bu}$	forecast of the aggregated profile, obtained with a bottom up approach
$\mathscr{F}$	Fourier transform
$\rho_{i,j,k}$	absolute value of the correlation coefficient between forecast errors of profile i and j at the k_{th} step ahead
θ_{el}	sun elevation angle
$f_m(\cdot)$	forecast function for NWP scenario generation

Chapter 5 - Distributed energy markets

α	repartition coefficient
λ	Lagrangian multiplier
$e(\cdot)$	system-level objective function

Introduction

Motivations of the thesis

The incremental deployment of decentralized stochastic generators at the low and medium voltage levels of the distribution grid introduces a number of new challenges for the safe operation of the network, like for example voltage fluctuations and line congestions [1; 2; 3; 4]. Among the proposed solutions to this problem, an active control of distributed loads and storage, also known as demand side management (DSM), has been widely suggested [5; 6; 7; 8]. For an effective management of the network through the actuation of distributed flexible loads and storage systems, it is imperative to be able to coordinate their actions efficiently. Substantial, authoritative work addresses the theme of DSM and demand response (DR) from a theoretical standpoint [9; 10; 3; 11; 12], analyzing the best strategies for agent coordination towards an optimal aggregate behaviour on single grid levels. But, as flexibility will be offered at different levels and will provide a number of services, from voltage control for the distributed system operator (DSO) to control energy for the transmission system operators (TSOs), it is important to make sure that these services will not interfere with each other. It is clear that offering services to a TSO or a balance responsible party (BRP) can very realistically have an impact on power quality, both locally and at a distance. While TSOs objectives are focused on the flawless operation of the bulk grid, (e.g. reserve scheduling and congestion management in the transmission grid), DSOs and independent system operators (ISOs) are more concerned with power quality in terms of bounded nodal voltages, line congestion, islanding of local portions of the grid and local dispatch. The coordination needed for such services implies substantial load correlation and openly challenges the traditional assumption of "statistical smoothing", under which DSOs have designed their grids; let alone the effects of the ever-increasing penetration of photovoltaic (PV) systems at the low voltage level. At the same time, the stochasticity and volatility of distributed energy resources (DERs) power generation is pushing electricity markets towards smaller clearance times, with respect to the day ahead planning used in the Day Ahead Market. Furthermore, as DSM and DR rely on prosumers owned assets with small starting time compared to centralized power plants, the solution paradigm is shifting from solving unit commitment (UC) problems to the one of solving multistage stochastic problems, using a rolling horizon fashion. This paradigm shift

in the modality of grid usage requires a certain degree of system overhaul. In particular, one needs to ensure that DSM does not create congestions or voltage violations at any point of the distribution grid. A comprehensive approach towards the actuation of flexibility, taking into account its effects at different grid levels, is proposed in this thesis. In this work, I chose to investigate distributed control techniques for the problem of coordinating the flexibilities in the grid. Compared to centralized techniques, distributed control offers the advantage of being more scalable and allowing to preserve partially the privacy of the agents. On the other hand, distributed control presents a number of challenges, among which the need to decompose the optimization problem and distribute it among the agents, who need to solve parts of it locally, often on a hardware, which has limited computational power. To successfully apply distributed control techniques for the coordination of a set of agents represented by electrical loads or batteries, one needs to rely on the forecasts of each agents' electrical consumption/production with relatively high time resolutions. Since both power consumption and DERs power production profiles have a daily seasonality, a 24 hours ahead planning is typically used. The high number of time steps, the frequency at which the problem must be solved, the number of agents to be coordinated (in the range of hundreds) and the limited computational power of the devices on which the distributed control problem is solved (smart meters), require a careful selection of both optimization strategy and forecasting algorithm. The efficient forecast of a high number of relatively small loads and generators is particularly challenging. Part of this thesis is specifically dedicated to the design and evaluation of forecasting techniques for production and consumption. Both the accuracy and the computational requirements of the proposed techniques is evaluated. A particular focus went on PV forecasting, since PV represents most of DERs. Another challenge when forecasting relatively small loads and generators is the lack of monitoring equipment. Often, PV production is not measured separately and it simply adds negatively to the aggregated power consumption. In this work, disaggregation techniques that allow disaggregating PV production from power consumption are proposed and evaluated. When multi objective optimization techniques are applied to optimize for both a local (e.g. single household level) and a global (aggregate power of a neighborhood) objective function, consistency in the forecasted power profiles is required. This means that the forecasts of the single agents should sum up to the forecast of their aggregate. This consistency is particularly challenging to obtain when considering probabilistic forecasts. For this reason I propose a method to retrieve probabilistic forecasts which are aggregate-consistent by construction, without the need of empirically modeling a large number of probabilistic interdependencies between forecasters, as the state of the art suggests.

Contributions and structure of the thesis

In what follows, the main contributions of the thesis are presented, referencing the chapters in which they are treated, grouped by the four problematics introduced in the motivation.

Low computational complexity methods for predictive control In chapter 2, control and forecasting techniques with low computational requirements are compared to more sophisticated and computational intensive methods. In particular, we investigated the loss of performance of a deterministic model predictive controller (MPC) using forecasters of a computationally cheap parametric forecaster, with respect to a tree based stochastic model predictive controller (TBSMPC) exploiting conditional probability density functions provided by a quantile random forest (QRF).

- In section 2.2, a nonuniform stepsize MPC formulation is presented. Logarithmically spaced steps are used, and their number is systematically increased while evaluating the controller performance using synthetically generated residential power profiles. The aim of the control step reduction is twofold, lowering the controller computational time and both the forecasters training dataset and training computational time . We show that, when re-weighting the objective function for the length of the averaging bins, the maximum relative increase in the objective function with respect of the full formulation using 96 steps, is well below 1%, using only 10 steps.

- In section 2.3, two new forecasters with low training computational time and memory requirement are presented: a detrended Holt-Winter (HW) forecaster and a bagging of extreme learning machines. The HW provided strictly better a-priori performance for the first step-ahead forecasts, with respect to all the other models. When evaluating the forecasters a-posteriori, by means of closed loop performances of an MPC controller, the Holt-Winter forecasters shows performances which are close to the one provided by the best forecaster, with the advantage of being computationally cheaper. When the a-posteriori evaluation is done using a TBSMPC controller, the tests shown that the forecaster based on a QRF regressor consistently provide better performance, at the price of a higher computational time for the model training and memory requirements.

Influence of PV modeling on forecasting As a growing number of roof-mounted PV system is being installed, in section 3.2 we investigate the effect of exploiting physics-based PV models in order to forecast their power production. Forecasting PV power is of great importance for the future electrical grid, as more accurate power predictions allows to better handle abrupt rump up in regional power flow, due to change in cloud configuration, and ultimately permit to increase the number of PV plants which can be hosted in the distribution grid. Estimating physical models of existing PV systems in an automatic and unsupervised way, has the additional benefit of turning PV panels into irradiance sensors. In section 3.1.2 we introduce a new method to reconstruct the GHI using only AC power from a PV power plant. Chapter 2, treating these subjects, is partially based on the annexed papers:

[A] *F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised Disaggregation of Photovoltaic Production from Composite Power Flow Measurements of Heterogeneous Prosumers," IEEE Trans. Ind. Informatics, 2018.*

[B] *L. Nespoli and V. Medici, "An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements," Solar Energy, 2017.*

The main outcomes of the thesis on this topic, contained in chapter 3, are the following:

- A methodology to blindly identify a physical model of PV power plants, starting from composite power signals, are introduced in sec 3.1. The different methods are explained in detail in paper A.

- In section 3.1.2, a new unsupervised method for estimating the GHI from AC photo-voltaic power measurements is introduced. The detailed procedure is presented in paper B, and its improved accuracy with respect to satellite-based irradiance estimations is reported, for two case study.

- It is shown how, combining physical models and QRF, the accuracy of predicting PV output from meteorological conditions increases significantly. Moreover, blindly identify the PV model starting from composite power measurements does not significantly decrease the prediction accuracy.

- It is shown how modeling PV does help to increase the forecast accuracy, only for steps ahead between 30 minutes and 12 hours, period in which NWP are more reliable.

Hierarchical forecasting techniques   When applying DSM, we are typically interested in minimizing an objective function which depends on the aggregated power profiles of a group of agents in the distribution grid, while respecting grid constraints. This requires to separately forecast agents' power profiles. In this case hierarchical forecasting techniques can be used to improve the forecasts' accuracy. When we are interested in probabilistic forecasts, the problem becomes harder, requiring in general a multidimensional integral. The main contribution in hierarchical forecasting of this thesis, contained in chapter 4, are the following:

- In section 4.1, a new distributed method to reconcile forecasters at different levels of a hierarchical structure is presented. This method can be used to make forecasts done by different entities aggregate-consistent, thus usable in distributed control. The main advantage in redistributing the reconciliation is that private information which could be used by the base forecasters, is not disclosed. Furthermore, informations at upper levels of the hierarchical structure is only available by means of aggregate power profiles.

- In section 4.2, a new method to obtain aggregated consistent pdfs for hierarchical power forecasts is presented. We show that nontrivial methods for summing the bottom level forecasts' pdf are needed especially in the case of high penetration of PV. In this case forecasting errors becomes dependent, due to imperfect NWP.

Distributed energy market   Two of the challenges of distributed DSM are the ability of providing multiple ancillary services at different voltage levels of the distribution grid, and trustless coordination of agents. For example, an independent group of prosumers can provide congestion management at the MV level and voltage control at LV level for a DSO, sell its aggregated flexibility to BRPs to help them meeting previously committed power profiles on the spot or intraday market, or sell provide secondary or tertiary control to the TSO. Another interesting case is the one of self consumption community (SCC), in which a group of prosumers connected to the main grid through a single point of coupling can pay its electricity bill as a single entity. This means that SCCs are pushed to increase their self consumption in order to low their total energy expenses. These thematics are briefly introduced in chapter 5. We propose a multilevel hierarchical distributed algorithm, which makes use of voltage sensitivity coefficients in order to respect grid constraints. Since distributed control algorithm based on problem decomposition are prone to malicious attacks and manipulations [13; 14; 15], in this chapter we analyze the proposed algorithm and show it can be easily turned into a non-cooperative game with a unique Nash equilibrium. Moreover, we propose a method to enforce individual rationality, which is, the condition for which the energy market participants are always better off opting in.

The two subjects are separately treated in these two annexed papers:

[C] *L. Nespoli and V. Medici, "Constrained hierarchical networked optimization for energy markets," IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2018.*

[D] *L. Nespoli, M. Salani, and V. Medici, "A rational decentralized generalized Nash equilibrium seeking for energy markets," in 2018 International Conference on Smart Energy Systems and Technologies, SEST 2018 - Proceedings, 2018.*

List of related publications   The following publications are related to, but not included in this thesis.

- *V. Medici, M. Salani, L. Nespoli, A. Giusti, N. Vermes, M. Derboni, E. Rizzoli, D. Rivola, "Evaluation of the Potential of Electric Storage Using Decentralized Demand Side Management Algorithms," in Energy Procedia, 2017.*

- *L. Nespoli, A. Giusti, N. Vermes, M. Derboni, E. Rizzoli, and L. M. Gambardella, "Distributed demand side management using electric boilers," Computer Science - Research and Development, 2016.*

- *L. Nespoli, V. Medici, and R. Rudel, "Grey-Box System Identification of Building Thermal Dynamics Using only Smart Meter and Air Temperature Data," 14th IBPSA Conf., 2015.*

1 Background and state of the art

Part of this thesis concerns the evaluation of different forecasting techniques by means of optimal (stochastic) control performance. This evaluation requires the knowledge of a broad variety of topics, which however, share the common mathematical background of optimization. For instance, in chapter 2 we will use a quadratic cost function for the evaluation of different probabilistic forecasters, which is justified by the common need of evaluating proximal operators for different class of distributed control algorithms. Lagrangian duality framework is used in chapter 4 to distribute the technique of hierarchical forecasting reconciliation, as well as to obtain the distributed control algorithms in chapter 5. Monotone operator theory, on which the alternating direction method of multipliers (ADMM) is based, is used to demonstrate the uniqueness of Nash equilibrium for the class of non-cooperative games generated by sharing problems in chapter 5. The concept of rooted trees is used in chapter 5 in order to formulate a multilevel hierarchical coordination algorithm spanning multiple voltage level of the distribution grid, as well as in the tree based stochastic MPC framework to have a compact description of the evolution of the control problem's uncertainties. For this reason, this chapter gives an overview on the interconnections between forecasting, stochastic and distributed control, and aims at grouping the common mathematical tools which have been used in different parts of the thesis.

1.1 General patterns - optimal power flow

One of the key problems faced by independent system operators (ISOs) when considering intra day operation or day-ahead market clearing is the unit commitment problem (UC) [16], in which the operations of a set of generators has to be planned in advance, and the ISO commits to the optimal scheduling for the next day. Usually, the UC incorporates some kind of optimal power flow problem (OPF) [17; 18; 19; 20]. Based on the specific constraint which are considered, the OPF problem is known by different names, as network constrained, security constrained, alternate current, direct current OPF. A review on the different kinds of OPF formulations can be found in [21]. As the power generation get more decentralized and uncertain, due to DERs and REs, ISOs and and distribution system operators (DSOs) have

moved from the traditional deterministic OPF and UC formulation, to ones which incorporate higher levels of uncertainty and to decentralized solution strategies. A comprehensive review on the different state of the art methods to deal with uncertainty power system studies and stochastic OPF (SOPF) can be found in [22; 23], while [24] present a more specific review on stochastic formulations of UC problems. In [25] a partial survey on distributed approaches to UC, OPF, optimization and approximated PF formulations is presented.

Several formulations of the OPF exist, based on the type of the considered grid (high or medium voltage, radial or meshed, balanced or unbalaced) and control objective. In order to motivate the modeling choice that I use in chapter 5, in the following I introduce the general formulation of the OPF. Given an electrical grid composed by a set of n buses (or nodes), we refer to the optimal power flow problem to the task of minimizing an objective $f(S_c)$, which is function of the complex powers $S_c \in \mathbb{C}^{n_c}$ injected in the set of controllable buses $\mathcal{N}_c = 1, 2.., n_c, n_c < n$, subject to bus voltage consistency, power balance and operational constraints.

$$
\begin{aligned}
&\min_{S_c \subset S} f(S_c) \\
&s.t. \ \ I = YV \\
&\qquad S = V \odot I^* \\
&\qquad V \in \mathcal{V}, \quad S \in \mathcal{S}, \quad I \in \mathcal{I}
\end{aligned}
\tag{1.1}
$$

where $\odot$ is the Hadamard product, Y is the admittance matrix, I^* stands for the complex conjugate of the currents' vector and $\mathcal{V}, \mathcal{S}$ and $\mathcal{I}$ are operational constraint sets for the voltages, complex power and currents. While the OPF described in (1.1) uses the so called bus injection model, an equivalent formulation known as branch flow model can be found in the literature [26]. The latter is especially useful for the formulation of convex relaxations of the OPF [27], or for its distributed computation [28]. Unfortunately, while (1.1) can be reliably solved for high and medium voltage grids, where the lines' parameters needed to obtain the admittance matrix Y are usually known, this is not the case for low voltage networks. Furthermore, solving (1.1) requires the knowledge of current and voltage phasors for the controlled nodes in the grid. In this thesis I assume that phasors' measurements are not available, and that the OPF can only be solved relying on magnitude measurements of V and I, provided by residential smart meters. For these reasons, instead of solving the OPF directly, chapter 5 considers a relaxed formulation of (1.1), in which the knowledge of phasors' angles is not needed. The relaxed formulation is based on the voltage sensitivity coefficients, which are the first order approximation of the power flow equations. These are introduced in details in subsection 5.2.1, where limitations of this formulation are also discussed. Under these assumptions, the

relaxed OPF becomes:

$$
\begin{aligned}
\min_{P,Q} \ & f(P_c) \\
s.t. \ & |V| = V_0 + K_p \Delta P + K_q \Delta Q \\
& |V| \in \mathcal{V}, \quad P \in \mathcal{P}, \quad Q \in \mathcal{Q}
\end{aligned}
\tag{1.2}
$$

where K_p and K_q are voltage sensitivity coefficients matrices associated with active (P) and reactive (Q) power, V_0 is a reference voltage magnitude vector and Δ is the first order discrete time difference operator.

1.2 Mathematical preliminaries and notation

Through the thesis we will make use mathematical concepts for which different notations are reported in the literature. For example, the nomenclature for probability spaces and vector maps are often inconsistent through the literature. Here the mathematical notation used in the rest of the thesis, which was made as consistent as possible, is reported.

Random variables x defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ where Ω is the sample space $\mathscr{F}$ is a $\sigma-$field, and $\mathbb{P}$ is a probability measure, are reported without subscripts, whereas the same variable with a subscript, x_k, denotes a realization, which is, a random draw from the probability density function (pdf) $p(x)$. The cumulative density function (cdf) of a continuous pdf is

$$
F(z) = \mathbb{P}(x < z) = \int_0^z p(x) \mathrm{d}x
\tag{1.3}
$$

while the empirical cdf is defined as

$$
\hat{F}(z) = \frac{1}{1 + n_{obs}} \sum_{k=1}^{n_{obs}} \mathbb{I}_{\{x_k < z\}}
\tag{1.4}
$$

where $\mathbb{I}_{\{\cdot\}}$ is the indicator function and n_{obs} is the number of observed realizations of x. The α quantile of F_x is defined as the generalized inverse of F

$$
q_\alpha = F^{-1}(\alpha) = inf\{z \in \mathbb{R}, F(z) \geq \alpha\}
\tag{1.5}
$$

Given a multivariate random variable, $F(x|y)$ and $p(x|y)$ are the conditional cdf and pdf, respectively, for which the Bayes' theorem holds:

$$
p(y|x)p(x) = p(x|y)p(y)
\tag{1.6}
$$

where $p(x, y)$ is the joint pdf and $p(x)$ is also known as the marginal pdf of x. The expectation operator is denoted as $\mathbb{E}[\cdot]$, and $\mathbb{E}_D[\cdot]$ denotes the expectation with respect to the dataset D. The set of integers $\{k_1, k_2, ... k_N\}$ is denoted as $\mathbb{N}_{[k_1:k_N]}$, and $[x_k]_{k=1}^N = [x_1^T, x_2^T ... x_k^T]^T$ is the collection of x_k from $k = 1$ to N. $f : \mathbb{R}^n \to \mathbb{R}$ indicates a real value function, while $f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ indicates an operator, also known as a map, multi-valued function or correspondence, mapping $\mathbb{R}^n$ onto itself. For example $\left[\partial_k V(x, y_k)\right]_{k=1}^N$ indicates the subdifferential of the map $V(x, y) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. The scalar product is denoted as $\langle x, y \rangle$, or equivalently $x^T y$ for the Euclidean space. The notation $\|x\|_2^2$ stands for the sum of squares of x, while $\| \|_p$ indicates the p norm.

1.3 Stochastic model predictive control formulations for power systems

As anticipated in the introduction, since DSM and DR rely on prosumers' owned assets with small start up and shot down times and costs compared to highly inertial traditional power plants, in order to reliably control these assets in the LV or MV gird, there is no need of committing to a certain scheduling of the latter [29]. For this reason in this thesis we will focus on the optimization of intra-day operations, and we won't solve the UC problem, which is usually formulated as a two stage stochastic program and solved via L-shaped or Bender decomposition method, but we will restrict the study to stochastic OPF formulations. Considering multiple timesteps, both the previously introduced versions of the OPF, (1.1) ad (1.2), can be expressed as:

$$x_k^* = \operatorname*{argmin}_{x \in \mathscr{X}} \mathbb{E}\left[\sum_{t=1}^T l(x_{t+k|k}, \epsilon_{t+k|k})\right] \tag{1.7}$$

where $\mathbb{E}$ is the expectation operator over the random variable ϵ, which is defined over the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, x is the decision vector (containing both states and control actions), $\mathscr{X}$ is a bounded set, representing (probabilistic) operational constraints, T is the number of timesteps in the control horizon and l is a loss function to be minimized. We stress out that, since the objective function (and possibly $\mathscr{X}$) depends on ϵ, the optimal decision variable x_k^* will also depend on the disturbance ϵ. For intra-day optimization, (1.7) is intended to be applied in a receding horizon fashion, thus that we can draw from the vast literature on stochastic MPC (SMPC). In general, due to lack of knowledge and unboundedness of the sample space Ω, problem (1.7) is intractable, and must be therefore approximated. Three main strategies to approximate the stochastic control problem (1.7) are present in the literature and we briefly review them in the following. Namely, they are chance constraint (CC) analytic approximation, scenario based (SB) SMPC and tree-based (TB) SMPC. It must be noted that system operators are usually interested in solving a robust optimization problem rather than a stochastic one. Indeed the distinction between robust and stochastic optimization rely on how

we define the disturbance ϵ and the constraint set $\mathscr{X}$ (as a function of ϵ). Robust optimization usually considers an unknown distribution of ϵ, but a bounded sample space Ω, such that $\epsilon \in [\epsilon_{min}, \epsilon_{max}]$. A solution is than found for which $x \in \mathscr{X}(\epsilon)$ is satisfied for all the values of $\epsilon \in \Omega$. In other words, robust optimization find the optimal solution x^* considering worst case scenarios in terms of the disturbance, which is a wise optimization strategy in the case in which a violation of the constraint set implies irreversible damage to the operated system, as is usually the case in power grids. The CC approach tries to enlarge the feasible set of the solution, allowing probabilistic violations of the constraint set, making some assumptions on the pdf of ϵ. These assumptions are avoided in the SB approach, which uses a number of scenarios to model ϵ. Basic SB optimization do not uses chance constraints, that is, all the scenarios add hard constraints to the main problem. Anyway, is easy to see that a link exists between the number of considered scenarios and the probability to respect the constraints in the worst case. Indeed, as the number of scenarios increases, the solution under SB optimization approaches the one obtained using robust optimization. For the three aforementioned strategy, only one vector x^* is retrieved, which contains the optimal actions to apply to the controlled system up to time $k + T$. A third way of taking into account uncertainty is to encode ϵ into a disturbance tree, spreading with increasing time. This technique, TBSMPC, allows to represent the natural evolution of uncertainty, which typically increases as we consider further points in time, and to retrieve a rooted tree of control actions, of which only the first component is actually applied. In the following we briefly introduce them in order to motivate the use of TBSMPC in the rest of the thesis.

Analytical approximations of chance constraints   Instead of strictly requiring $x \in \mathscr{X}$ for all possible realizations of ϵ, we can relax this hard constraint and reformulate it in terms of probabilistic constraints, also known as chance constraints:

$$
\begin{aligned}
x_k^* &= \underset{x}{\operatorname{argmin}} \, \mathbb{E} \left[\sum_{t=1}^{T} l(x_{t+k|k}, \epsilon_{t+k|k}) \right] \\
&s.t.: \quad \mathbb{P}[x \in \mathscr{X}] > 1 - \delta
\end{aligned}
\tag{1.8}
$$

Unfortunately, also problem (1.8) is in general computationally intractable [30], especially when $\mathscr{X}$ contains more than one constraint (e.g., it's a polytopic), joint probabilistic constraints would require the computation of a multidimensional integral over an unbounded probability space, and must be therefore approximated, for example, using convex analytic approximations, as in [31]. A more straightforward result is available in the case of a polytopic CC:

$$
\mathbb{P}\left[g^T x \le h\right] \ge 1 - \delta
\tag{1.9}
$$

If the expected value and covariance matrix W of the disturbance are known, the Chebyshev - Cantelli inequality can be used [32; 33], which guarantees that (1.9) is (conservatively) satisfied

if

$$g^T x \leq h - \sqrt{g^T W g} f(\delta) \tag{1.10}$$

where $f(\delta) = \sqrt{(1-\delta)/\delta}$. Condition (1.10) can be made less conservative under Gaussian assumption of ϵ, using $f(\delta) = \mathcal{N}^{-1}(1-\delta)$. Given the results in [34], under the assumption of lognormal distribution, in [35] joint polytopic CC are conservatively reformulated as a set of single deterministic constraints. Another notably approximation, given the results in [30], in the case of disturbance generated by i.i.d. Gaussian random variables and SMPC with affine disturbance feedback is presented in [36].

Scenario based stochastic optimization   When no assumption on the disturbance ϵ can be made, or the CC formulations result in a too conservative control problem, one can revert to sampling based methods [31]. In scenario based SMPC, a number of scenarios is drawn from a stochastic process, or obtained by sampling with the methods explained in section 1.4. In scenario based SMPC, the probabilistic constraint in 1.8 is replaced with a set of deterministic ones:

$$
\begin{aligned}
x_k^* = \underset{x}{\text{argmin}}\, \mathbb{E} &\left[\sum_{t=1}^{T} l(x_{t+k|k}, \epsilon_{t+k|k}^i) \right] \\
s.t.: \quad x + \epsilon_{t+k|k}^i &\in \mathscr{X} \quad \forall i \in \mathbb{N}_{[0,N_s]}, \forall t \in \mathbb{N}_{[0,T]}
\end{aligned}
\tag{1.11}
$$

where $\epsilon_{t+k|k}^i$ is the i_{th} scenario of the disturbance at time t. It is easy to see that, as the number of scenarios increases, if the scenarios are drawn from the exact pdf of the disturbance, x_k^* will tend to the solution of a robust optimization problem. In order to loose the constraints, and obtain a less conservative solution, the number of scenarios must be carefully evaluated. It can be shown that the number of scenarios N_s which guarantees that the probabilistic constraints in (1.8) is satisfied with reliability δ, can be found analytically [37; 38]. The authors in [39] present a review on methods to encode chance constraints in SBMPC.

TBSMPC   Instead of approximating problem (1.7) with a set of scenarios, we can describe the temporal evolution of the disturbance ϵ with a rooted tree, denoted as τ. In this way we can retrieve a tree of control actions, with the same structure of the disturbance characterization. In other words, this method returns n_s optimal control vectors x^*, where n_s is the number of leaves in the tree structure. Intuitively, this approach would result in a less conservative control with respect to the SB approach, since we can optimize over a greater number of control actions, each of which considers a branch in the evolution of the disturbance tree. This allows to take less conservative actions as the number of considered scenarios (or tree leaves) increases, as opposed to the SB approach. On the other hand, as in the SB approach, constraints in each branch of the tree are treated as (hard) deterministic constraints, meaning that increasing the number of leaves lead to consider extreme events as in robust optimization.

This method originates from the multistage stochastic programming [40; 41] literature, which is an extension of two-stage stochastic programming. When applied to MPC, this technique is also known as TBSMPC, and has been successfully applied in the literature to a broad class of problems, from intra-day energy management, to dynamic option hedging and drainage water systems control [42; 43; 44; 45; 46]. In [47; 35] two comparisons with numerical results of SBMPC and TBSMPC are presented. Different approaches can be adopted in order to retrieve the scenario tree used to describe the disturbance of the problem. This is usually obtained by clustering a high number of scenarios, which can be obtained with the method described in 1.4. The problem of generating an optimal tree from a set of scenarios, in the sense of minimizing the distance between distributions, is NP hard [48], and thus the scenario reduction could result in a high computational time when the number of original scenarios N_s and the length of the considered horizon H are high. In the rest of the thesis we have adopted the backward reduction algorithm described in [49; 41], which uses a greedy strategy in order to minimize the Kantorovich distance between the trees and the original sets of scenarios. The probabilities of reaching each node from the root, π_i, are found aggregating the probabilities of the originally merged scenarios, which by construction guarantees that:

$$\sum_{i=1}^{\mu_k} \pi_{k,j} = 1 \tag{1.12}$$

where k is the step ahead, and μ_k is the number of nodes for the step k. In order to give a flexible description of the resulting rooted tree, the algorithm has been rewritten in python and coupled with the networkx package, which encodes the disturbance structure in a directed acyclic graph (DAG), allowing to use high level graph searches queries (e.g. list the ancestors of a given node). An example of scenario tree representation is shown in Fig. 1.1. The code is freely available at `https://gitlab.com/supsi-dacd-isaac/scenred`.

1.4 Multivariate probabilistic forecasting

Conditional pdfs can then be used to create scenarios, which are used in scenario based stochastic optimization, as explained in 1.3. Scenarios can be produced in different ways, depending on the type of algorithm used for obtaining the posterior pdf. In chapter 2 we will use three different techniques, which we introduce here. We don't want to model the error with a particular probability distribution, since this will require to introduce assumptions in the analysis of the forecasters performance. We will only use non-parametric techniques for the production of the scenarios.

Bootstrap   Bootstrap [50] was introduced to perform non-parametric inference on independent and identically distributed (iid) data, and lately become a very popular method for assessing statistical accuracy and model fitting [51]. The bootstrap consist in repeatedly drawing with replacement from a dataset D, in order to estimate some relevant statistical properties. Since in this thesis we are interested in power forecasts, which typically shows a
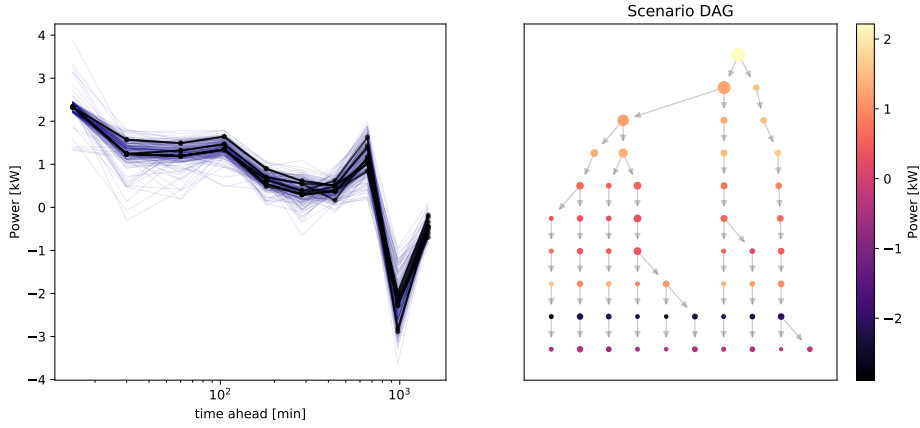
Figure 1.1 – Example of scenario reduction results and encoding of the scenario tree into a DAG. Left: 200 scenarios generated for the prediction of 24 hours ahead power profile of a building. Blue lines: original scenarios. Black lines: scenario tree. Right: the same scenario tree, depicted as a DAG. The size of the dots its proportional to the intra-step probabilities.

daily seasonality, we slightly modify the bootstrap technique to take this into account. In order to generate statistically consistent scenarios for the power prediction, we fit a forecaster model $\mathcal{M}$ on a training dataset D_{tr}, and then we retrieve the forecaster's error on D_{tr}, $\epsilon_{tr} \in \mathbb{R}^{n_{obs,tr} \times T}$, where $n_{obs,tr}$ is the number of observations in the training set and T is the horizon of the prediction. The scenarios for each of the $n_{obs,te}$ entries of the test dataset D_{te}, on which the control policy is then evaluated, are retrieved performing bootstrap on ϵ_{tr}, based on the hour of the day.

$$y_{t,scen} = \hat{y}_t + [\hat{\epsilon}_h]_{s \in \mathbb{N}_{[0,n_s]}} \qquad t \in \left[t_{start}^h, t_{end}^h \right] \tag{1.13}$$

where $\hat{y}_t \in \mathbb{R}^H$ is the univariate forecast of length H at time t, $y_{t,scen} \in \mathbb{R}^{H \times n_s}$ is the matrix of scenarios for time t, $[\hat{\epsilon}_h]_{s \in \mathbb{N}_{[0,n_s]}}$ is the set of errors bootstrapped from the training set at hour h and $\left[t_{start}^h, t_{end}^h \right]$ are the starting and ending time of hour h, based on the used sampling time. We the bootstrap technique for retrieving conditional pdfs for the detrended Holt-Winter (HW) forecaster. Bootstrapping for HWs is suggested also by [52], and a bootstrapping option is implemented in the R library fpp2 for ARIMA and HW forecasters.

Bagging of multivariate forecaster   When combined to model fitting, bootstrap aggregation is also known as bagging: the same class of models $\mathcal{M}$ is fitted n times, applying bootstrap to the dataset D. Applying bootstrapping at the training dataset D_{tr}, and fitting a model for each of the n_b bootstrapped dataset, will result in a set of models $\left[\mathcal{M}_{D_{tr,b}}\right]_{b \in \mathbb{N}_{[0,n_b]}}$. The prediction of the bagging model is then the average output of the n models:

$$\hat{y} = \frac{1}{n_b} \sum_{b=1}^{n_b} \mathcal{M}_{D_{tr,b}}(D_{te}) \tag{1.14}$$

We will use this is the technique used in chapter 2 to build bagging of extreme learning machine (ELM) forecasters. Since ELMs can predict multivariate output, that is, the output of each ELM is the T-steps long power forecast, we could directly use the prediction of each n_b models as a scenario.

Copula coupling   Some forecaster algorithms cannot be used to generate multivariate output, but only to predict an univariate response. This is the case of econometric models, such as AR, ARMA, ARIMAX. The usual approach for scenario generation of these class of forecasters, and state space based models in general, is to assume that the noise follows some class of random process (e.g. Brownian motion) and generate as many scenarios as needed, through a Monte Carlo simulation, as done in [53] and suggested in [54]. Since we don't want to make hypothesis on the error pdf, and use non-parametric methods to estimate it, we use instead the method described in [55], which generates scenarios linking the posterior pdfs of the prediction with a multivariate Gaussian covariance structure. This method is more flexible, since can be applied to forecasters which return skewed (non-Gaussian) pdfs as predictions, and do not assign a prior distribution to the error's pdf.

In brief, the method can be described as follows. Given a set of T posterior cdf, $F_t(x)$ $t \in \mathbb{N}_{[0,T]}$, one for each step in the prediction horizon, we cannot directly build scenarios by Monte Carlo simulation independently sampling from each distribution. In fact, this will cause the disruption of the inter-temporal statistical properties of the forecasts. Instead, we can estimate the temporal correlation of scenarios by means of the sample mean vector $\mu \in \mathbb{R}^T$ and sample covariance matrix $\Sigma \in \mathbb{R}^{T \times T}$, obtained by the historical values in the training set D_{tr}, and then draw from the Gaussian multivariate distribution the realization of the random variable $z \sim \mathcal{N}(\mu, \Sigma)$. This provide us with a consistent inter-temporal structure between timesteps, which we can then superimpose to the $F_t(x)$ by means of the so called probability integral transformation, which essentially states that sampling a cdf F using a random variable drawn from the pdf of F, $z \sim p$, will result in a random variable which is uniformly distributed:

$$F(z) \sim U[0,1] \tag{1.15}$$

where $U[0,1]$ is the uniform pdf on the interval $[0,1]$. Given 1.15, we can then sample from the multivariate Gaussian $z \sim \mathcal{N}(\mu, \Sigma)$, apply the Gaussian cumulative function Φ to obtain T univariate uniform distributions, which we can then use to sample the posterior cdfs $F_t(x)$. Formally, the s_{th} scenario is obtained using:

$$\hat{y}_{t,s} = \left[F_t \left(\Phi(z_t) \right)^{-1} \right]_{t=1}^T \tag{1.16}$$

where z_t is the t_{th} component of the realization drawn from x.

We stress out that this procedure is different from assuming a multivariate Gaussian distribution for the errors, which won't allows to retain different pdfs for each step ahead. This methods can also be interpreted as a multivariate Gaussian copula approach, as described

in [56], and ensemble copula coupling [57; 58]. In section 4.2 we will mention again copulas as means of estimating pdfs of sum of random variables, starting from marginal pdfs. These methods can also be interpreted as sample reordering techniques [59]. A review on other techniques for scenario generation, based on the forecasting method, can be found in the working paper [60].

1.5 Distributed coordination of agents

In this section, we briefly introduce proximal operators and monotone operators. These concepts are used in various parts of the thesis, especially in chapter 5 for the coordination of distributed agents.

1.5.1 Proximal operators

The proximal operator, also known as Moreau's proximal mapping, has been widely used for the decomposition of convex problems. Proximal algorithms have analogies with Newton's methods, but sit at a higher level of abstraction [61]. Formally, given a function $f(x) : \mathbb{R}^n \to \overline{\mathbb{R}}$, its (rescaled) proximal operator is defined as:

$$\mathbf{prox}_{\gamma f}(v) = \operatorname*{argmin}_{x} f(x) + \frac{1}{2\gamma} \|x - y\|_2^2 \tag{1.17}$$

where γ is a scaling parameter, weighting the second part of the objective function. The proximal operator can be interpreted intuitively in terms of its role in minimizing the function f iteratively. At each iteration, $\mathbf{prox}_{\gamma f}$ maps $\mathbb{R}^n$ to a point inside the domain of f. The solution of the iteration is a compromise between minimizing f and moving from the current point. This, among the fact that $\mathbf{prox}_{\gamma f}$ has always a unique minimizer, being a closed, bounded and strictly convex function, gives the proximal operator suitable properties for being used in optimization algorithms. For instance, the proximal operator can be seen as a single step of the backward Euler method [61], and can be thus described as an implicit method. In this thesis, we will mainly use two properties of proximal algorithms. The first one states that, if f is fully separable, i.e. $f(x) = \sum_i^n f_i(x_i)$, then also the proximal operator can be computed in a separable way:

$$\left(\mathbf{prox}_{\gamma f}(v) \right)_i = \mathbf{prox}_{\gamma f_i}(v_i) \tag{1.18}$$

This makes the proximal operator suitable for decomposition algorithms. The second notable property is that the proximal of an indicator function, $\mathbb{I}_{x \in \mathcal{X}}$, reduces to the Euclidean projection into the set $\mathcal{X}$:

$$\mathbf{prox}_{\mathbb{I}_{x \in \mathcal{X}}}(v) = \operatorname*{argmin}_{x \in \mathcal{X}} \|x - v\|_2 \tag{1.19}$$

1.5.2 Monotone operators

The theory of monotone operators is strictly linked with the derivation and analysis of distributed convex algorithm. However, in this thesis we are mostly interested in their role in solving variational inequalities and their implications for non-cooperative n-persons games. We will only consider operators mapping $\mathbb{R}^n$ onto itself. For a more general introduction, see [62]. An operator, or relation, or set-valued mapping $\mathscr{F} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, is a map from $\mathbb{R}^n$ to the set $\mathscr{F}(x) \subset \mathbb{R}^n$. An operator is called monotone, and respectively β-strongly monotone, if:

$$
\begin{aligned}
\langle u - v, x - y \rangle &\geq 0 \quad \forall\, x \neq y \in \mathbb{R}^n, \quad u \in \mathscr{F}(x), \quad v \in \mathscr{F}(y) \\
\langle u - v, x - y \rangle &\geq \beta \|x - y\|_2 \quad \forall\, x \neq y \in \mathbb{R}^n, \quad u \in \mathscr{F}(x), \quad v \in \mathscr{F}(y)
\end{aligned}
\tag{1.20}
$$

In paper D, we will deal with a particular operator, called the game mapping, or pseudo-gradient. Given a game, defined by the agents' costs $u_i(x_i, x_{-i}) : \mathbb{R}^n \times \mathbb{R}^{n(N-1)} \to \mathbb{R}$, where N is the number of players, the game mapping is given by:

$$
\mathscr{F}(x) = \left[\partial_{x_i} u_i(x_i, x_{-i}) \right]_{i=1}^N
\tag{1.21}
$$

which represents the vector containing the subdifferentials of the agent's costs, when considering the actions of the other agents as fixed. If $\mathscr{F}(x)$ is (strictly) monotone, it can be shown that the game admits a (unique) generalized variational Nash equilibrium (GNE) [63; 64]. This means that there exists a (unique) solution x^* to the variational inequality:

$$
\langle \mathscr{F}(x^*), x - x^* \rangle \geq 0 \quad \forall x \in \mathscr{X} \subset \mathbb{R}^{nN}
\tag{1.22}
$$

where $\mathscr{X}$ is a closed and convex set. In other words, given the definition of $\mathscr{F}$, there exists a value of x, which encodes all the agents' actions, x^*, for which no agents can reduce their own costs if all the others agents keep their own actions fixed and inside the feasible set $\mathscr{X}$. This is the definition of a GNE. To summarize, the monotonicity of the game mapping $\mathscr{F}$, along with the convexity with respect to x_i of the cost functions $u_i(x_i, x_{-i})$, is a sufficient condition for the existence of a a solution of the variational inequality (1.22), which is a GNE for the considered game. However, the inverse does not hold in general, and for this reason this resolution concept is known to be a 'refinement' of the GNE. The monotonicity of the game mapping $\mathscr{F}$ will be used in paper D to show that the so called sharing problems generate games for which a unique variational GNE exists.

2 Multi-step-ahead forecasting for demand response applications

In this chapter, we investigate the performance of different classes of forecasters for the prediction of the 24 hours ahead power generation and production of residential users. The comparison is done evaluating the forecasters in k-fold cross validation, both a-priori and a-posteriori. The a-priori evaluation is done by means of KPIs on the distance of the forecasts from the actual observed values, while the a-posteriori evaluation is done by means of closed-loop control performance, which is obtained using both a deterministic and a TBSMPC controller. In section 2.1 the dynamic models used to generate a residential power profile dataset are presented. The modeling choices were made with the aim of keeping a low complexity in the parametrization of the models, in order to obtain representative but generic and adaptable components, while producing component-level power profiles with arbitrary time resolution which could be used to evaluate the impact of smart controllers on the state of the grid using a bottom-up approach. In section 2.2, a nonuniform stepsize strategy for MPC and TBSMPC is presented, aiming to reduce the computational time for both control and forecasting. The feasibility of the proposed strategy and performance losses are evaluated using perfect forecasts. In section 2.3 different techniques for multistep ahead forecasts are presented, and performance investigated.

The main outcomes of this chapter are summarized in the following:

- In section 2.2, a nonuniform stepsize MPC formulation is presented. Logarithmically spaced steps are used, and their number is systematically increased while evaluating the controller performance using synthetically generated residential power profiles. We show that, when re-weighting the objective function for the length of the averaging bins, the maximum relative increase in the objective function evaluated in closed loop, reduces from 35% to 1.2% in the case in which 7 timesteps are used, while passing from 25% to 0.5% in the 10 timesteps case.

- In section 2.3, four different methods are used to forecast residential power profiles. We propose two new forecasters: a detrended Holt-Winter forecaster and a bagging of extreme learning machines. The first forecaster provided strictly better a-priori

performance for the first step-ahead forecasts, with respect to all the other models. When evaluating the forecasters a-posteriori, by means of closed loop performances of an MPC controller, the Holt-Winter forecasters shows performances which are close to the one provided by the best forecaster, with the advantage of being computationally cheaper. When the a-posteriori evaluation is done using a TBSMPC controller, the tests show that the forecaster based on a quantile random forest regressor consistently provide better performance, at the price of a higher computational time for the model training.

2.1 Synthetic power profile generation

To our knowlege, one of the most fungible datasets for residential energy consumption forecasting is the Apartment dataset of the UMass Smart* Dataset [65]. Unfortunately, these data come from american consumers, which are not very representative of European energy consumptions. Moreover they do not possess PV installations, which we would like to include in the analysis. Due to the scarcity of energy consumption datasets, researchers in this field have recently tried to produce synthetic datasets for forecasting and analysis, based on statistical analysis or on simulation [66][67].
We decided to follow this strategy, producing our dataset through a dynamic simulation. For this task we built the code from scratch relying on the standard scipy ODE integrator.
In the following we give a detailed description of the reference system we have considered for the simulation. Since it is not of general interest to describe all the possible configurations of the simulated systems, we just describe a typical configuration, on which all the simulated systems are based. In order to obtain a representative dataset for Switzerland, we used the STASCH6 standard [68] and its variants as a reference for the heating system and the control logic.

Heating system and control logic

The STASCH6 standard comprehends 3 main components: an heatpump (HP), a water tank used as an energy buffer, and a heating element delivering heat to the building. The HP control logic is based on two temperature sensors placed at different heights of the water tank, while the circulation pump connecting the tank with the building's heating element is controlled by an hysteresis on the temperature measure by a sensor placed inside the house.
We describe the control logic in a sequential way, following the heating components of the system. The first decision is taken by the building central controller, which decides its working mode, that is, if the building needs to be cooled or heated, based on a moving average of the

historical data of the external temperature:

$$
\begin{cases}
wm_t & = -1 \quad \textbf{if} \quad T_{ma,t} > T_{max,ma} \\
wm_t & = 1 \quad \textbf{if} \quad T_{ma,t} < T_{min,ma} \\
wm_t & = 0 \quad otherwise
\end{cases}
\tag{2.1}
$$

where the working mode wm_t is negative when the building requires to be cooled, positive when heating is required, and 0 when no actions are needed. $T_{max,ma}$ and $T_{min,ma}$ represent the maximum and minimum values of the external temperature's moving average, which is based on the past 7 days. The actual activation of the heating element is controlled by the hysteresis on the internal temperature of the building, T_z. If the working mode is positive, this is given by:

$$
\begin{cases}
s_{hy,t} & = 1 \quad \textbf{if} \quad \left(T_z < T_{min,hy} - \Delta T/2 \right) \\
& \qquad \textbf{or} \quad \left(T_z < T_{min,hy} + \Delta T/2 \quad \textbf{and} \quad s_{hy,t-1} \right) \\
s_{hy} & = 0 \quad otherwise
\end{cases}
\tag{2.2}
$$

where $s_{hy,t}$ is the state of the hysteresis at time t, 1 meaning that the circulation pump of the heating element must be activated, and DT was chosen to be equal to $1°C$. For completeness, we report also the control logic when the building is in cooling mode:

$$
\begin{cases}
s_{hy,t} & = 1 \quad \textbf{if} \quad \left(T_z > T_{max,hy} + \Delta T/2 \right) \\
& \qquad \textbf{or} \quad \left(T_z > T_{max,hy} - \Delta T/2 \quad \textbf{and} \quad s_{hy,t-1} \right) \\
s_{hy} & = 0 \quad otherwise
\end{cases}
\tag{2.3}
$$

The incoming water temperature in the heating element is then modulated linearly through a 3-way valve between a maximum and minimum value, based on the external temperature, both in the heating and cooling modes. When operative, the heating element requests hot or cold water to the water tank, which control logic is based on two temperature sensors located in two different layers. When the building is in heating mode, the control logic is a simple hysteresis based on the temperature of the sensor in the uppermost layer, which is identical to the one in (2.2). When in cooling mode, the control logic is the following:

$$
\begin{cases}
s_{hy,t} & = -1 \quad \textbf{if} \quad \left(T_{up} > T_{max}^c + \Delta T/2 \right) \\
& \qquad \textbf{or} \quad T_{low} > T_{max}^c + \Delta T/2 \\
s_{hy,t} & = 0 \quad \textbf{if} \quad \left(T_{low} < T_{min}^c \right) \textbf{ or } \left(T_{up} < T_{max}^c - \Delta T/2 \right) \\
s_{hy,t} & = s_{hy,t-1} \quad otherwise
\end{cases}
\tag{2.4}
$$

where T_{up} and T_{low} are the temperature measured by the upper and lower sensors, respectively, and T_{min}^c and T_{max}^c are the minimum and maximum desired temperatures of the water in the tank while in cooling mode.

The value of $s_{hy,t}$ is then communicated to the HP. In the case in which the HP is also used for

the domestic hot water (DHW), the DHW tank is always served with priority by the HP.

Building model

We modeled the building thermal dynamics with a simple one state RC equivalent model, as done in [69]. The main reason for this choice is that it is hard to generalize RC models with higher number of states, since no values can be found in the literature for the needed parameters. Estimating an RC model from data requires different measurements of temperatures, internal and solar gains, at a resolution of at least 10 minutes. This kind of datasets are extremely hard to find, and limited to only a few, often undwelled, cases. These equivalent RC circuit parameters could, in theory, be estimated starting from first principles, however recently proposed studies show that this can give worse results then estimating a model from data [70]. The second reason is that, while a higher order model leads in general to smaller one step ahead residuals compared to a lower order model, the loss of accuracy passing from a one state model to an higher order one when considering a longer period of simulation is much lower [71]. Last, when considering RC models for buildings with a number of states higher than 3, the chances of overfitting are high, and additional measurements such as the heat fluxes between thermal zones are required to guarantee observability. Alternatively, pseudo-random binary sequences can be applied to the heating systems in order to excite the system in a wide range of frequencies [72], while being uncorrelated with other exogenous inputs. This technique induces high changes in internal temperature of the building and cannot clearly be applied to occupied buildings.

We adopted the following methodology to retrieve representative R values for the single state RC equivalent model. We retrieved the distribution of year of construction for residential buildings in Switzerland from the swiss Federal Statistical Office [73]. We then combined them with the estimated mean heating needs per squared meter, based on construction year [74]. In this way we obtained a distribution of energy demand for heating, $E_d \left[kWh/m^2/year \right]$. We then retrieved the U values per squared meter (the inverse of the R parameter per squared meter), dividing E_d for 1500 equivalent hour of the building's heating system per year. This gave us the estimated U value distribution in $\left[kW/m^2 \right]$. In Fig. 2.1, the final distribution for Switzerland is shown. For comparison, we also plotted the distribution of the declared U values from the EU28 members, which is available for the year 2014 at [75]. For additional comparison, we identified a one state RC equivalent circuit from a monitored building located in Biel-Benken. The model is the following;

$$ C\frac{\partial T_z}{\partial t} = \frac{T_{ext} - T_z}{R} + kQ_h + A_{eq}I_s \tag{2.5} $$

where T_{ext} is the the external temperature, R is the equivalent thermal resistance for the building, k is a parameter weighting the estimated power coming from the heating system Q_h, I_s is the incoming solar radiation and A_{eq} is the estimated equivalent window area.
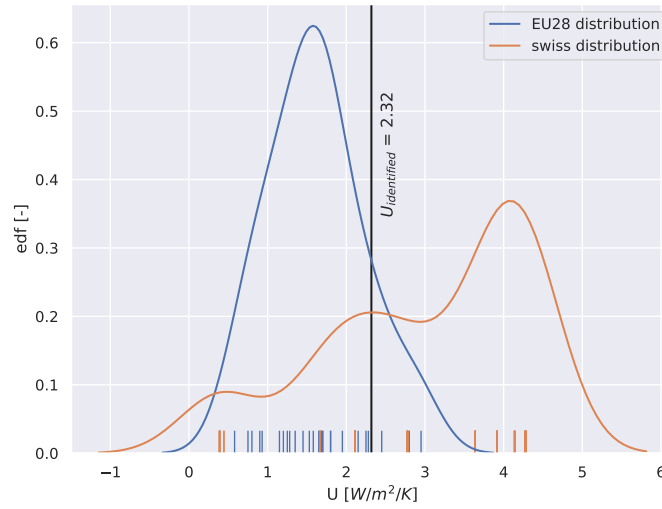
Figure 2.1 – Comparison of U value distributions for Switzerland and for the member of EU28. The vertical line shows the identified U value for a monitored building located in Biel-Benken

Floor heating

Modeling floor heating requires to simulate an N-states system, since the temperature of the water in a given point of the serpentines depends in a non-trivial way on all the temperatures of the previous portion of the serpentines and of the surrounding floor. Furthermore, simulating the temperature of the water in the serpentine in a dynamic way could lead to prohibitive computational time (considering we want to simulate hundreds of buildings), due to the Courant–Friedrichs–Lewy condition. Considering a typical mass flow in the serpentine of $0.1 \left[kg/s \right]$, a radius of the tubes of $2 \left[cm \right]$, and a discretization of 1 meter along the serpentine, the maximum allowable time-step is in the range of 2 seconds (considering implicit solution of the transport equation inside the tube). Since we do not simulate thermal activated building structures (TABS), in which the water of the heating system flows inside the building's concrete structure, but only underfloor heating pipes, which effects due to thermal inertia are less significant, we chose to neglect the thermal transient of the screed layer. Considering a fixed and uniform temperature for the ground and the building internal temperature at each time-step and stationary conditions, we can retrieve the analytical expression of the temperature profile along the pipe, through the energy balance on an infinitesimal element of the pipe. This can be expressed as:

$$\frac{\partial T_x}{\partial t} = \Phi_x - \Phi_{x+\partial x} + \dot{q}_{up} + \dot{q}_{down} \tag{2.6}$$

where x is the distance from the pipe entrance, T_x is the temperature of the water inside the pipe at x, Φ are enthalpy flows at the entrance and exit of the considered infinitesimal volume, $\dot{q}_{up}$ and $\dot{q}_{down}$ are the heating powers from the building and from the ground. Expressing the

latter through equivalent resistance taking into account convective and conductive effects, the balance in steady state can be rewritten as:

$$\frac{\dot{m}c_p}{\rho^*}\frac{\partial T_x}{\partial x} = \frac{R_{down}T_z + R_{up}T_g}{R_{down} + R_{up}} - T_x = T^a - T_x \tag{2.7}$$

where T^a is the asymptotic temperature and where:

$$R_{down} = \frac{1}{h_{in}w} + \frac{1}{h_{u,eq}w} + R_u \tag{2.8}$$

$$R_{up} = \frac{1}{h_{in}w} + R_g \tag{2.9}$$

$$\rho^* = \frac{R_{up} + R_{down}}{R_{up}R_{down}} \tag{2.10}$$

where w is the diameter of the tube, h_{in} is the internal coefficient of heat transfer, which can be retrieve using available empirical relation for fully developed flow with fixed temperature at the boundary conditions [76], $h_{u,eq}$ is the heat transfer coefficient between the floor and the building air including both the effect for natural convection and radiation. The values of $h_{u,eq}$ can be found in the literature [77],[78]. The value of the thermal resistances R_u and R_g, towards the floor and the ground, can be found in the literature as well. We can reformulate (2.7), making it adimensional through a change of variable:

$$\frac{\partial \Theta}{\partial \mathcal{X}} = -\Theta \tag{2.11}$$

from which solution we can retrieve the temperature profile of the water inside the pipe:

$$T_x = T^a + (T_0 - T^a)e^{\frac{-x\rho^*}{\dot{m}c_p}} \tag{2.12}$$

where T_0 is the temperature of the water at the pipe inlet. We can use (2.12) to retrieve the heating power flowing into the building, integrating $\dot{q}_{up}(x)$ along the pipe.

$$\dot{Q}_{up} = \int_0^L \dot{q}_{up}(x)\mathrm{d}x = \int_0^L \frac{T(x) - T_z}{R_{up}}\mathrm{d}x \tag{2.13}$$

where L is the length of the serpentine. Integrating, we obtain

$$\dot{Q}_{up} = \frac{(T^a - T_z)L - (T_L - T_0)\frac{\dot{m}c_p}{\rho^*}}{R_{up}} \tag{2.14}$$

where T_L is the temperature of the water at the outlet of the serpentine. Note that the equation (2.14) tends to $(T_L - T_0)\dot{m}c_p$ when R_{down} increase and R_{up} is kept fixed.
The nominal mass flow of the heating system and the length of the serpentine are found as the

solution of the following optimization problem:

$$\underset{L,\dot{m}}{\mathrm{argmin}} \left(\dot{Q}_{up}(L) - \dot{Q}_{nom}\right)^2 + 10^{-3} (\dot{m} - \dot{m}_{nom})^2 \tag{2.15}$$

where $\dot{m}_{nom}$ is a reference mass flow, equal to $0.1 \left[kg/s\right]$ and $\dot{Q}_{nom}$ is the power required to keep the building internal temperature constant under reference conditions (we used an external temperature of -4°C and a desired internal temperature of 20 °C):

$$\dot{Q}_{nom} = \frac{\Delta T_{ref}}{R} \tag{2.16}$$

where R is the resistance of an equivalent RC circuit describing the heating dynamics of the building.

Water tanks and boilers

The water tank connected with the floor heating, which is used as a buffer by the heat pump, and the boiler for the DHW, are modeled as a N-states fully-mixed stratified tanks. Despite not being able to model buoyancy driven effects such as heat plumes and transient de-stratification, this kind of models are suitable for 1D simulations and control [79].

The dynamic equation describing the evolution of the temperature of the tank's layers is the following:

$$C\frac{\partial T_i}{\partial t} = \dot{Q}^u_{buo,i} + \dot{Q}^d_{buo,i} + \dot{Q}_{h,i} + \dot{Q}_{loss,i} + \dot{Q}^u_{cond,i} + \dot{Q}^d_{cond,i} + c_p\dot{m}(T_{i-1} - T_i) \tag{2.17}$$

where T_i is the temperature of the i_{th} layer, $Q^u_{buo}, Q^d_{buo}, Q^u_{cond}, Q^u_{cond}$ are the thermal powers due to buoyancy and conduction, from the lower and upper layer, respectively. The last term represents the enthalpy flow due to mass exchange, while C is the thermal capacity of the layer, in $[J/K]$ and $Q_{h,i}$ is the thermal power due to an electric resistance (for the boiler) or an heat exchange (for the heating system buffer). The expression for the above thermal power are the following:

$$\dot{Q}^u_{buo,i} = k\max(T_{i+1} - T_i, 0)N, \quad 0 \quad for \quad i = N \tag{2.18}$$

$$\dot{Q}^d_{buo,i} = k\max(T_{i-1} - T_i, 0)N, \quad 0 \quad for \quad i = 1 \tag{2.19}$$

$$\dot{Q}^u_{cond,i} = u_{amb}(T_{i+1} - T_i), \quad 0 \quad for \quad i = N \tag{2.20}$$

$$\dot{Q}^d_{cond,i} = u_{amb}(T_{i-1} - T_i), \quad 0 \quad for \quad i = 1 \tag{2.21}$$

$$\dot{Q}_{loss,i} = u_{amb}(T_{ext} - T_i) \tag{2.22}$$

$$\dot{Q}_{h,i} = \dot{Q}_{tot}/n_h \quad if \quad i \in \mathscr{I} \tag{2.23}$$

$$\tag{2.24}$$

where N is the number of layers, u_{amb} is the equivalent thermal loss coefficient with the

ambient and $\mathscr{I}$ is the set of the n_h layers heated by the heat exchange (or electric resistance). The buoyancy model is the one proposed in the IDEAS library [80]. Detailed description of the parameters for the boiler model can be found in [81].

Heat pump model

The heat pump is modeled by means of interpolated tables, in which heating and electrical power are available as a function of the evaporator and the condenser temperatures. The tables were taken from the energy simulation software Polysun (Vela Solaris AG, Winterthur, Switzerland). When the heat pump produces heat for both the heating system and the domestic hot water, its control logic prioritizes the latter, meaning that the buffer is heated as long as the DHW tank temperature sensor reaches the upper bound of its hysteresis control.

PV model

The sun azimuth and elevation are calculated based on the current time and the altitude, longitude and latitude of the given location. For this task we have used the Sandia National Laboratories PV Collaborative Toolbox [82], which is based on the 1985 Grover Hughes' Engineering Astronomy course at Sandia National Laboratories. The direct normal irradiance (DNI) is then calculated by means of the empirical disc model [83] . The diffuse horizontal irradiance at time t, DHI_t is then calculated as:

$$DHI_t = GHI_t - \cos\left(\theta_{z,t}\right) DNI_t \tag{2.25}$$

where $\theta_{z,t}$ is the zenith angle of the sun at time t. DHI is then used to estimate the projection of the diffuse radiation on the given surface I_d, using the Hay and Davies' model [84]. The overall radiation on the given surface is then given by the sum of the diffuse, direct and ground-reflected radiation.

$$I_t = I_{b,t} + I_{d,t} + I_{g,t} \tag{2.26}$$

where I_g is the ground reflected component, calculated as:

$$I_{g,t} = \rho GHI_t \frac{(1 - \cos(\alpha))}{2} \tag{2.27}$$

where ρ is the albedo, which was fixed to a typical value of 0.2, and α is the tilt angle of the PV field. The direct irradiation on the oriented surface I_b is obtained from the DNI:

$$I_{b,t} = DNI_t \cos(AOI_t) \tag{2.28}$$

where AOI is the angle of incidence of direct light on the PV oriented surface. To calculate DNI and $I_{d,t}$ we used the PV Performance Modeling Toolbox by Sandia National Laboratories [82]. Since reflection losses can significantly increase at small AOI [85], we applied an AOI

correction, independent from the module technology [86]:

$$I_t^{AOI} = IAM_t I_{b,t} + 0.95(I_{d,t} + I_{g,t}) \tag{2.29}$$

where IAM is the incidence angle modifier. We use the following ASHRAE approximation [87]:

$$IAM_t = \max\left(1 - k_1 \left(\cot(\min(AOI_t, \pi/2)) - 1\right), 0\right) \tag{2.30}$$

and k_1 is 0.05. Finally, in order to obtain an estimation for the electrical power produced by a field with the ith orientation, we apply a correction taking into account the ambient temperature and the inverter and module efficiencies. The cell temperature is first estimated from the ambient temperature, then a linear correction is applied [88]:

$$T_{cell,t} = T_{a,t} + \phi I_t^{AOI} \tag{2.31}$$
$$I_t^{AOIT} = I_t^{AOI}\left[1 + \gamma(T_{cell,t} - T_{ref})\right] \tag{2.32}$$

$T_{ref} = 25°C$ a reference temperature, ϕ and γ two coefficients. In this study, ϕ and γ are not estimated and are set respectively to the values of 3.14e-2 $[Km^2/W]$ and -4.3e-3 $[1/K]$, which represent crystalline silicon framed PV modules. Finally, the estimation of power production are corrected for the module and inverter efficiencies, using the following equation:

$$P_t = \eta_t I_t^{AOIT} \tag{2.33}$$

where η_t is the combined module and inverter efficiency. In order to reduce the number of parameters, we modeled it as a function of the irradiance I_t^{AOIT} using the following equation:

$$\eta_t = k_2 + k_3 \ln(I_t^{AOIT}/I_{STC}) + k_4(\ln(I_t^{AOIT}/I_{STC}))^2 \tag{2.34}$$

where $I_{STC} = 1000W/m^2$ is the reference irradiance and k_2, k_3, k_4 are free parameters. By fitting equation 2.34 to typical inverter and polycrystalline module data, we obtained the following values: $k_2 = 0.942$, $k_3 = -5.02e-2$, $k_4 = -3.77e-2$.

2.2 Multi-step-ahead forecast for control with nonuniform step size

Systems with an internal buffer, such as batteries, boilers and the thermodynamics systems composed by an heatpump and a building, can increase their performance when shifting from a reactive control to a predictive one. Furthermore, when the dynamics of the controlled system can be modeled accurately, we can use the system model to generate prediction on the system state into the future, as it's done in Model Predictive Control (MPC). In any case, the predictive control requires a multi-step-ahead forecast of the disturbances when the typical rate of change of the system rate due to control actions and disturbances is meaningful effect we get from controlling the system, has a typical time scale which is smaller than the rate of change of the system's state (e.g. the state of charge of a battery) when subject to a typical

control action.

The idea behind the proposed strategy is that, we could change the size of the time-step when forecasting distant events into the future, since these should have a lower impact on the optimal control action we must apply to the system at the present time, with respect to the events occurring in the next timestep. A similar reasoning is used in reinforcement learning, where a discount factor is applied to the reward of future actions [89]. If this is the case, instead of predicting one day ahead with 1 minute timesteps, we could instead predict averaged values of the system's disturbance on wider timesteps during the end of the day.

The advantage of using a nonuniform step-size would be twofold. Firstly, it will reduce the computational time needed for both the training of a forecasting algorithm, and for the solution of the control problem. Even when considering an optimistic computational time for the optimization problem of $\mathcal{O}(n)$, where n is the number of steps, passing from an horizon of 1440 steps (one minute intervals) to one of 14 steps reduces the computational time by a factor of 100. Secondly, in general is more difficult to predict disturbances at several steps ahead into the future than predicting aggregates for the same time horizon. Aggregation of distant time-steps will reduce the variance of the signal to be predicted. In fact, the sample mean of a signal is the best predictor in the case of least squares loss functions, i.e. is the value minimizing its variance. The variance reduction of signals due to aggregation is exploited in hierarchical forecasting to try to increase the prediction accuracy of the forecasts in the bottom level of the hierarchy [90; 91]. In [92] the same reasoning is used with temporal hierarchies, to reconcile solar forecasts.

An alternative solution to the proposed approach would be to set up a stack of hierarchical receding horizon controller operating at different time scale, where the higher level controller sets up state constraints for the lower level controllers, in cascade. The concept of nested controllers for functional, temporal and spatial decomposition is not new [93], but in they are usually limited to up to 3 layers, mainly due to the rising in the overall controller complexity. For example in [94] a three layer temporal hierarchy is proposed for the management of wastewater networks. A similar approach using two temporal layers is presented in [95], where the master controller set constraints for the lower level MPC, operating with different objective functions. The same approach is followed in [96], where the lower level controllers are operated by PIDs. A related work in the context of dispatchability of distribution feeders is presented in [97], where a battery commits in advance to a day-ahead dispatch plan, and is then operated using a shrinking horizon MPC.

A similar approach for speeding up MPC is the so called move blocking. This technique consists in fixing the control action for a given amount of time steps. Formally, the control vector $u = [u_i]_{i \in \mathbb{N}_{[1,N]}} \in \mathbb{R}^{Nm}$ must satisfy:

$$u = (T \otimes I_m)\, u_{mv} \tag{2.35}$$

where $u_{mv} \in \mathbb{R}^{Mm}$, with $M < N$, is the actual decision vector, which is expanded by the product

with $T \in \mathbb{R}^{N \times M}$ and I_m, where $\otimes$ is the Kronecker product. In [98] a review of different move blocking strategies is presented, while in [99], a way for the optimal design of T is presented. It must be noted that these strategies reduce the overall number of decision variables, but does not decrease the resolution of the forecast needed to operate the controller.

2.2.1 Nonuniform stepsize MPC

In this section the loss of accuracy when approximating a typical MPC control problem with one having a lower number of variables and inputs is investigated. As seen in section 1.5, since coordination strategies deriving from decomposition of centralized problems using ADMM, DW decomposition and similar techniques result in the evaluation of proximal operators [100], I consider an objective function which includes a quadratic tracking. In particular, I restricted the analysis to the case in which an electrical battery is operated to minimize/maximize the energy cost/revenues of a single user, while trying to track a constant power profile equal to zero, which corresponds to the case of peak shaving. Consider the battery's control law to be the solution of the following optimization problem:

$$
\begin{aligned}
u^* = \underset{x, u \in \mathcal{U}}{\text{argmin}} \; & \alpha \|\hat{P} + Su\|_2^2 + \sum_{t=1}^{T} c(\hat{P} + Su) \\
s.t.: \; & x_{t+1} = A_d x_t + B_d u_t \quad \forall t \in \mathbb{R}_{[1:T]} \\
& x_t \in \mathcal{X} \quad \forall t \in \mathbb{R}_{[1:T]}, \quad \mathcal{X} = [x_{min}, x_{max}] \\
& u_t \in \mathcal{U} \quad \forall t \in \mathbb{R}_{[1:T]}, \quad \mathcal{U} = [u_{min}, u_{max}] \\
& x_0 = x_{start}
\end{aligned}
\tag{2.36}
$$

where x_t and $u_t \in \mathbb{R}^2$ are the state of charge of the battery and the vector of control actions (charging and discharging power) of the battery, $S \in \mathbb{R}^{T \times 2T}$ is a summation matrix, summing charging and discharging battery operations with appropriate signs, $\mathcal{X}$ and $\mathcal{U}$ are the constraints sets containing the battery operational limits, $\hat{P} \in \mathbb{R}^T$ is the forecasted uncontrolled power at the electrical main of the user, and $c(\cdot)$ is the energy cost function defined as:

$$
c(z_t) = \begin{cases} p_{b,t} z_t, & \text{if} \quad z_t \geq 0 \\ p_{s,t} z_t, & \text{otherwise} \end{cases}
\tag{2.37}
$$

where $p_{b,t}$ and $p_{s,t}$ are the buying and selling tariffs, respectively, at time t. A_d and B_d are the exactly discretized dynamic matrices, that consider self-discharge, and (asymmetric) charging and discharging efficiencies, which are derived from the continuous matrices using:

$$
\begin{aligned}
A_d &= e^{At} \\
B_d &= (A_d - I) B A^{-1}
\end{aligned}
\tag{2.38}
$$

where A and B are the continuous dynamic matrices and t is the discretization time-step.

I compare problem (2.36) with the one in which the forecasted power profile $\hat{P}$ is only available as average values, performed with an increasing number of aggregation steps. The rationale behind this, as previously explained, is that the controller shouldn't incur in a high performance loss considering only the mean value of the power profile at distant timesteps. In other words, that the first order moment of the disturbance for the last timesteps is the only relevant information for the controller, and knowing other statistical properties won't result in an increase of the controller's performance. This means that instead of having fixed dynamic matrices, we will have to consider a set of n_a matrices, where n_a is the number of aggregation steps:

$$
\begin{aligned}
A_{d,k} &= e^{A t n_k} \quad \forall k \in \mathbb{R}_{[1:n_a]} \\
B_{d,k} &= (A_{d,k} - I) B A^{-1} \quad \forall k \in \mathbb{R}_{[1:n_a]}
\end{aligned}
\tag{2.39}
$$

where n_k is the number of original steps used for the k_{th} aggregation. I have used logarithmically spaced bins for the reduction of the control horizon. Given an original timestep of 10 minutes, the series of aggregation steps n_k is found setting $n_1 = 1$ and solving for γ the following system of equations:

$$
\begin{cases}
n_{k+1} = \lfloor \gamma \, n_k \rfloor \quad \forall k \\
\sum_{k=1}^{14} n_k = 144
\end{cases}
\tag{2.40}
$$

where $\lfloor \cdot \rfloor$ is the floor operator. An example of the logarithmically spaced aggregation is shown in Fig. 2.2.

The resulting nonuniform stepsize optimization problem is the following:

$$
\begin{aligned}
u^* = \underset{x, u \in \mathcal{U}}{\mathrm{argmin}} &\sum_{k=1}^{n_a} w_k \left[\alpha \left(\hat{P}_k + S u_k \right)^2 + c(\hat{P}_k + S u_k) \right] \\
s.t.: \; &x_{k+1} = A_{d,k} x_k + B_{d,k} u_k \quad \forall k \in \mathbb{R}_{[1:n_a]} \\
&x_k \in \mathcal{X} \quad \forall k \in \mathbb{R}_{[1:n_a]}, \quad \mathcal{X} = [x_{min}, x_{max}] \\
&u_k \in \mathcal{U} \quad \forall k \in \mathbb{R}_{[1:n_a]}, \quad \mathcal{U} = [u_{min}, u_{max}] \\
&x_0 = x_{start}
\end{aligned}
\tag{2.41}
$$

where w_k is a coefficient which re-weights the importance of the objective function of the k_{th} step, and $S = [1, -1]$. Differently from the formulation used in [101], where no re-weighting was considered, we argue that the w_k should re-weight the terms of the objective function based on the number of original steps used for the k_{th} aggregation, that is $w_k = n_k / T$.

I stress out that it is not guaranteed that problem (2.41) is a good approximation for (2.36), due to the presence of a sum of squares. If the objective function was linear, it is easy to see that the process of averaging future steps using bins would result in a mathematically equivalent form of the objective function. In this case, the only difference between (2.41) and (2.36) will be in the formulation of the system dynamics. In the presence of a sum of squares, we can see
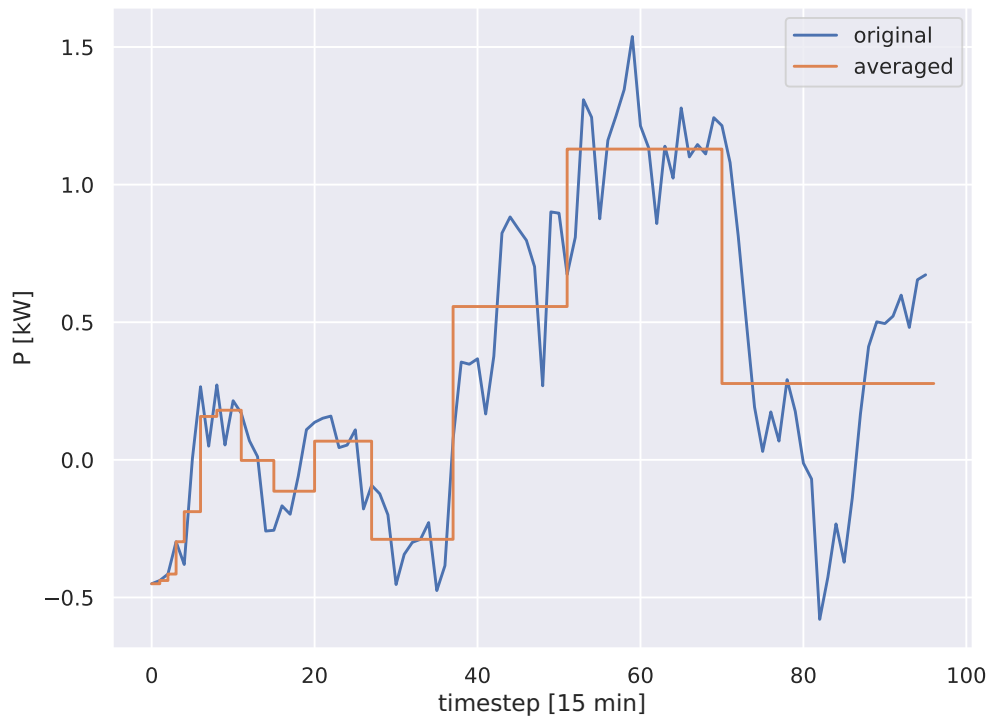
Figure 2.2 – Example of logarithmically spaced aggregation, using 14 steps, of a de-trended random walk. Blue: the original profile of 96 timesteps of 15 minutes each. Orange: logarithmic aggregations.

that the two objective functions would in general differ, recalling that, for any random variable x:

$$\mathbb{E}\left[x^2\right] = Var\left[x\right] + \left(\mathbb{E}\left[x\right]\right)^2 \tag{2.42}$$

This means that, since the variance is always positive, problem (2.41) will tend to underestimate the influence of future (averaged) steps in incrementing the objective function. For instance, the two formulations would be equal only in the case of zero variance of the power profile. Anyway, we are not directly interested in the difference between $\mathbb{E}\left[x^2\right]$ and $\left(\mathbb{E}\left[x\right]\right)^2$, but rather in the difference of controller performances when the aggregation bins are logarithmically spaced. This is evaluated by means of numerical tests in the following subsection.

2.2.2 Estimation of sub-optimality for the nonuniform stepsize MPC

Since estimating sub-optimality of the nonuniform stepsize MPC doesn't imply any consideration on the accuracy of the forecasts, in this section we just consider to be in possess of perfect forecasts of the disturbance. We compare two controllers, the first one using a fixed 10 minutes step for the description of the disturbance, and a second one using a reduced number of nonuniform logarithmically spaced steps. To perform this comparison, we used synthetically generated power profile, obtained as described in section 2.1, and tested the controller closed loop performances. In particular, we have generated 6 different annual profiles each of which composed by an uncontrolled power profile, an heat pump and a residential PV. For each power profile we randomly chose 6 sets of three consecutive days, in order to solve problem (2.41) on the first two days of each set. That is, the observations of the last day are only used as perfect forecasts for the last step of the second day, and the closed loop performance is just evaluated on the first two. We then solved the problem in a receding horizon fashion, where only the solution of the optimal control action of the first time step is actually applied. This means that for each profile we solved 12-days long receding horizon problems, which means that for each profile we solved problem (2.41) 864 times, retrieving an equal number of optimal control actions $u_{i,t}^*, t \in [1, 288]$. We systematically changed the two main parameters of the battery, which are the capacity $[kWh]$ and the E-rate $[kW/kWh]$. The capacity was changed between $[0.5, 2]\bar{E}_d$, where $\bar{E}_d$ is the average daily generated power exceeding the demand, that is:

$$\bar{E}_d = \mathbb{E}_D\left[\hat{P}_s^T\left(\hat{P}_s < 0\right)\right]/N_d \tag{2.43}$$

where N_d is the number of days for each profile, $\mathbb{E}_D$ is the expectation with respect to the randomly sampled profiles, and $\hat{P}_s \in \mathbb{R}^{144N_d}$ is the s_{th} random profile. This scaling was done to provide a realistic dimensioning of the battery for each scenario.

The KPI used is the ratio between the expectation over the dataset D_s of the closed loop evaluation of the objective function in (2.41) and the expectation of (2.36), when considering
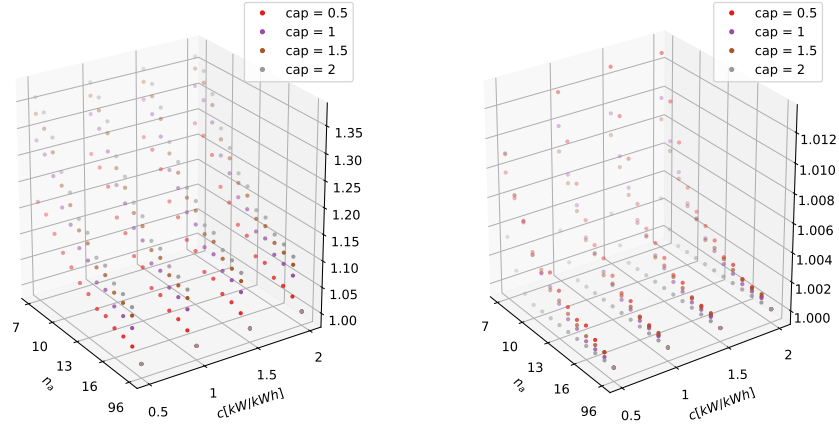
Figure 2.3 – Performance gap shown in terms of K^*, which is the ratio of the objective function value obtained by problem (2.41) and the non averaged formulation (problem 2.36), when using unitary w_k (left) or equal to n_k/T (right). The scores are plotted as a function of the battery capacity and E-rate, for an increasing number of aggregation steps n_a. For both the cases the increase of capacity is more relevant reducing the performance gap, with respect an increase in the E-rate.

the same combination of capacity and E-rate:

$$K^*_{i,j,n_a} = \frac{K_{c_i,cap_j,n_a}}{K_{c_i,cap_j}}$$

$$K = \mathbb{E}_{D_s}\left[\alpha\|P + Su^*\|_2^2 + \sum_{t=1}^{T} c(P + Su^*)\right]$$

(2.44)

where K_{c_i,cap_i,n_a} is the KPI obtained with the nonuniform formulation in (2.41), with n_a averaging steps, u^* are the optimized control actions, and P is the realization of the power profile, which in this case is equal to the one used in the optimization, since we are using perfect forecasts. In Fig. 2.3 K^*_{i,j,n_a} is plotted as a function of the capacity and the E-rate, when considering unitary unitary w_k (left box) or equal to n_k/T (right box). As we can see, for both the formulations, the normalized KPI K^* exponentially approaches the performance obtained by the uniform step-size formulation (2.36). It is clear that the parameter which most affects the KPI K^* is the capacity of the battery, while the c parameter effect is negligible. As we can see, the non re-weighted formulation incurs in a higher performance loss, with an increase in K, with respect to the uniform formulation, spanning from around 30 % to more than 5%, going from $n_a = 7$ to $n_a = 18$. The re-weighted formulation is far more stringent, ranging from a maximum of 1.2 % to under 0.1 % for the same range of n_a. The gain in performance is made more evident in Fig.2.4, where the K^*_{i,j,n_a} KPIs are collapsed in the same boxplot for all the values of capacity and E-rates.
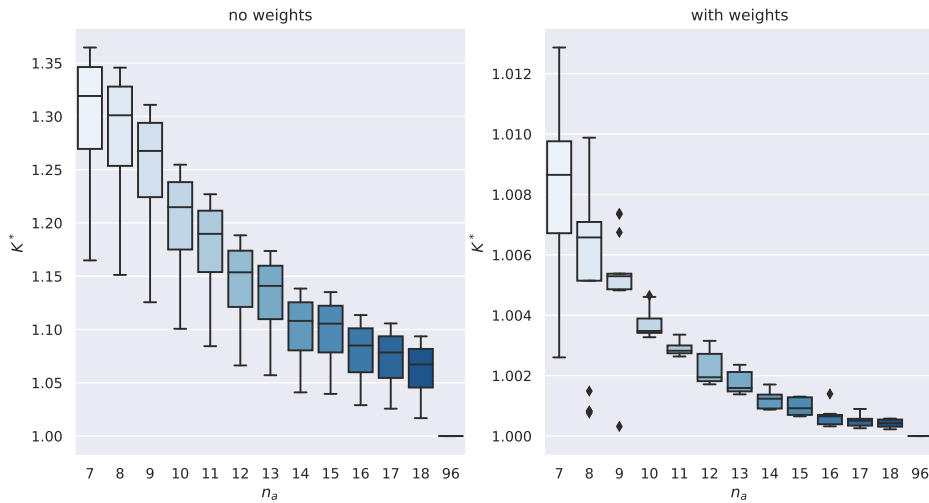
Figure 2.4 – Comparison of controller performances, when using unitary w_k or equal to n_k/T. Each boxplot contains 16 observations, which are the fold-averages of the KPI, with respect to the capacity and c combinations. The scores are plotted with increasing number of aggregation steps n_a. The score K^* is normalized with the KPI obtained by the non averaged formulation (problem 2.36).

Fig. 2.5 shows the relative increase in computational time when using the nonuniform step-size formulation, for increasing n_a. Once again, all the results from different values of capacity and E-rate are collapsed in the boxplots. We stress out that we are just interested in seeing the linear scaling of the computational time, while we are less interested in the absolute computational time, since the code is not pre-compiled and could be further optimized. Just to give an idea of the problem complexity, we report that the median computational time for solving (2.41) with $n_a = 7$ was of 0.028 seconds on a $Intel Core i7 - 5500U CPU@2.4GHz$.

2.3 Evaluation of multi step ahead forecasters for net power prediction

Demand side management algorithms use forecasts of the overall power consumed or generated at the point of common coupling with the grid in order to properly plan an optimal control signal. They typically need up to 24 hours ahead forecasts, since residential power profiles have a strong daily seasonality. Forecasting a signal several steps ahead in the future poses technical and conceptual challenges, depending on the adopted technique. Some simple methods, like state space methods, ARMAX and exponential smoothing, are usually used to perform multi-step ahead forecasting through a recursive technique, in which the 1 step ahead forecast is used as an input to the same model to predict the next step ahead [102]. Although this procedure is very simple, since it does not change the model at each time step, it can be prone to instabilities: since errors tend to accumulate and cannot be

Figure 2.5 – Boxplots of the computational time of the nonuniform stepsize formulation, for increasing number of aggregation steps n_a. The computational time is normalized with the CPU time for the formulation with $n_a = 7$. The median computational time for $n_a = 7$ was 0.028 seconds.

compensated, the recursive technique strongly depends on the quality of future exogenous inputs, that are usually forecasted using some other methods. Another possible way to apply a general purpose univariate interpolator to multivariate regression, is to perform some sort of embedding of the time step ahead one is trying to predict, and pass it as an additional input to the interpolator. The so called direct way is to fit T different interpolators where T is the number of steps ahead to be predicted, keeping the set of regressors fixed, while changing the target variable to be the t_{th} step ahead target. The latter could provide better accuracy when compared to multivariate regression, also known as multiple input multiple output (MIMO) regression, but it lacks of interdependency modeling between time-steps, which is needed in the case we are interested in producing scenarios for stochastic control out of the predictions [48]. Anyway, the interdependency structure can always be superimposed later, through time dependent copulas, for example with the dual-ensemble copula-coupling approach [58]. A comparison of several multi-step ahead strategies for forecasting, applied to very simple univariate time series of the NN5 competition [103], using non-probabilistic methods, is presented in [104].

2.3.1   Models and methodology

Forecasting multi step-ahead power profiles for residential households is an extremely complex task, mainly due to the high dimensionality of the multivariate output, the even higher dimensionality of the inputs, which make us incur in curse of dimensionality [105] related problems, the volatility of the target and its low signal to noise ratio. The last two points doom any point forecasting method to return large errors when compared to the task of forecasting more predictable time series (like the power flow at a medium voltage transformer). Furthermore, a decision maker or a control algorithm is, generally speaking, not interested in point forecasts, but rather in probabilistic ones, that is, in the a-posteriori probability distribution of the target given the present conditions. As such, we only focused on methods able to cope with those aspects, and evaluated them also based on the predicted quantile of the conditional distributions. For 3 of them we directly predict time aggregates of the target variable, as motivated in subsection 2.2.1, in the attempt of mitigating the difficulties of predicting an output with high dimensionality. For the recursive quantile random forest model, we firstly retrieve the whole set of 96 step predictions (24 hours ahead with a timestep of 15 minutes) and we perform a-posteriori.

We introduce the methodology used for the creation of the training and testing datasets D_{tr} and D_{te}. Each of these datasets contain a predictors (or regressors) matrix X, and a target matrix Y. Formally, given datasets with N observations, and a prediction horizon of T steps ahead, we obtain the Hankel matrix of targets $Y \in \mathbb{R}^{(N-T)\backslash T \times T}$, where \ stands for integer division, and the Hankel matrix of the regressors, $X \in \mathbb{R}^{(N-T)\backslash T \times n_x T}$, where n_x is the number of

regressors. Verbosely, X and Y are the following:

$$X = \begin{bmatrix} x_{1,t-T} & x_{1,t-T+1} & \dots & x_{1,t} & x_{2,t-T} & x_{2,t-T+1} & \dots & x_{n_x,t} \\ & & & \dots & & & & \\ x_{1,t-T+1} & x_{1,t-T+2} & \dots & x_{1,t+1} & x_{2,t-T+1} & x_{2,t-T+2} & \dots & x_{n_x,t+1} \\ x_{1,N-2T} & x_{1,N-2T+1} & \dots & x_{1,N-T} & x_{2,N-2T} & x_{2,N-2T+1} & \dots & x_{n_x,N-T} \end{bmatrix} \qquad (2.45)$$

$$Y = \begin{bmatrix} y_{t+1} & y_{t+2} & \dots & y_{1,t+T} \\ & & \dots & \\ y_{N-T+1} & y_{N-T+2} & \dots & y_N \end{bmatrix} \qquad (2.46)$$

where $x_{1,t}$ stands for the first explanatory variable at time t. In order to further reduce the dimensionality of the problem, for some of the forecasting methods we test, we aggregate columns of X and Y with logarithmically spaced bins. That is, referring to table 2.4, the first 31 (465 divided by 15) columns of matrix X, which are the farthest in time with respect to the one step ahead prediction, are averaged together, and so on for all the columns. This reduces the size of X from $(N-T) \times T n_x$ to $(N-T) \times 10 n_x$. This strategy assumes that, since we are requiring forecasts with a loss of temporal resolution as we move forward in time from t, we can as well get rid of the high temporal resolution of historical values, as we move backward from it.

Formally, we can define the reduction of X and Y through a block diagonal matrix M:

$$\begin{aligned} X_{red} &= \left[X_{red,i} \right]_{i \in \mathbb{N}_{[1:n_x]}} \\ X_{red,i} &= X_i M_x \\ M_x &= \bigoplus_{k \in \mathcal{K}} \mathbb{1}_{n_k} \quad \mathcal{K} = [K:1] \\ Y_{red} &= Y M_y \\ M_y &= \bigoplus_{k \in \mathcal{K}} \mathbb{1}_{n_k} \quad \mathcal{K} = [1:K] \end{aligned} \qquad (2.47)$$

where $\bigoplus$ is the direct sum operator, $X_{red,i}$ and Y_{red} are the i_{th} regressor's reduced matrix and the output reduced matrix, K is the number of aggregated steps, each of which includes n_k original steps (reported in table 2.2) and $\mathbb{1}_{n_k}$ is the column vector of ones, of length n_k. An example of the resulting aggregation for the past and future values of the target is shown in Fig. 2.6.

In the following we describe the four methods we employed for this analysis. Each method belongs to a different class of multi-step ahead forecasters introduced before.

Quantile regression forest and recursive quantile regression forest   Random forests are basically an ensemble of (high variance and low bias) decision trees, trained with random subsets (independently chosen) of the explanatory variables and samples of the original
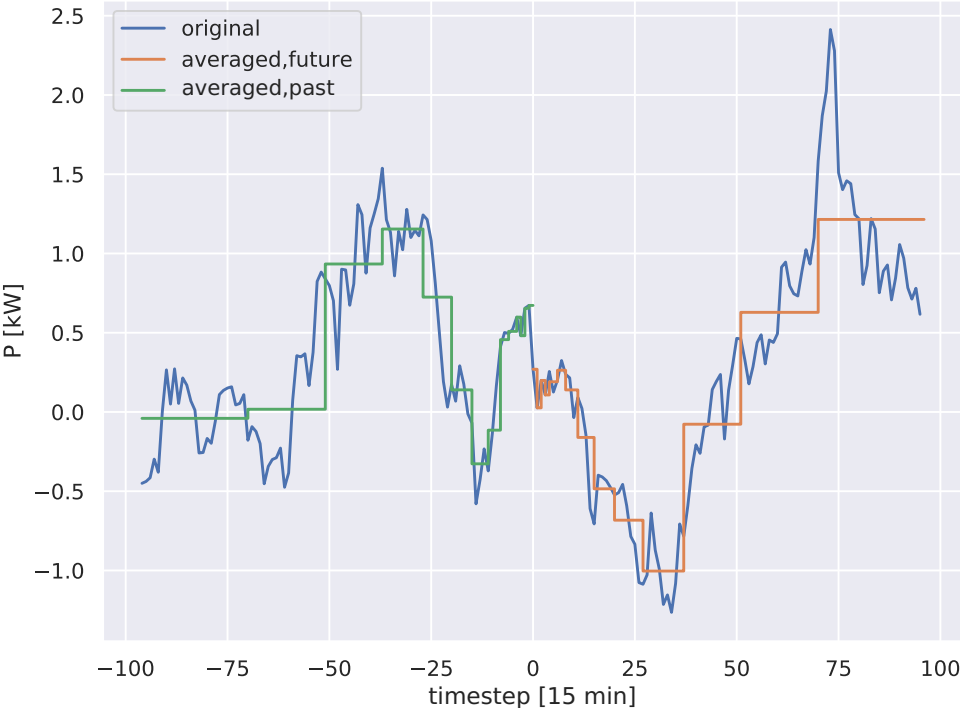
Figure 2.6 – Example of logarithmically spaced aggregation, for the bins reported in 2.2, for the regressor (green) and the target (orange).

training dataset. It is known that aggregating multiple forecasters leads to variance reduction of the prediction, when base learners (the to-be aggregated forecasters) are trained using random samples of the dataset. This technique is known as bagging, bootstrap aggregation [106], or weak learners aggregation. Each tree is trained following a greedy strategy, choosing which variable to split and at which level in a sequential way, minimizing a loss function (usually squared error). In order to get the quantiles out of a random forest, one could simply estimate the quantiles of the prediction of the N trees, assuming each tree has been independently optimized. Anyway, a better alternative exists, which consists of keeping track of all the data in the trees' leaves, and estimating empirical quantiles out of them. This algorithm is known as quantile regression forest (QRF) [107]. The first method based on QRF uses a direct MISO approach, which uses past variables up to time $t - T$ to predict aggregated output at step k. The input of the QRF is the logarithmically time averaged set of regressors X_{red}. We train K different random forests, each of which predicts a different (aggregated) step ahead. For the k_{th} step ahead forecaster:

$$\hat{Y}_{red,k} = f_k(X_{red}) \quad \forall k \in \mathbb{N}_{[1:T]} \tag{2.48}$$

where $\hat{Y}_{red,k}$ is the prediction for the k_{th} column of Y_{red}, that is, the forecast for the k_{th} step ahead. The final set of K forecast steps at time t, is then retrieved joining the output of all the f_k forecasters.

The second method still relies on QRF, but uses a recursive-like logic to predict the $t + k$ step ahead. The first step ahead is simply predicted taking the full Henkel matrix X as regressors, that is, the one described by equation (2.45) and with 15 minutes sampling time. The second forecaster, predicting $\hat{Y}_{red,2}$, will make use of a regressors matrix, $X_{rec,2}$, generated replacing the first column of X with the prediction of $f_k(X)$ for the first step ahead, $\hat{Y}_{red,1}$. The k_{th} forecaster will replace again n_k original steps from $X_{rec,k-1}$, where n_k is the number of original steps for the k_{th} aggregation, with its predicted output $\hat{Y}_{red,k}$. The procedure is iterated since $10th$ step ahead, for which all the columns referring to the historical values of y but the last 31, have been removed from the original X. The rationale behind this strategy is that the progressive reduction of the dataset and the inclusion of the prediction of the previous step ahead could increase the accuracy of the QRF.

Tree bagging Extreme learning machines   This method uses a MIMO regressor, the so called extreme learning machine (ELM) [108], to predict all the 10 step ahead at once. Since ELMs were originally thought as interpolators, we perform an ensemble through bagging in order to obtain the conditional distribution of the prediction. The ELM is a 2-layers perceptron, in which the first matrix of weights linking the inputs to the layers of neurons, and the neurons biases, are randomly initialized. As such, the only weights that need to be learned are the ones linking the neurons with the (multivariate) output. This can be done efficiently by means of linear regression exploiting the Moore-Penrose inverse. The main advantage of the ELM is that, since there is no need of training the first matrix of weights and biases, we can

bypass backpropagation (gradient descent) or other types of iterative optimizers, which are usually needed due to the presence of the nonlinear activation function, and just retrieve the optimal values of the second matrix of weights. In practice, this method firstly augment the space of explanatory variables by applying random nonlinear transformations, and retrieves their optimal combination by linear regression. Since the dimensionality of the explanatory variables is highly increased, we can penalize the weights adding a Ridge punishment in order to regularize the output and decrease the forecaster variance. Formally, the ELM is described by:

$$
\begin{aligned}
Y^T &= \theta O + \epsilon \\
O &= \sigma(W X^T + b)
\end{aligned}
\tag{2.49}
$$

where X and Y are the previously described Henkel matrices, σ is the activation function (which is usually a sigmoid), $W \in \mathbb{R}^{(N-T)\backslash T \times n_n}$ and $b \in \mathbb{R}^{n_n}$ are the randomly initialized matrix of weights and biases of the first neuron layer, where n_n is the number of neurons. $O \in \mathbb{R}^{n_n \times (N-T)\backslash T}$ is the output of the first layer and $\theta \in \mathbb{R}^{n_n \times T}$ is the matrix of final weights which connects the output of the first layer O to the matrix of targets Y. As in normal linear regression, assuming Gaussian noise ϵ and applying regularization, we retrieve the analytical solution for θ:

$$
\theta^* = \left(OO^T + \lambda I\right)^{\dagger} (OY)
\tag{2.50}
$$

where λ is the Ridge regularization parameter, I is the identity matrix of appropriate dimensions and $\dagger$ stands for the pseudoinverse. Due to the presence of the sigmoid activation function, inputs must be normalized to be centered in the interval $[-1, 1]$. We normalized X to unit variance; informally:

$$
X_n = \hat{\sigma}_s^{-1}(X - \hat{\mu}_s)
\tag{2.51}
$$

where $\hat{\sigma}_s$ is the vector containing the estimated sample standard deviation of the columns of X, and $\hat{\mu}_s$ is the vector of the estimated sample means of the columns of X. In order to speed up the activation function computation, we have used a piecewise linear function instead of the analytic sigmoid. Notice that in this case the nonlinearity reduces to an upper and lower threshold. For the tests we have adopted 400 neurons ELM, with λ equal to $1e-1$. The value of λ was chosen minimizing the quantile skill score, introduced later in section 2.3.2, on a set of residential profiles, and kept fixed. To estimate the conditional pdf and reduce the variance of the predictions, we have applied bagging to 100 ELMs. For each ELM, we selected 70% of variables and 70% of the original observations, at random.

Detrended multiple Holt-Winters   The Holt-Winters (HW) method [109] is a special class of the exponential smoothing [110], which consists of three smoothing equation, such that the final prediction is a combination of the level a, trend b and seasonality s. We tested different flavors of the HW families and based on performance we adopted a double seasonality additive

HW:

$$\hat{y}_{t+h} = (a_t + hb_t) + s_{1,t-p(1)+1+(h-1)\backslash p_1} + s_{2,t-p_2+1+(h-1)\backslash p_2}$$
$$a_t = \alpha(y_t - s_{1,t-p_1} - s_{2,t-p_2}) + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{2.52}$$
$$s_{1,t} = \gamma_1(y_t - a_t - s(2, t-p_2)) + (1-\gamma_1)s_{1,t-p_1}$$
$$s_{2,t} = \gamma_2(y_t - a_t - s(1, t-p_1)) + (1-\gamma_1)s_{2,t-p_2}$$

where α, β, γ_1 and γ_2 are parameters to be learned from data, while $p_1 = 96$ and $p_2 = 672$ are the periods of the seasonalities, and \ is the modulo operator. The data we used is sampled with a 15 minutes timesteps, so that p_1 and p_2 corresponds to a daily and weekly period. The model (2.52), and HW in general, do not include exogenous inputs. Since quantities like external temperature and irradiance are important explanatory variables in load forecasting, we included them with an a-priori linear detrend, such that the new target is $y = y - X\beta_d$, where X is a three column matrix containing GHI, T and the unit vector (for the intercept), and β_d is the vector of linear coefficients. Usually, a single set of α, β, γ_1 and γ_2 values is fitted, and the prediction of each step ahead is obtained applying equations 2.52 recursively, as usually done for state space systems. That is, the HW can be classified as a SISO forecaster, which uses recursion in order to generate the univariate predictions for the whole prediction horizon. In order to increase the accuracy of the method, we instead fitted 96 sets ,α, β and γ parameters, based on the step ahead. To identify them, we used synthetic generated power profiles, using the methodology described in section 2.1. Due to the linear detrend we applied to the target, the fitted β values where close to 0 for all the steps ahead, and thus we decided to fix this parameter to 0. The identified parameters are shown in Fig. 2.7. It can be seen that the variation of the parameters with the step ahead is very regular, and hence the three set of parameters could be replaced by 3 smoother interpolating function. The only steep variation of the parameters is located in the first steps ahead, where the α and γ_1 parameters experience a quick drop. The effect of the high values of these two parameters in the first step ahead on the HW prediction, is that of having a more persistent-like response. In fact, as can be seen from the second equation in 2.52, the α value directly weight the previous value of the time series y_t and the two seasonality values $s_{1,t-p_1}$ and $s_{2,t-p_2}$. For the HW the conditional probability distribution of the prediction was obtained a-priori, using data from the training set. That is, for each fold, we obtained the distribution of the error on the training set, based on the step ahead and the hour of the day:

$$\hat{q}_{\alpha_i,h,k} = q_{\alpha_i}(e_{h,k}) \tag{2.53}$$

where $e_{h,k}$ is the set of training errors related to the hour h and to the k_{th} step ahead and α_i is the level of the quantile.

Table 2.1 summarizes for clarity the differences between the tested forecasting methods.

Table 2.1 – Characterization of the used forecasting methods.

forecaster	regressors	target	class	pdf
QRF	X_{red}	Y_{red}	MISO	conditional
RQRF	$X_{rec,k}$	Y_{red}	MISO	conditional
Detrended HW	original signals	Y, averaged ex-post	MISO	bootstrap
Bagging of ELM	X_{red}	Y_{red}	MIMO	conditional
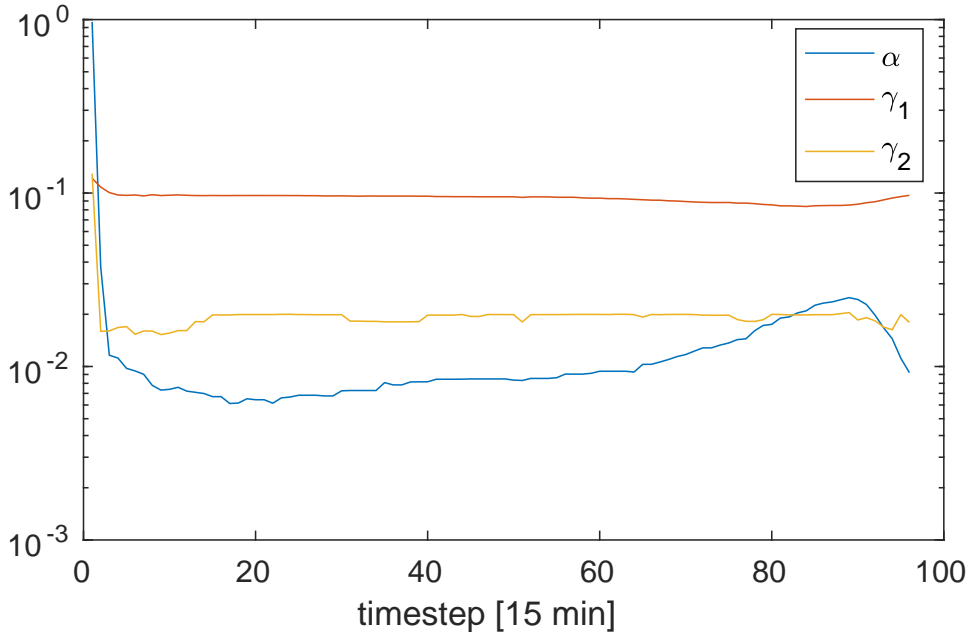
Figure 2.7 – Fitted parameters for the HW model, as a function of the step ahead

2.3.2 A priori evaluation

We compared the performance of the four different forecasters in predicting the power demand/production of a group of 100 prosumers, at different levels of aggregation, using one year synthetic data generated as described in 2.1. We used 15 minutes-sampled time series, but we did not predict 96 timesteps. Rather, we predicted 10 steps ahead with different levels of aggregations, using logarithmically spaced bins, as explained in section 2.3.1. The length of each bin is reported in Table 2.2. In table 2.3 the regressors for which the averaged historical values are passed to the forecasters, as well as the ones for which only the value at the current prediction time t is given.

Each user power consumption is obtained combining 3 different appliances: an heat pump (HP), a roof-mounted PV and uncontrolled loads (UN). The simulated prosumers are composed as described in Table 2.4. The forecasters are evaluated in cross validation, using 6 folds of 2 months each. We adopt the same cross validation approach later described in section

Table 2.2 – Number of minutes for each step ahead.

k	1	2	3	4	5	6	7	8	9	10
n_k	1	1	2	3	5	7	10	15	21	31
minutes	15	15	30	45	75	105	150	225	315	465

Table 2.3 – Regressors for which the averaged history is passed to the forecasters (first row), the ones for which the averaged future values is given (second row, NWP variables), and the ones for which only the value at the current time is given (third row)

24 steps-back averaged regressors	P, GHI, T
24 steps-forward averaged regressors	GHI_{nwp}, T_{nwp}
not averaged regressors	dayhour, day of week

3.2.1, in which the training set in each of the 6 folds is composed by groups of 3 consecutive days (see Fig. 3.2). As explanatory variables we use the historical values of the power, GHI and T. Additionally, we gave the regressor the perfect forecast of GHI and T. This was motivated firstly by the lack of NWP for the typical meteorological year we have used to generate the synthetic load, and secondly because we wanted to evaluate the best possible performance of the forecaster, without dealing with NWP accuracy. In fact, NWP services provide forecast with different accuracy, based on the presence of on-ground measurements used for prediction calibration.

Table 2.4 – Number of houses per type of appliances.

PV+HP+UN	PV+UN	HP+UN	UN
20	20	10	50

A-priori KPIs

The main two KPIs for the evaluation of forecasters performances are the root mean squared error (RMSE) and the mean absolute error (MAE):

$$RMSE_{sa} = \left[\frac{1}{n} \sum_{t=1}^{n} \left(y_{t,sa} - \hat{y}_{t,sa} \right)^2 \right]^{\frac{1}{2}} \tag{2.54}$$

$$MAE_{sa} = \frac{1}{n} \sum_{t=1}^{n} |y_{t,sa} - \hat{y}_{t,sa}| \tag{2.55}$$

where sa is the step ahead and n is the number of observations in the current test fold. These KPIs are the most used in the literature for performance evaluation. However, in some case is better to use a normalization factor to fairly compare the performance. For example, when evaluating PV predictions and forecasts, we have to keep in mind that the PV power signal is

highly variable, getting close to zero when the sun elevation is low, and zero during nighttime. For this reason before evaluation of the KPIs we have discarded all the values for which the sun elevation is bigger than zero: $\theta_{s,el} \leq 0$. Moreover, due to the periods in which the signal is close to zero, using the mean absolute percentage error is misleading, since dividing the prediction error by the true value would result in high (and highly non-linear) results. This is not only true for PV signals, but for signals with several zero crossings. For this reason, when evaluating PV forecasts we will also make use of the normalized variants of RMSE and MAE, as reported hereafter. In this case, the error is previously normalized using the mean of non-zero values for each step ahead, that is:

$$nRMSE_{sa} = \left[\frac{1}{n} \sum_{t=1}^{n} \left(\frac{y_{t,sa} - \hat{y}_{t,sa}}{\bar{y}_{sa}} \right)^2 \right]^{\frac{1}{2}}$$ (2.56)

$$nMAE_{sa} = \frac{1}{n} \sum_{t=1}^{n} \frac{|y_{t,sa} - \hat{y}_{t,sa}|}{\bar{y}_{sa}}$$ (2.57)

where $\bar{y}_{sa} = \frac{1}{n} \sum_{t=1}^{n} y_{sa,t}$ is the mean of non zero values of y for the current fold, at the sa step ahead. Note that this normalization does not requires to re-weight for the different length of the step ahead bins.

For probabilistic forecasts, different KPIs can be used to evaluate the predicted quantiles accuracy. Quantile skill score S which compares the quality of the predicted quantiles of the a-posteriori probability distribution returned by the forecasters, with the empirical one, estimated on the test set.

Formally, S is defined as:

$$S(\hat{y}_{t+k|t}, y_{t+k}) = -\sum_{i=1}^{m} \left(\mathbb{I}_{\{y_{t+k} \leq \hat{q}_{\alpha_i, t+k|t}\}} - \alpha_i \right) \left(y_{t+k} - \hat{q}_{\alpha_i, t+k|t} \right)$$ (2.58)

where $\hat{q}_{\alpha_i, t+k|t}$ is the predicted α_i level quantile for $t + k$ at time t and $\mathbb{I}_x$ is the indicator function on the condition x, and m is the number of evaluated empirical quantiles. The skill score (2.58) is a proper skill score [111], it is always positive, and lower values of S mean higher quantile prediction accuracy.

Numerical results

For this study, we have used $m = 10$, using linearly spaced quantiles on the $[0.05, 0.95]$ interval. Examples of the conditional pdfs returned by the forecasters are shown in Fig. 2.8. The first row refers to the first step ahead, while the second refers to the 10_{th} step ahead. The results for the individual profiles are shown in Fig. 2.9, by means of boxplots. Each boxplot contains the KPIs of 100 profiles, already mediated across the 6 folds. We evaluated the forecasters using the quantiles skill score 2.58 , RMSE 2.54 and MAE 2.55. Since all the boxplots show skew distributions for the KPIs, we additionally evaluated the nMAE 2.57, which results in
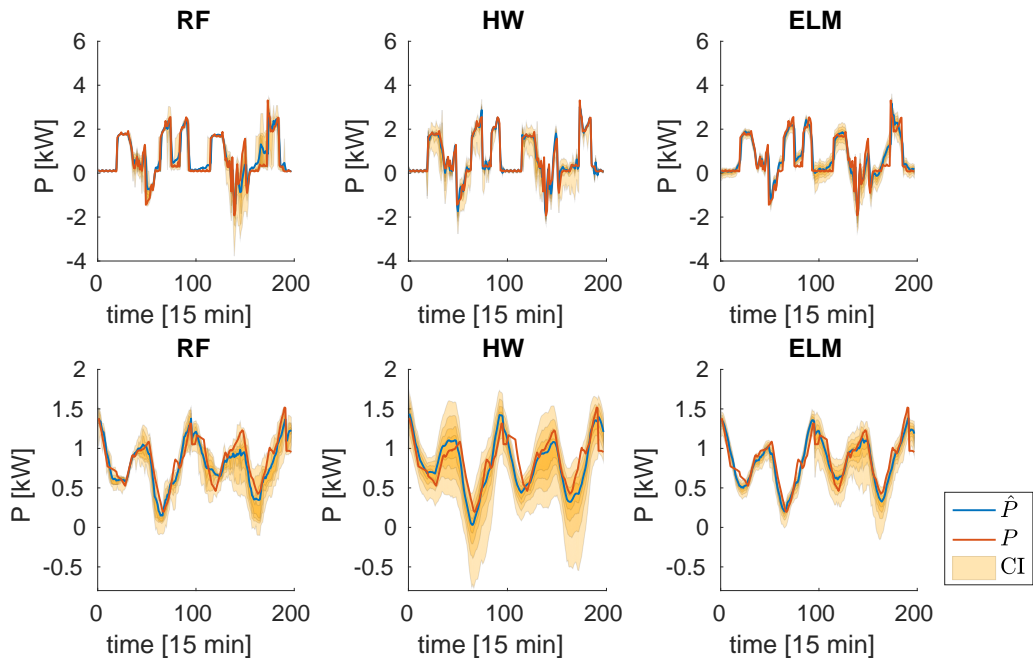
Figure 2.8 – Example of returned conditional pdfs, for the QRF, the HW and the bagging of ELM forecasters. First row: predictions for the 1_{st} step ahead, that is, the mean value of the next 15 minutes. Second row: predictions of the last step ahead, that is, mean values of 7.75 hours, going from 16.25 hours ahead up to 24 hours ahead.Confidence intervals are relative to 10 linearly spaced quantiles between $[0.05, 0.95]$
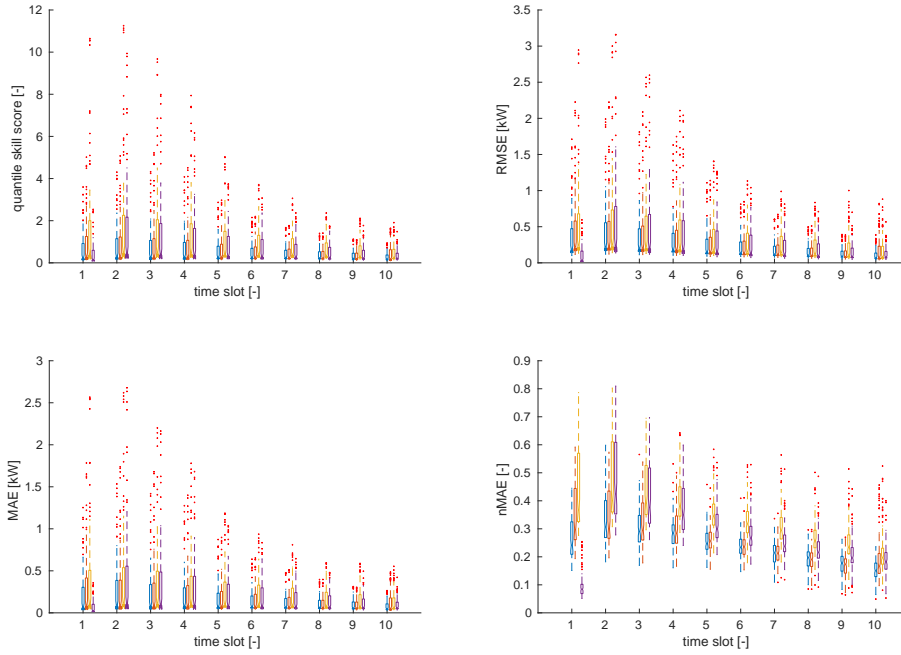
Figure 2.9 – Evaluation of different regressors for multiple step ahead forecasting, in terms of quantile skill score, RMSE,MAE and nMAE. Each boxplot contains 100 points, which are the results for each agent, mediated across the CV folds. Blue: QRF, direct. Red: RQRF, recursive. Yellow: bagging of ELM. Violet: detrended HW

not-skewed distribution. This indicates that the skewness was due to the original distribution of the mean absolute values of the power signals in the evaluated population. In any case, as expected, the renormalization did not change the rank of performance of the various forecasters. QRF and the RQRF clearly obtain better performance in all the indicators, with the exception of the one step ahead prediction, in which the HW is consistently the best forecaster. Besides the accuracy of the methods, we reports also the agent computational time for each forecasters, obtained on a Intel Core i7-4790K @ 4.00GHz with 32 GB of RAM. The values in the boxplots are, once again, already mediated across the folds, and refer to the training period on each fold, which was roughly 2 months. The median training time for each forecaster is reported in table 2.5. While the QRF and RQRF achieve greater accuracy in the forecasts, they are hardly embeddable in a smart meter due to both the high number of parameters and the high computational time. On the other hand, the ELM only requires matrix inversions in the training phase, and simple algebraic multiplication for the test phase. While the ELM requires $\mathcal{O}(n_n)$ parameters, where n_n is the number of neurons, the HW only requires 3×96 parameters, which can be reduced further if the parameters in figure 2.7 are approximated with splines or piecewise linear functions. More importantly, the HW does not require to keep historical values for the training dataset, being an adaptive state-space method. This means that all the information needed for performing a forecast are stored in the model parameters. This is

also true for the detrend of the exogenous variables when an adaptive linear fit is applied, as it was the case in this study. This means that the HW method can be easily embedded in a microcontroller.

Table 2.5 – Agent median computational training time for 2 months of data, for each forecaster.

forecaster	QRF	RQRF	ELM	HW
training time [s]	45.7	241.1	6.3	-

2.3.3   A posteriori evaluation

Beside evaluating the various a-priori through the KPIs introduced in section 2.3.2, we would like to evaluate them by means of performance of a controller planning for charging and discharging operations of an energy storage. As explained in section 2.2.1, we consider the case in which a residential user is operating its own battery in order to minimize its billing costs, while coordinating with some other batteries in order to minimize a common objective . When the coordination is done using decomposition techniques, this would results in the evaluation of a proximal operator [100], which in general leads to a linear-quadratic objective function. This is further explained in chapter 5 and in papers C and D. We used both a deterministic and a TBSMPC controller. For both of them, we used the logarithmically spaced stepsizes reported in table 2.2, as in section 2.2.1 we have estimated that the performance gap in this case is below 0.5 % for all the considered combinations of battery capacity and E-rate.

The TBSMPC controller was formulated using 4 sets of variables for each node in the disturbance tree (the state of the battery, the charging and discharging operations, and an additional variable y which performs as the upper bond for the cost function 2.37), leading to a total of $4n_{nodes}$ optimization variables, where n_{nodes} is the total number of nodes in the disturbance tree. This reduces the overall number of optimization variables with respect to the case in which scenarios are treated separately, as in [35], and non-anticipativity constraint (causality of the control actions) is enforced by means of equality constraints. In fact, in this case the overall number of variables would be $4n_sH$, where n_s is the number of scenarios and H is the length of the optimization horizon. Since the two formulations are mathematically equivalent,

we adopt the one with the lowest number of optimization variables.

$$
\begin{aligned}
u^* = \operatorname*{argmin}_{x,u\in\mathcal{U}} &\sum_{k=1}^{n_a} w_k \sum_{j=1}^{\mu_k} \pi_{k,j} \left[\alpha \left(\hat{P}_{k,j} + S u_{k,j} \right)^2 + y_{k,j} \right] \\
s.t.: \; &x_{k+1,d} = A_{d,k} x_{k,j} + B_{d,k} u_{k,j} \quad \forall k \in \mathbb{R}_{[1:n_a]}, \quad \forall j \in \mu_k, \quad \forall d \in \mathcal{D}_{k,j} \\
&x_i \in \mathcal{X}, \quad \forall i \in \mathcal{N}, \quad \mathcal{X} = [x_{min}, x_{max}] \\
&u_i \in \mathcal{U} \quad \forall i \in \mathcal{N}, \quad \mathcal{U} = [u_{min}, u_{max}] \\
&y_i \geq p_b \left(\hat{P}_i + S u_i \right) \quad \forall i \in \mathcal{N} \\
&y_i \geq p_s \left(\hat{P}_i + S u_i \right) \quad \forall i \in \mathcal{N} \\
&x_0 = x_{start}
\end{aligned}
\tag{2.59}
$$

where $\pi_{k,j}$ is the probability of reaching the j_{th} node of timestep k starting from the root node, μ_k is the number of nodes at timestep k, $\mathcal{D}_{k,j}$ is the set of descendant of the j_{th} node at timestep k, and $\mathcal{N} = \mathbb{N}_{[1:n_{nodes}]}$ is the set of all the nodes in the tree.

Since the RQRF was strictly worse in terms of a-priori KPIs, with respect to the QRF, while being more computationally demanding, we evaluated only the a-posteriori performance for the RF, the bagging of ELMs and the detrended HW. For each forecasters, we generated scenarios using one of the techniques described in section 1.4. Namely, we used multivariate Gaussian copula coupling for the QRF, estimating different multivariate copula based on the hour of the day. This means that we estimated 24 covariance matrices from the historical data of the power signal, $\Sigma_h \; \forall h \in \mathbb{N}_{[1:24]}$, which were then used to remap the empirical conditional cdf returned by the QRF. For the HW forecaster, bootstrapped historical data, always based on the hour of the day, was used. Finally, the individual weak learner (single ELMs) predictions was used for the bagging of ELMs. For each forecaster, we retrieved 200 scenarios, which were reduced into a disturbance tree using the technique and code presented in section 1.3.

The performance of the controller was evaluated in closed loop, using cross validation. We used the forecasts obtained in the previous section for one power profile composed by uncontrollable loads, an HP and a PV power plant. We systematically changed the capacity and E-rate of the battery, with an equally spaced sampling grid, as done in section 2.2. The values assumed by the E-rate were $\{0.5, 1, 1.5, 2\}$, while the values of the battery capacity were $\{1, 1.5, 2\}\bar{E}_d$, where $\bar{E}_d$ is the mean daily generated energy, as defined in (2.43). From the yearly profile, which was divided in 6 folds, as described earlier, we tested the controller on only 4 days belonging to the original forecast test set. This means that the control problem (2.59) was evaluated $6 \cdot 4 \cdot 96$ times during the performance assessment, for each capacity and E-rate combination.

A-posteriori KPIs

For both the deterministic and TBSMPC, we evaluated the performance by means of relative difference from the objective function obtained by a prescient controller K_{pre}^*. The prescient controller uses perfect forecasts, and thus provides us with the lower bound for the optimized objective function used in 2.41. The first KPI is the ratio between the expectation of the performance gap and the expected optimal objective function of the prescient controller.

$$
\begin{aligned}
K_{re}^* &= \frac{\mathbb{E}_{D_s}\left[K - K_{pre}\right]}{\mathbb{E}_{D_s}\left[K_{pre}\right]} \\
K &= \alpha\|P + Su^*\|_2^2 + \sum_{t=1}^{T} c(P + Su^*)
\end{aligned}
\tag{2.60}
$$

where K was evaluated for both the deterministic and stochastic controller.

Additionally, we also computed the expectation of the same ratio:

$$
K_{er}^* = \mathbb{E}_{D_s}\left[\frac{K - K_{pre}}{K_{pre}}\right]
\tag{2.61}
$$

This second KPI is more unstable than the first, since folds for which the optimal objective function value retrieved by the prescient controller is close to zero (which are the folds in which a higher rate of self consumption is attainable) make this estimator more pessimistic than (2.60).

In order to clearly see the increase of performance due to the stochastic formulation, we defined the additional KPIs $K*_{diff,re}$ and $K*_{diff,er}$:

$$
\begin{aligned}
K_{diff,re}^* &= 1 - \frac{\mathbb{E}_{D_s}\left[K_{stoc} - K_{pre}\right]}{\mathbb{E}_{D_s}\left[K_{det} - K_{pre}\right]} \\
K_{diff,er}^* &= 1 - \mathbb{E}_{D_s}\left[\frac{K_{stoc} - K_{pre}}{K_{det} - K_{pre}}\right]
\end{aligned}
\tag{2.62}
$$

which correspond to the amount of performance gap reduction made possible by the TBSMPC formulation, with respect to the deterministic case. Once again, we both investigated the ratio of the expectation and the expected ratio.

Numerical results

The TBSMPC was evaluated using four different scenario tree structures τ^{μ_1,μ_H}, where μ_1 is the number of scenarios in the first step ahead, and μ_H is the number of scenarios in the last step ahead. The steps ahead in between the number of scenarios were linearly scaled between μ_1 and μ_H. Based on this notation, the four different evaluated trees were $\tau^{1,20}, \tau^{1,40}, \tau^{10,20}$ and $\tau^{10,40}$. In Fig. 2.10 K_{re}^* and K_{er}^* are plotted for all the combinations of battery's capacity and E-

rate, for both the deterministic and stochastic formulations (for the case of $\tau^{1,40}$), represented by dots and crosses, respectively.

From the left panel, in which K_{re}^* is reported, we can see that for both deterministic and TBSMPC formulations, the bagging of ELM gives the worst results, for all the parameters' combinations. Both QRF and the detrended HW provide significantly better results, showing a performance gap with respect to the prescient controller between 10 and 15 % for most of the parameters' combinations. The deterministic HW forecaster provides performance which are very close, and sometimes better than, the QRF forecaster. In particular we can see that for the deterministic formulation, the ranking of performance between the QRF and the HW changes, passing from a normalized capacity of 1 to 2, for any E-rate. For a normalized battery capacity of 1, the QRF is always better than the HW, while for a normalized battery capacity of 2, the opposite is true.

On the other hand, the higher accuracy of the QRF in modeling the conditional pdf can be appreciated when solving the problem using TBSMPC. In this case the QRF solutions are always better than the ones provided by the HW forecasters, in terms of expected performance gap from the prescient solution. For all the parameters' combinations, the K_{re}^* for QRF is below 10 %.

The right panel of Fig. 2.10 shows the K_{er}^* KPI. From this panel we can see that for the HW, passing from the deterministic to the stochastic formulation increase this KPI, while on the contrary, both the ELM and the QRF show a decrease. This could mean that the HW estimated pdf is not good enough for periods in which the power signal is close to zero. In Fig. 2.11 and Fig. 2.12, the KPIs $K_{diff,re}^*$ and $K_{diff,er}^*$ are shown, respectively. From the first figure, we can see how the expected reduction of the performance gap with the prescient solutions, is significantly higher for the QRF formulation, for which, in the case of a normalized battery capacity greater than 1.5, $K_{diff,re}^*$ is always higher than 19%, for all the four tree configurations. The performance gap reduction is less significant for the HW, for which is around 10% in the case of $\tau^{10,40}$, while being even less significant for the ELM forecaster. From Fig. 2.12 we can draw the same conclusions reported for Fig. 2.10, for the HW forecasters, which in some cases shows a reduction of $K_{diff,er}^*$.

In Fig. 2.13, the boxplots of the computational times are reported for the deterministic and the four stochastic formulation. The increase of computational time is consistent. We must notice that the TBSMPC formulation we used adopted a non fixed tree structure between the solving horizon. This means that the structure of τ is not fixed, although having a fixed number of nodes per step ahead μ_k. Fixing the structure, for example specifying the number of splits per node at each step ahead, can readily halve the computational time, since the optimization problem does not need to be re-formulated after each run.

In order to systematically investigate the effect of the number of scenarios in the tree structures, μ_H, we used 20 annual synthetically generated profiles, in combination with the QRF forecasts. Of these, a first set includes 10 residential profiles, which are composed by a PV power plant,
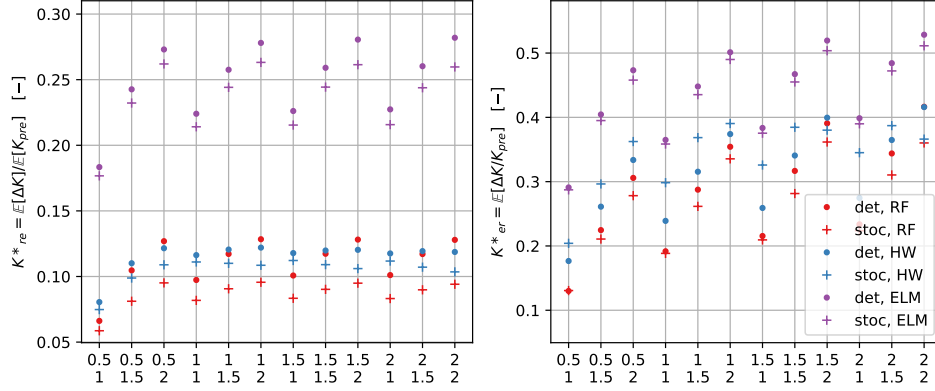
Figure 2.10 – KPI K^*_{re} (left), and K^*_{er} (right), for different forecasters, as a function of combinations of normalized battery capacity and E-rate. On the x axis, the first row refers to the E-rate, while the second one to the battery capacity normalized with E^*_{nom}. Dots refer to the solutions of the deterministic solver, while crosses to the solution of the TBSMPC controller.

a HP and uncontrolled loads. In the second set, the profiles are similarly composed, but the HP is not present. We linearly increased the number of scenarios from $\mu_H = 20$ to $\mu_H = 80$, using seven steps. The tested disturbance trees are $\tau^{1,20}, \tau^{1,30}, ... \tau^{1,80}$. For this test we kept the capacity and E-rate fixed to E^*_{nom} and 1 respectively. In Fig. 2.15

In Fig. 2.15, the results of the cross validation are shown. The left panel reports the KPI $K^*_{diff,re}$, as a function of the number of scenarios μ_H. Each line represents a different power profile. The red ones are the ones including a HP, while the blue ones belong to the second set. We can see how the reduction in the performance gap for the red profiles is in general higher with respect to the second group of profiles. This can be explained in terms of volatility of the forecasted power profiles in the two cases. In Fig. 2.14, an example of one profile from each set is shown for the first test fold of 4 days. We can see how the presence of the HP increases the variance (and decreases forecastability) of the profiles from the first group. It is thus reasonable that the TBSMPC provides higher benefits for this set of profiles. We can see that for a few cases, the benefit of considering more scenarios is not monotone, but start being detrimental after a certain number. This phenomenon has been observed also in [45; 35]. The right panel reports boxplots of the computational time per single solved horizon. Each boxplots contains 16 points, which represent the mean computational time of 16 single cases, mediated across all the solved horizons. These computational times were obtained on a IntelCore 8 x i7-4790K CPU @4.0GHz, while running 8 solvers in parallel (for this reason the computational times of the last 4 cases were not included in the boxplots). We stress out again that the code was not optimized nor ready to run on an embedded system, and we just want to report how the computational time scales with respect to the number of scenarios.

Figure 2.11 – KPI $K_{diff,re}^*$, for different forecasters, as a function of combinations of normalized battery capacity and E-rate, and different numbers of scenarios. On the x axis, the first row refers to the E-rate, while the second one to the battery capacity normalized with E_{nom}^*.
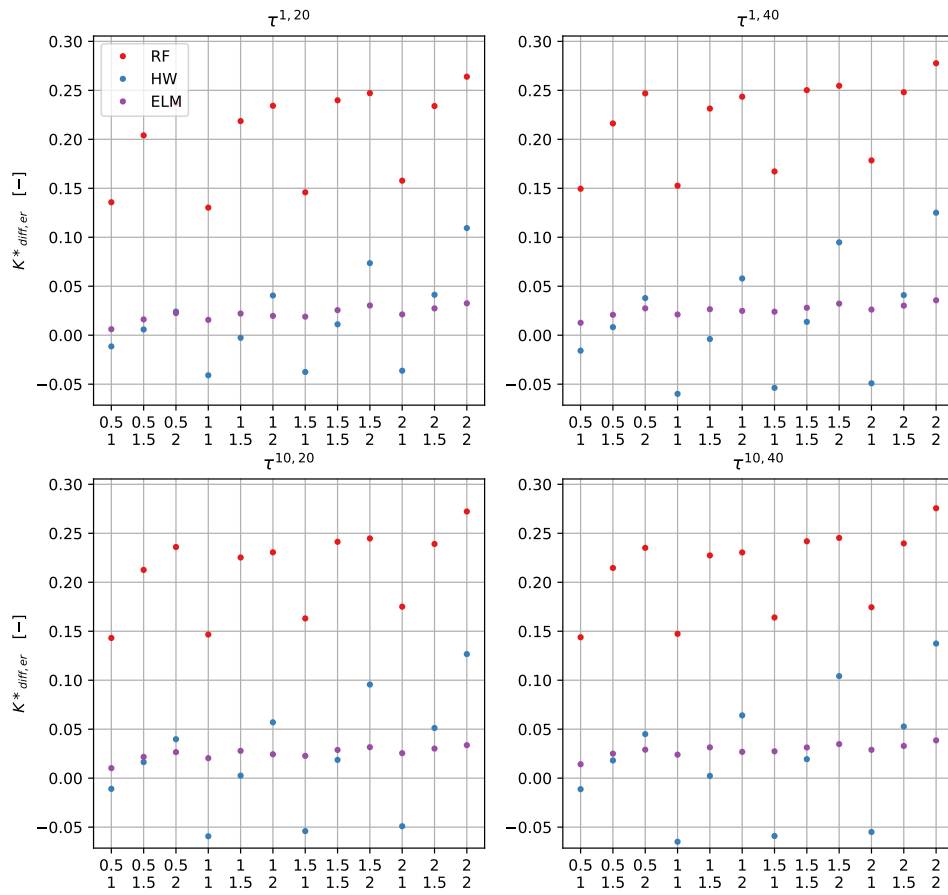
Figure 2.12 – KPI $K^*_{diff,er}$, for different forecasters, as a function of combinations of normalized battery capacity and E-rate, and different numbers of scenarios. On the x axis, the first row refers to the E-rate, while the second one to the battery capacity normalized with E^*_{nom}.

Figure 2.13 – Boxplot distributions of CPU times for the deterministic and stochastic formulation using different tree structures (first row: μ_1, last row μ_{10}).
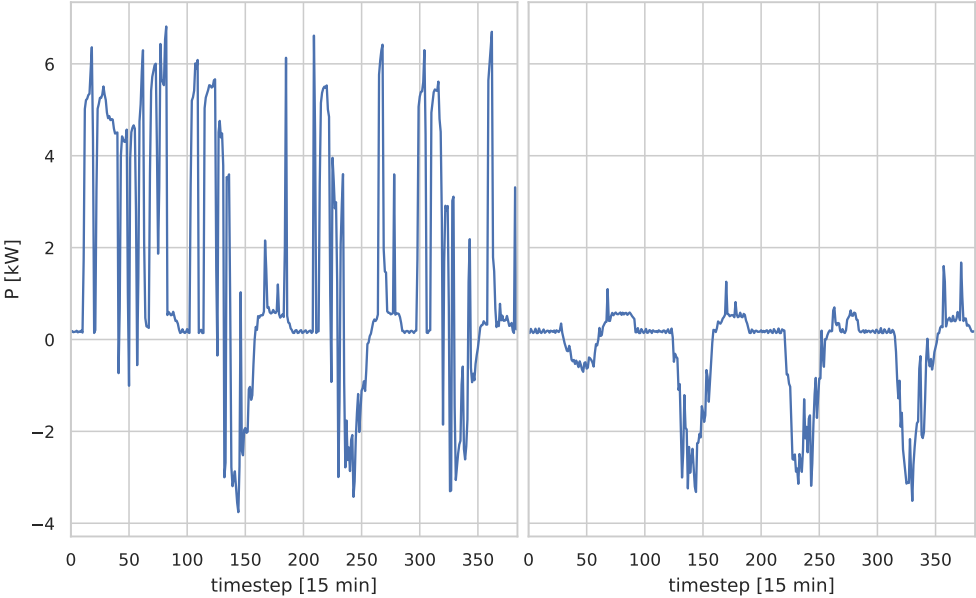
Figure 2.14 – Example of power profiles from the two sets, during winter. On the left, a power profile from the first set, composed by a PV, an HP and uncontrollable loads. On the right, a profile from the second set, composed only by a PV and uncontrollable loads.
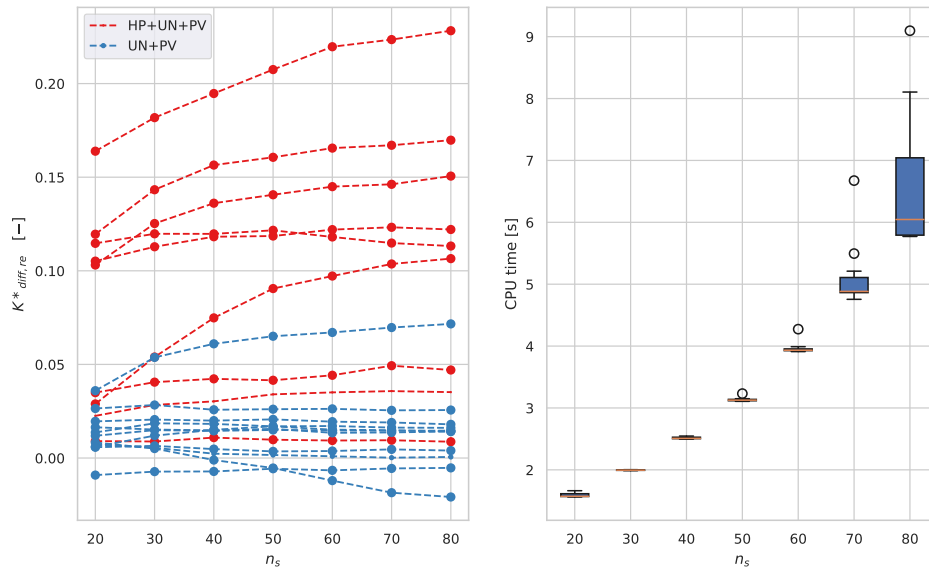
Figure 2.15 – Left: $K^*_{diff,re}$ for increasing number of scenarios. Each line represents a different power profile. The red ones include an HP, while the blue ones are only composed by a PV plant and uncontrolled loads. Right: computational time boxplots for increasing number of scenarios, for the solution of a single optimization horizon.

3 PV modeling for power forecasting

In this chapter the influence of using a physics-based model for predicting and forecasting AC PV power output, based on meteorological conditions, is investigated. In this chapter, new methods to identify a PV model without irradiance measurements are proposed. Moreover, we present new algorithms which are able to identify the presence, and estimate a PV physical model, using only composite AC power flow measurements. Additionally, the PV physical models can be exploited to estimate the global horizontal irradiance (GHI) with higher accuracy with respect to satellite-based services. This would allow to use already installed PV power plants as a distributed sensor network for retrieving GHI measurements. Finally, we evaluate the influence of the identified models on the accuracy of the forecasted PV power production.

The main outcomes of this chapter are summarized in the following:

- A methodology to blindly identify a physical model of PV power plants, starting from composite power signals, is introduced in sec 3.1. The different methods are explained in detail in paper A.

- A new unsupervised method for estimating the GHI from AC photovoltaic power measurements is introduced. The detailed procedure is presented in paper B, and its improved accuracy with respect to satellite-based irradiance estimations is reported, for two case studies.

- It is shown how, when combining physical models and QRF, the accuracy of predicting PV output from meteorological conditions increases significantly. Moreover, blindly identifying the PV model starting from composite power measurements does not significantly decrease the prediction accuracy.

- It is shown how modeling PV does help to increase the forecast accuracy, only for steps ahead between 30 minutes and 12 hours, which corresponds to the period in which NWP are more reliable.

3.1 Blind PV model identification and GHI unsupervised estimation

3.1.1 Problem formulation

The physics of PV power production, as well as the physics of solar geometry is well known. Accurate empirical models, which assess the total incoming irradiation on an oriented surface, given the GHI are also available [112],[113]. Under the assumption that the AC power produced by the PV power plant is not curtailed, we can therefore build a function that links the GHI and T to a given PV plant power output:

$$\widehat{P}_{pv} = g(GHI, T, t, geo, \Theta)$$
$$\Theta = [\theta_t, \theta_a, P_{nom}]$$

(3.1)

where $\widehat{P}_{pv} \in \mathbb{R}^T$ is the estimation of the power generated by a given PV plant, $GHI \in \mathbb{R}^T$ and $T \in \mathbb{R}^T$ are the vectors of the observed GHI and temperatures at time t, θ_t and θ_a are the vectors containing the tilts and azimuths of the modules, geo is a vector containing the geographical coordinates of the plant, namely latitude, longitude and elevation, and P_{nom} is the vector of the nominal powers of the modules. Usually, the physical model is estimated by minimizing the distance from the observer PV AC power signal and the one predicted by g and knowing GHI, solving the following nonlinear optimization problem:

$$\Theta^\star = \underset{\Theta}{\mathrm{argmin}} \, \|P_{pv} - \widehat{P}_{pv}\left(GHI, T, t, geo, \Theta\right)\|_2$$

(3.2)

We are interested in retrieving Θ, both in the case in which GHI and T local measurements are not available, and in the case in which only the overall power measurement of the building to which the PV power plant is connected.

3.1.2 On how to identify PV models without GHI measurements, and use PV panels as irradiance sensors

In order to find a good functional form for g, the following consideration must be taken into account:

1. We want to estimate PV plants orientation without knowing the actual GHI seen by the panels

2. PV plants can be composed by groups of panels with different orientations, e.g. plants with an east-west configuration

3. The presence of shadows affects the relation between the projection of GHI on an oriented surface and the PV power output

4. Problem (3.2) is non convex, due to the nonlinear transformations required to map GHI

to the plane of array.

If the GHI seen by the panels is unknown, the estimation of the panel's orientation would result in a blind identification problem. We exploit the fact that we can retrieve a good approximation of GHI for clear-sky periods, using a model for the extra terrestrial irradiation and for the air mass index, e.g. the McClear clear sky model [114]. We can thus identify the PV plants orientations if we could select clear-sky periods using only the PV plant's power output.

In order to select clear sky periods for the identification, we propose the following method. The distributions of the PV power signals as a function of the sun position typically presents a bimodal behavior, induced by the presence of clouds during data acquisition. On the other hand, a unimodal distribution could indicate a systematic shadow for the corresponding sun position. To select clear data periods, we fit a Gaussian mixture pdf with two components $X \sim \mathcal{N}(\mu, \sigma)$, with $\mu \in \mathbb{N}^2, \sigma \in \mathbb{N}^2$, for each sun position, with a discretization of $5°$. Then, for each sun position we identify the observations lying in the one sigma interval around zero of the Gaussian distribution with the largest μ as clear observations.

However, it is not always guaranteed that during the selected periods, the GHI seen by the PV plant would be exactly the one predicted by the clear sky model. To overcome this and the other aforementioned problems, instead of solving problem (3.2) and directly estimate Θ, we recast it as a robust linear regression:

$$\omega^* = \underset{\omega \in \mathbb{R}_+^N}{\operatorname{argmin}} \mathcal{L}\left(P_{pv} - I_N \omega\right) \tag{3.3}$$

where $\mathcal{L}$ is a robust loss function, $I_N \in \mathbb{R}^{T \times N}$ is a matrix whose columns contain estimations of the power produced by N differently oriented virtual PV panels with unitary nominal power, $\omega \in \mathbb{R}^N$ is the vector storing the importance of the virtual PV panels in predicting P_{pv}. The I_N matrix is obtained starting from GHI and T as described in 2.1, through equation 2.33. An example of signals contained in the I_N matrix is shown in Fig. 3.1, for $N = 21$, starting from clear sky GHI obtained using the McClear clear sky model. The additional requirement $\omega \in \mathbb{R}_+^N$ forces vector ω^* to have all positive entries. Thus, we can interpret the vector ω^* as coefficients describing the significance of each virtual PV panel in explaining the power output of the PV plant, rescaled for its nominal power.

Note that this strategy is very similar the method adopted by the ELM forecaster presented in section 2.3: we have turned a nonlinear problem into a linear one, avoiding solving a non-convex problem, providing an a-priori set of explanatory variables. This is the same for ELM, in which a-priori nonlinear regressors are obtained from random sigmoidal transformations, avoiding fitting the model using back-propagation. The difference is that in this case, physical information from the PV model is encoded in the regressors I_N. Furthermore problem 3.3 forces ω to be sparse and it is robust with respect to the presence of partial shadows.
If we use a loss function of an M-estimator for $\mathcal{L}$, we can solve problem 3.3 using an efficient iterative reweighted least squares algorithm. The presence of constraints $\omega \in \mathbb{R}_+^N$ calls for a
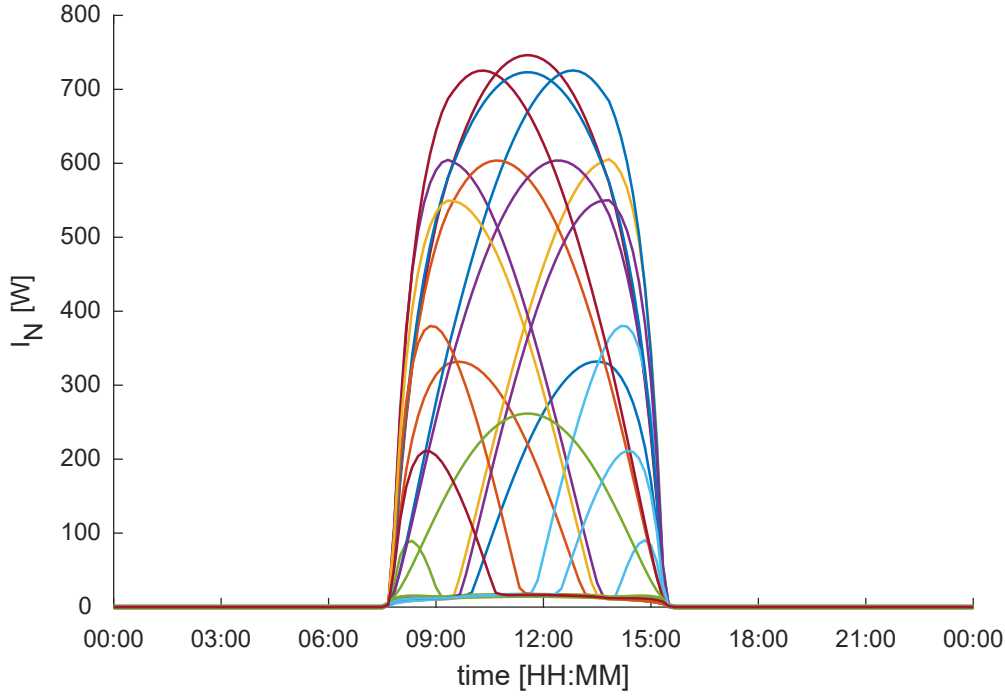
Figure 3.1 – AC power produced by 21 differently oriented virtual PV panels, during the first of January, for the location of Biel-Benken, CH. The virtual PV panels' orientations are obtained by generating a triangular mesh of an icosahedron on a unit sphere.

modification of the leverage weighting function, that can be achieved using a constrained linear regression, as explained in paper B.

Using the presented method, we can estimate ω without knowing the actual GHI seen by the panels, selecting only clear sky periods. Once ω has been estimated, we can finally use it to reconstruct the GHI signal seen by the PV panels. This can be done solving the following optimization problem:

$$GHI^* = \underset{GHI}{\operatorname{argmin}} \|P_{pv} - I_N(GHI)\omega\|_2 \qquad (3.4)$$

Problem (3.4) can be solved iteratively as explained in paper B. This means that we can retrieve the GHI that better explains the AC power produced by the identified model, i.e. the identified ω. We could use more than one PV signal to reconstruct GHI. In this case we need to identify a set of ω coefficients for each signal, and a merit function taking into account the influence of shading, as explained in paper B.

3.1.3 Identification of PV models from composite power flow measurements

As opposed to the case of industrial PV power plants, direct measurements of AC power are not always available for residential PV installations. In this case is more common to have only the aggregated measurement at the electric main of the building. As it has been reported in the literature in the case of wind forecasting [115], performing separated forecasts of load and production signals outperforms the accuracy of directly forecasting the net load. It has been recently shown that the same conclusion holds for the case of behind-the-meter PV [116]. In paper A, a way for identifying the physical model for the PV through unsupervised learning techniques is presented. The methods are then used to disaggregate the composite power flow measurements, retrieving the signal due to PV generation. Here we briefly explain the main idea behind the four presented methods. Let $P_m \in \mathbb{R}^T$ be the composite power measurement at the electric main, that is, the point of common coupling with the public grid (PCC), of a residential building. In the presence of a PV power plant connected at the PCC, we can describe P_m as the sum of of two components, generated and consumed power:

$$P_m = P_l + I_N \omega \tag{3.5}$$

where $P_l \in \mathbb{R}^T$ represents the power consumption, and the power production has been described in terms of the matrix I_N and unknown vector ω, as introduced in the previous section. All the proposed methods try to estimate ω assuming that the PV power signal is uncorrelated to, and independent form, the load P_l, at least for specific sampling time or in some range of frequencies. To better clarify this idea, we present one of the proposed method.

$$\omega^*, P_l^* = \underset{\omega, P_l}{\mathrm{argmin}} \, \|P_m - P_l - I_N \omega\|_2^2 + \lambda \|D P_l\|_1 \tag{3.6}$$

where $D = \mathbb{I}_T \otimes [-1, 1]$ is the discrete differentiation matrix, $\mathbb{I}_T$ is the identity matrix of length T and $\otimes$ is the Kronecker product. This method jointly estimates P_l and ω, assuming that the load is a piecewise constant signal.

The first part of the objective function is a least squares minimization between the estimated aggregated power and the observations at the PCC, P_m. However, this problem is under-determined since the free variables are more than the number of observations T. Therefore, we additionally punish temporal variations of the load. This results in a combination of linear regression and trend filtering problem, as for example in [117] [118]. The trend filtering problem can be seen as a fused lasso regression with a zero penalty term [119].

This method is used in the next section to blindly identify a PV model starting from P_m and a GHI measurement. This model is then compared, by means of prediction accuracy, to a PV model identified with direct observations of P_{pv}

3.2 Influence of PV modeling on forecasting

Several studies have tried to include physics based model in the forecasting task [120], while a plethora of methods relying on econometrics and machine learning have been proposed to forecast PV using a data-driven approach [121], [122]. However in these studies the authors tend to focus on one of the two approaches, and as such is hard to assess the contribution of physics-based models to PV power forecasting.

We split the analysis in two, starting from estimating the effect of PV models in predicting PV AC power, given observations of GHI and ambient temperature T and the current time t. That is, we want to find a function

$$\hat{P}_{pv,t} = g\left(GHI_t, T_t, t\right) \tag{3.7}$$

which maps the aforementioned variables to the estimated PV production, and see if the reconstruction accuracy increases when considering a physics-based model for the PV. We then investigate the effect of the same models when trying to forecast the PV power for the next 24 hours, that is, we want to find a second function

$$\hat{P}_{pv,[t:t+T]|t} = f\left(GHI_{[t-T:t]}, T_{[t-T:t]}, G\hat{H}I_{[t:t+T]|t}, \hat{T}_{[t:t+T]|t}\right) \tag{3.8}$$

where $x_{[t:t+T]|t}$ means all the values of variable x from t up to $t+T$, given the information available at time t, $G\hat{H}I_{[t,t+T]|t}$ and $\hat{T}_{[t,t+T]|t}$ being the forecasts at time t for irradiation and ambient temperature provided by a numerical weather prediction service, up to time $t+T$. The rationale behind this is that, if we possess accurate enough GHI and T forecasts for the next day, we can combine them with an estimated PV model to increase the forecast accuracy. For this task, we use real data coming from 4 PV roof-mounted power plants located in Biel-Benken, Switzerland. The PV power plants are composed by differently oriented folds; their description and metadata can be found in [123].

3.2.1 Evaluation of PV modeling for prediction

Since we want to estimate the effect of a physics-based model on the accuracy of g, we a-priori chose a family of regressors and kept it fixed during the analysis. We chose to use a random forest regressor due to its ability in modeling nonlinearities, its high resiliency to overfitting and its low prediction variance [106]. We compared the performance of g against physics-based models, and then we tried to increase its predictive power including the models' output as regressors in g. The final estimated models are the following:

1. We just use the random forest to map GHI and T to the PV generated power. Instead of

including time as a regressor, we use the solar azimuth instead.

$$^{1}\hat{P}_{pv,t} = g\left(GHI_t, T_t, \theta_{a,t}\right) \tag{3.9}$$

2. The PV power is estimated using only a physic based model, as described earlier in section 3.1. We used $N = 21$ different orientations, uniformly distributed on the unit sphere, disregarding the ones facing north.

$$^{2}\hat{P}_{pv,t} = I_{N,t}\hat{\omega} \tag{3.10}$$

3. The PV power is still estimated using only a physic based model, but this time the model is identified blindly, based only on the aggregated power measurements at the main of a residential building. Firstly, we use a disaggregation technique to estimate the generated AC power production of the PV, which is then used to identify the PV model, as described in paper A

$$^{3}\hat{P}_{pv,t} = I_{N,t}\hat{\omega}_{bl} \tag{3.11}$$

4. In the fourth model, the PV power estimated with the second model is passed as a regressor to the random forest g, along with the same inputs used for the first model. The idea here is to investigate if the non-trivial nonlinear transformations which project GHI on the plane of array, can help the random forest in predicting P_{pv}

$$^{4}\hat{P}_{pv,t} = g\left(GHI_t, T_t, \theta_{a,t}, {}^{2}\hat{P}_{pv,t}\right) \tag{3.12}$$

5. This model is equal to the previous one, but uses the prediction of model 3, which were obtained through the blind identification

$$^{5}\hat{P}_{pv,t} = g\left(GHI_t, T_t, \theta_{a,t}, {}^{3}\hat{P}_{pv,t}\right) \tag{3.13}$$

Results

We evaluated the methods through k-fold cross validation on a dataset of increasing size. In particular, we have investigated the prediction performance using 10 datasets, spanning from a minimum of of 40 days up to a maximum of 400 days. For each of these datasets we estimated the performance of the methods using a 10-folds cross validation. The cross validation is done in the following way: each dataset is divided in 10 folds; for each fold we extract a training and a test set. The training set is obtained taking sequences of 3 days out of 4, for all the length of the fold. The remaining data constitutes the test set. This fold selection was done to exclude seasonality effects from the analysis. In fact, it should be noticed that this methodology is more realistic compared to simply taking the first period of the data in each fold as the training test and the last part as the testing set, as the test sets can be as long as 100 days. This would
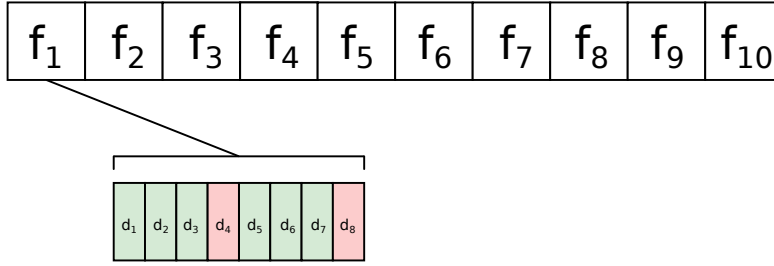
Figure 3.2 – Example of 10 folds cross validation on the 80 days dataset. The dataset is divided in 10 folds, each of 8 days. For each fold, the training set (green) is composed by 3 consecutive days each 4 days, while the test set (red) is composed by the remaining days.

lead to test the model on periods of the years which are significantly different from the data seen by g in the training set. On the other hand the aforementioned methodology mimics the behavior of estimating a model for the PV power plant once each three days, which is reasonable. Fig. 3.2 shows as an example the division of the cross validation folds for the 80 days dataset.

The used KPIs are the RMSE, MAE and their normalized versions, see equation (2.56) and (2.57) in chapter 2. The results, in form of boxplots containing all the cross validation results for the four different PV power plants (that is, each boxplot contains 40 points), are shown in Fig.3.3 and Fig.3.4 for RMSE and MAE, and their normalized versions, respectively.

From the results, we can draw the following conclusions: first, the accuracy of the identification does not show significant changes with the dataset size, for datasets larger than 120 days, for all the methods but the blind identification $^{3}\hat{P}_{pv}$. Secondly, while the random forest estimation $^{1}\hat{P}_{pv}$ has similar performance to the proxy based identification $^{2}\hat{P}_{pv}$ in terms of RMSE, the latter is slightly better for all the dataset periods when considering the MAE. Last, the most accurate methods for estimating P_{pv} are clearly $^{4}\hat{P}_{pv}$ and $^{5}\hat{P}_{pv}$, which combine the physics-based methods with the random forest regression. This could be explained by the fact that the I_N matrix used for the physics based regressions includes the projection of the GHI signal onto differently oriented planes. These projections are non-trivial, since they include the split of GHI in the diffuse and direct irradiance seen by these planes. At the same time, the physics behind these splits and projection is well-known. It seems reasonable that the random forest predictive accuracy increases when the information of these projections is (indirectly) included as a regressor.

In Fig. 3.5 the average values of the identified θ coefficient, mediated across the folds, for $^{2}\hat{P}_{pv}$ and $^{3}\hat{P}_{pv}$ are shown, based on the total number of training days. Each line of the figure refers to one of the four households hosting the PV power plants. We can see that the $^{2}\hat{P}_{pv}$ method, which only relies on the aggregated power profile, presents a dense pattern in the θ value, meaning that the true orientation of the PV folds is not accurately identified. On the other hand, the $^{2}\hat{P}_{pv}$ method, which makes use of values of measured GHI, presents a sparse and consistent pattern in the values of θ across the datasets, meaning that the dataset size is
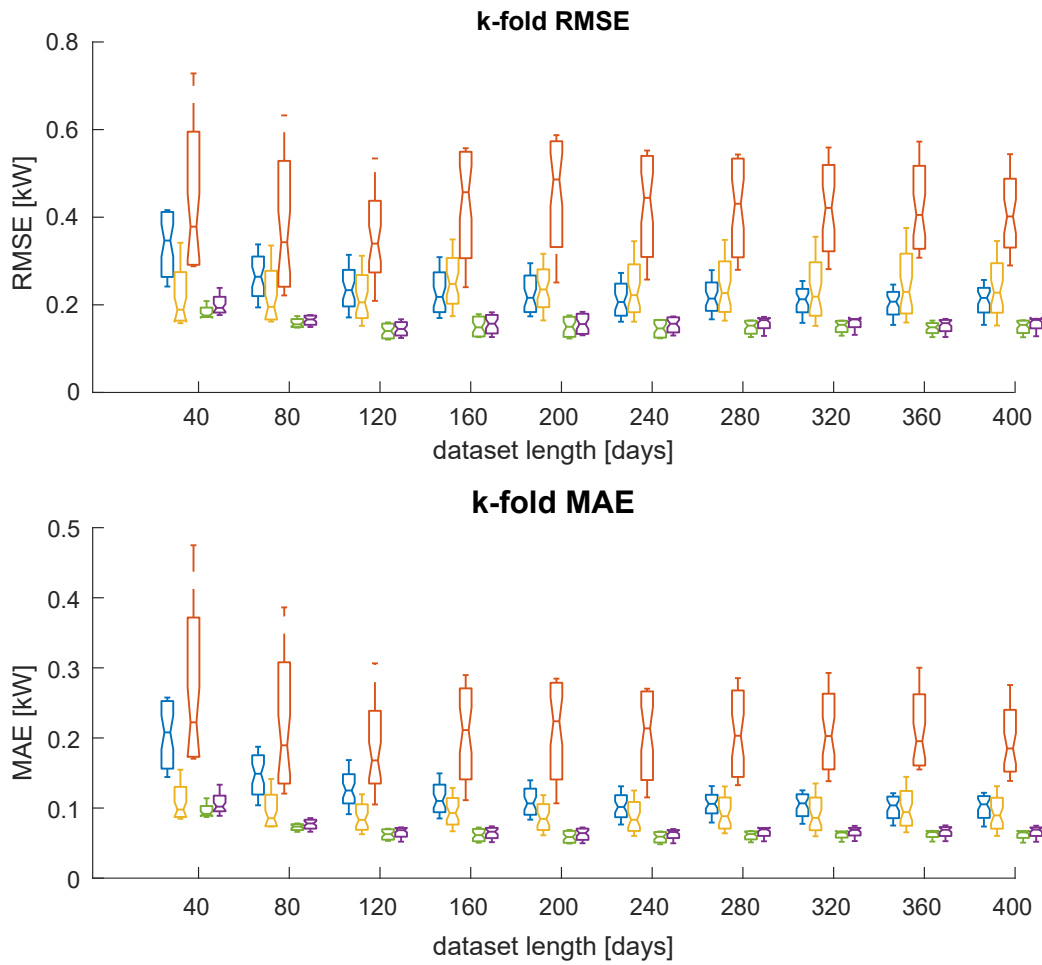
Figure 3.3 – Boxplots of the RMSE and MAE based on the number of days of the dataset and on the method of prediction. Blue: $^1\hat{P}_{pv}$, yellow: $^2\hat{P}_{pv}$, red: $^3\hat{P}_{pv}$, green: $^4\hat{P}_{pv}$, violet: $^5\hat{P}_{pv}$.
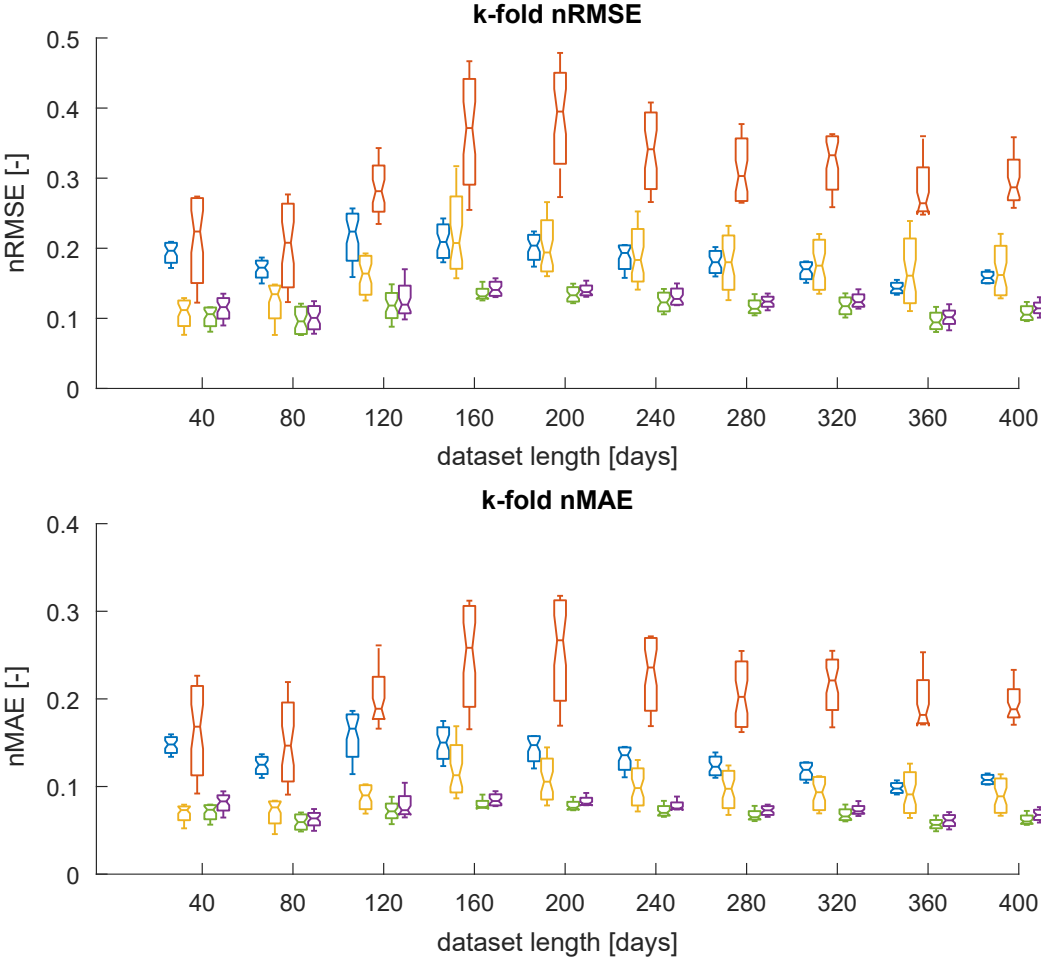
Figure 3.4 – Boxplots of the nRMSE and nMAE based on the number of days of the dataset and on the method of prediction. Blue: $^1\hat{P}_{pv}$, yellow: $^2\hat{P}_{pv}$, red: $^3\hat{P}_{pv}$, green: $^4\hat{P}_{pv}$, violet: $^5\hat{P}_{pv}$.
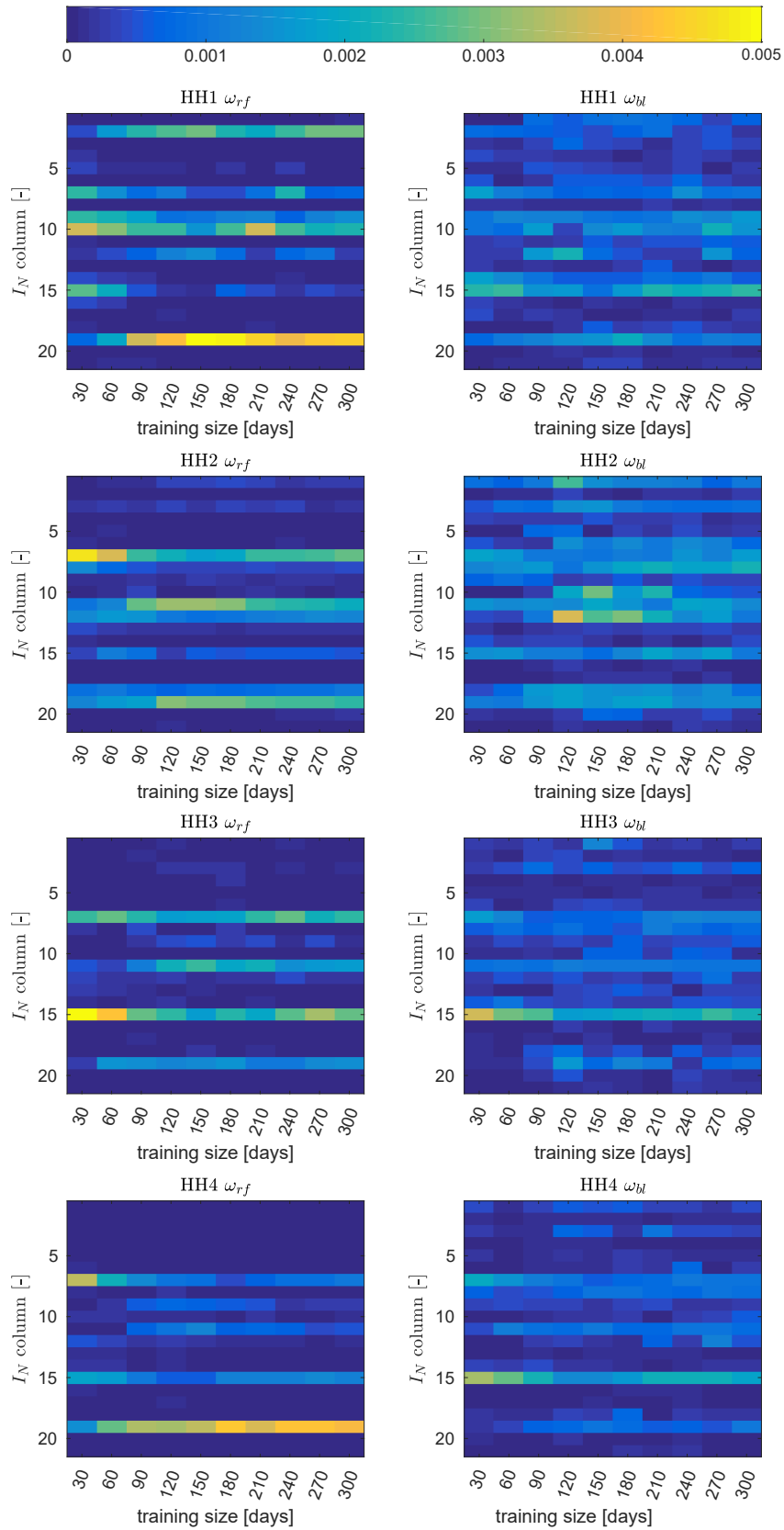
Figure 3.5 – Values of the identified ω coefficients for all the households, for the robust fit regression (left) and for the blind identification (right), based on he number of training days, mediated over the cross validation folds.

less relevant to the identified values.

3.2.2 Evaluation of PV modeling for forecasting

As mentioned previously, predicting the value of PV power output and forecasting PV production are two different task. As we have seen in the previous section, using a physics based model for PV increases the prediction accuracy when we try to estimate PV power production starting from known values of GHI, but whether this results in an increase of forecasting accuracy mainly depends on the NWP quality and resolution, as we argument in this section. Once again we used cross validation, this time using folds of 120 days, since no gain in accuracy was seen in 3.2.1 when selecting higher dataset size in PV power prediction. This time we used aggregating times with the minimum resolution of 10 minutes. The whole set of aggregation times is reported in table 3.1.

Table 3.1 – Number of minutes for each step ahead.

step	1	2	3	4	5	6	7	8	9	10	11	12	13	14
minutes	10	10	10	30	30	30	60	60	120	120	240	240	240	240
hours ahead	0.17	0.33	0.5	1	1.5	2	3	4	6	8	12	16	20	24

In order to estimate the influence of the PV modeling on forecasting, we evaluated the performance of a QRF, which was the forecaster with the higher accuracy among the one evaluated for the power forecast in the analysis presented in 3.2.1, on the Biel-Benken dataset. We used two variants of the ${}^{4}\hat{P}_{pv,t}$ model (equation 3.12) in order to increase the accuracy of the forecast. Namely, in each fold we estimated a physics based model, using the technique previously explained . Secondly, with the same data, we trained a random forest regressor to learn the map from GHI, T and the solar azimuth to the power production. We then fed this model with the NWP forecast for GHI and T, and then we use the result as an explanatory variable for the QRF. Formally, the forecaster can be described as:

$$
\begin{aligned}
\hat{P}_{pv,[t:t+T]|t} &= f(X) \\
X &= \left[GHI_{[t-T,t]|t}, T_{[t-T:t]|t}, {}^{4}\hat{P}_{pv,t}\left(\hat{\theta}, G\hat{H}I_{[t:t+T]|t}, \hat{T}_{[t:t+T]|t}\right) \right]
\end{aligned}
\tag{3.14}
$$

where ${}^{4}\hat{P}_{pv,t}$ is the best model for the PV power prediction 3.12. The difference from the prediction task is that now the model makes use of the NWP forecast $G\hat{H}I_{[t:t+T]|t}$ and $\hat{T}_{[t:t+T]|t}$. We used MeteoBlue [124] as NWP service, which provides local forecasts for GHI and T at hourly resolution for the next 48 hours, with 12 hours updates. As such the accuracy of the NWP forecasts is not constant with the step ahead, but is also dependent from the time of the day. Some authors use Kalman filters to reduce the forecast error, but this method needs to retrieve a dynamic model for the error, which in this case has discontinuities (twice per day, at the moment of the update), thus it is hard to model with a simple autoregressive

model. To perform the correction we decided once again to use a random forest which we fed with the perfect forecasts from historical values of GHI and T, and the hour of the day. We stress that the RFs for the NWP correction and the models for the PV prediction $\left(^4\hat{P}_{pv,t}\right)$ were trained and identified in each fold, in order to keep the results from different folds statistically independent.

In the second variant we assumed the lack of a pyranometer for the measurement of the local value of GHI. As such, we identified the PV model, $\hat{\theta}$, directly from the PV AC power measurements, using only observations from clear sky periods and exploiting robust regression. In this case, in order to correct the NWP forecasts, we reconstructed the local GHI seen by the PV panel starting from the identified model, as described in [123].

We started assessing the effect of the estimated PV models using perfect forecasts for GHI and T in order to have a lower bound for our evaluations. In Fig. 3.6 and Fig. 3.7 boxplots containing the $nRMSE$ and $nMAE$ for all the households and all the folds, as a function of the step-ahead, are shown. The blue boxplots refer to the base case predictions, where no PV models are estimated, while the red and yellow boxplots refer to the first and second variants of PV model estimations. We can see that the improvements in forecast accuracy due to the estimation of PV models are significant both in $nRMSE$ and $nMAE$, especially for the most aggregated time-steps. Additionally, the effect of estimating $^4\hat{P}_{pv,t}$ without knowing GHI is marginal.

When using NWP for GHI and T, PV modeling favorably affects the forecasts accuracy only starting from 1 hour ahead, as can be seen in Fig. 3.9, where the $RMSE$ ratio between the base case and the PV modeling cases are shown. The positive influence of PV models vanishes again starting from the 12_{th} step, corresponding to 12 hours ahead. This result can be better explained taking into account the accuracy of the NWP forecasts, as a function of the step ahead and aggregation. Fig. 3.10 shows the boxplot of the RMSE of GHI in J/m^2, normalized to the observed energy in J/m^2. The distributions in the boxplots are obtained randomly sampling one tenth of the dataset for 50 times. It can be seen that the $RMSE$ decreases with increasing aggregation steps, up to 12 hours ahead (the 11_{th} step), at which point the NWP starts increasing. This is due to the fact that the NWP forecasts are updated every 12 hours, at midnight and at noon.

We additionally investigate the loss of accuracy due to the resolution of the forecasts: since the NWP are available with a sampling time of one hour, we are interested in the loss of accuracy when the perfect forecasts are downsampled using the same resolution. In Fig. 3.8 the $nRMSE$ for the base case forecast, in which no PV models are used, are shown. The blue boxplots refer to the perfect forecasts, while the green and bordeaux refer to the 1 hour downsampled perfect forecasts and to the real forecasts, respectively. One can see that in the first timesteps, the distribution of the $nRMSE$ of the real forecasts is close to the distribution of $nRMSE$ of the downsampled perfect forecasts. This means that there is little or no bias in the corrected NWP forecasts. The results in Fig. 3.9 can be explained additionally considering the effect of PV
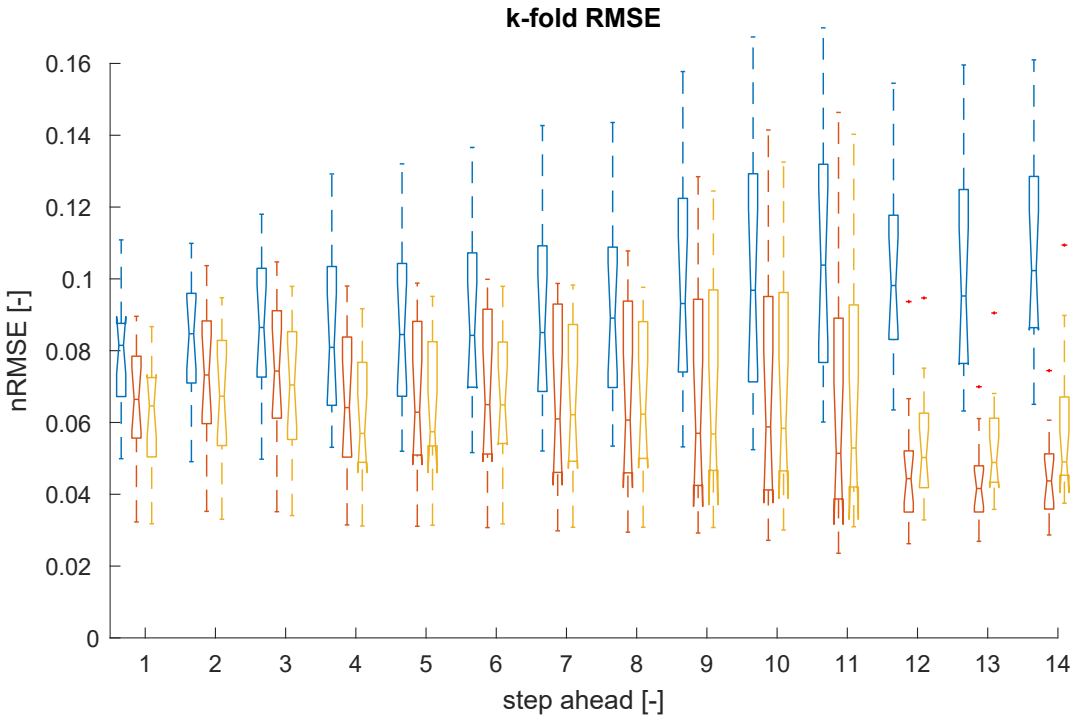
Figure 3.6 – nRMSE as a function of step ahead for perfect forecasts. Blue: base case. Red: with PV model. Yellow: with PV model estimated without GHI.
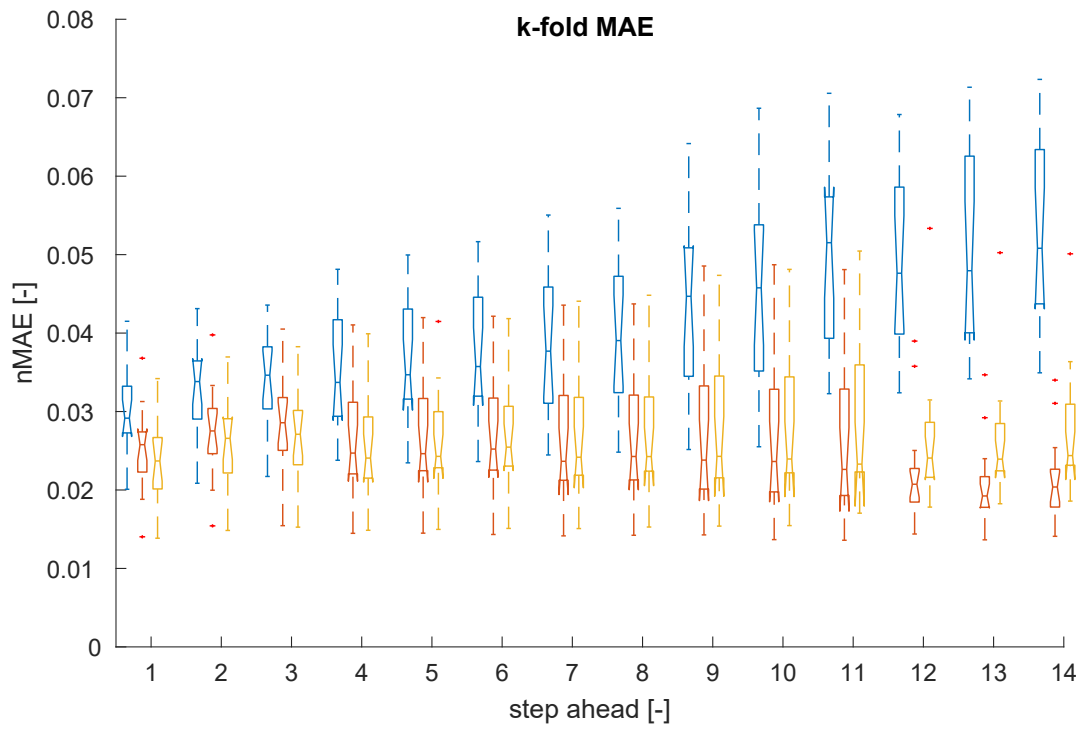
Figure 3.7 – nMAE as a function of step ahead for perfect forecasts. Blue: base case. Red: with PV model. Yellow: with PV model estimated without *GHI*.

modeling when using the 1 hour downsampled perfect forecasts, as shown in Fig. 3.11. The effect of modeling PV when using the 1 hour sampling time resolution is negligible for the first 3 step ahead, while increasing for the last steps ahead. We can conclude that the NWP forecasts accuracy and temporal resolution for GHI and T are not accurate enough to induce a decrease in the forecast error when using a PV model. Anyway, one can see from Fig. 3.8 that the accuracy of the real forecasts is already close to the 1 hour averaged perfect forecasts. This means that the (anyway modest) increase of accuracy that can be seen in Fig. 3.6 for the first step ahead is mainly due to the perfect knowledge of GHI. It seems not reasonable that the NWP accuracy can be improved for the first step ahead. This is mainly due to the fact that GHI signal has a typically high variance during overcast days. Any low-variance forecaster which uses a quadratic loss will tend to smooth out the high frequency components of the signal to be predicted, since under least squares error minimization criterion the conditional expected value of the signal is the best minimizer. On the other hand, one can see from Fig. 3.8 that PV modeling decreases the error starting from the 7_{th} step ahead (which corresponds to 2 hours ahead) up to 24 hours ahead, even when using 1 hour sampling time. At the same time we can see from Fig 3.8 that the last step ahead NWP forecast accuracy is distant from the perfect forecasts downsampled signal. This means that if the NWP forecasts accuracy increases for the last steps ahead, this will results in an increase of accuracy in PV power prediction when using PV modeling.

At last, we repeated the same analysis using only clear day samples. The clear days are identified as the 10% of days which shows the lowest error between the NWP forecasts and the expected extra-terrestrial irradiance:

$$
\begin{aligned}
\epsilon_{cl,t} &= \frac{\frac{1}{n}\sum_{k=t}^{t+H} G\hat{H}I_{k|t}}{\frac{1}{n}\sum_{k=t}^{t+H} E_k} \\
s_t &= \begin{cases} 1 & if \quad \epsilon_{cl} < q_{0.1}(\epsilon_{cl}) \\ 0 & otherwise \end{cases}
\end{aligned}
\tag{3.15}
$$

where s_t is the indicator for clear day selection, E_k is the extraterrestrial irradiance, which is known for a given time and geographical location, $G\hat{H}I_{k|t}$ is the NWP forecasted of GHI available at time t for time-step k, $\epsilon_{cl,t}$ is the normalized error at time t, ϵ_{cl} is the vector of all the normalized error for all the dataset, H is the number of step-ahead and q_α stands for the quantile of level α. Note that both signals are known in advance, so that this filter can be actually implemented to switch between different forecasting models. In Fig.3.12 the empirical cumulative distribution function (ECDF) of the nRMSE for the whole forecasting horizon, for the case of perfect forecasts, is shown. Formally, we plotted the ECDF of

$$
nRMSE_t = \left[\frac{1}{H} \sum_{sa=1}^{H} nRMSE_{sa} \right]^{\frac{1}{2}}
\tag{3.16}
$$

where $nRMSE_{sa}$ is the normalized RMSE for step ahead sa, described in equation (2.56).
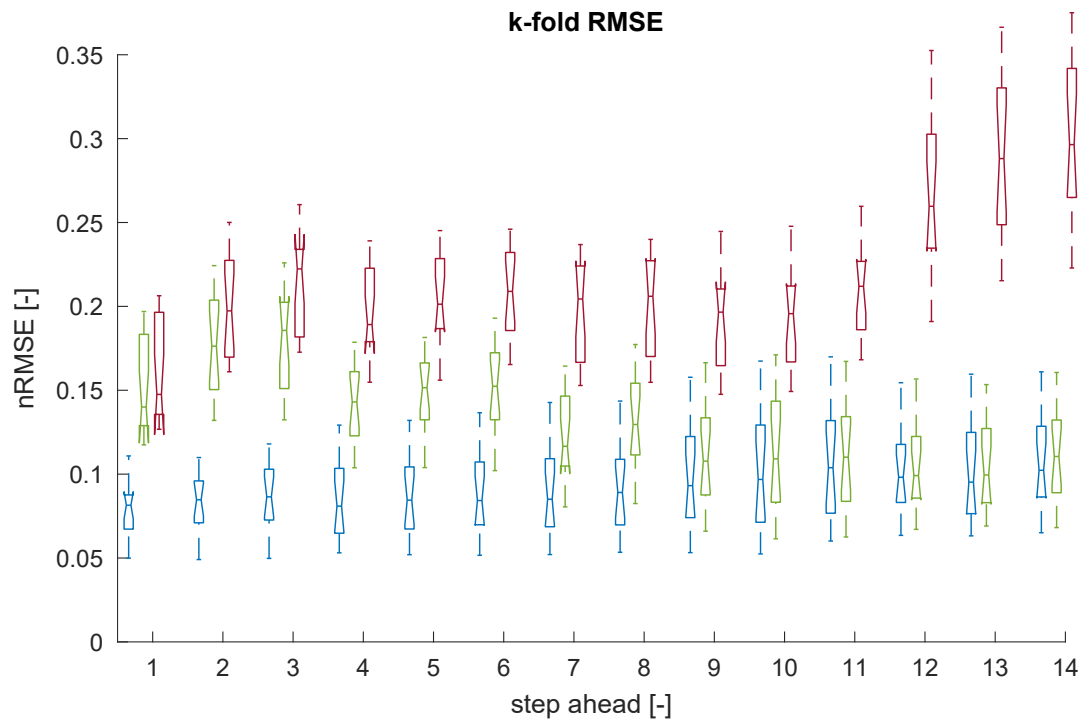
Figure 3.8 – nRMSE for perfect forecasts (blue), perfect forecasts downsampled (green) and real forecasts (bordeaux), for the base case (no PV models)
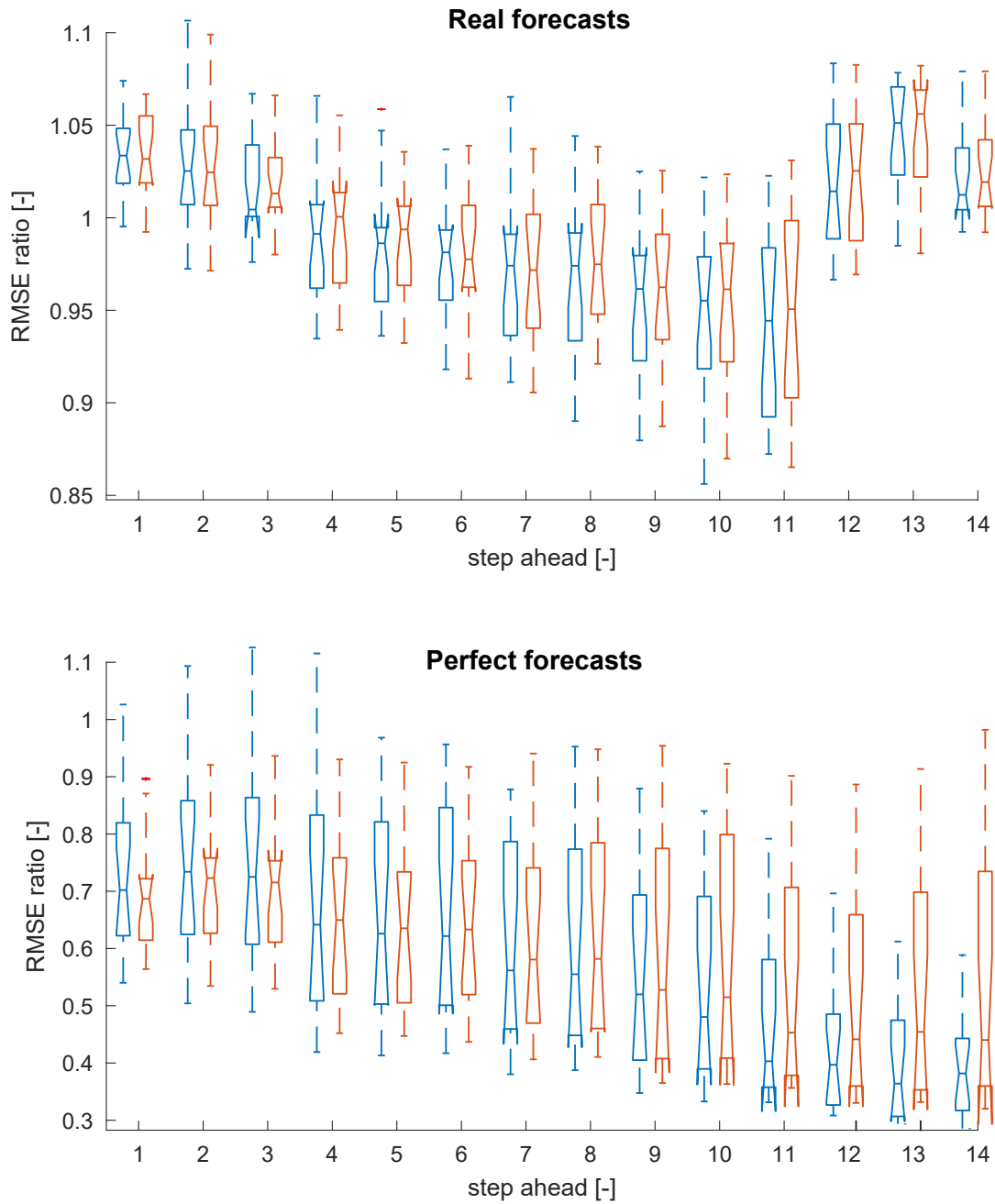
Figure 3.9 – RMSE of the forecast obtained using the PV models, normalized to the RMSE of the base case. Blue: with PV model. Red: with PV model estimated without GHI. When using NWP forecasts, PV modeling results beneficial for the step ahead in which the NWP accuracy is higher (top), while consistently increasing for the case in which perfect forecasts are used (bottom).
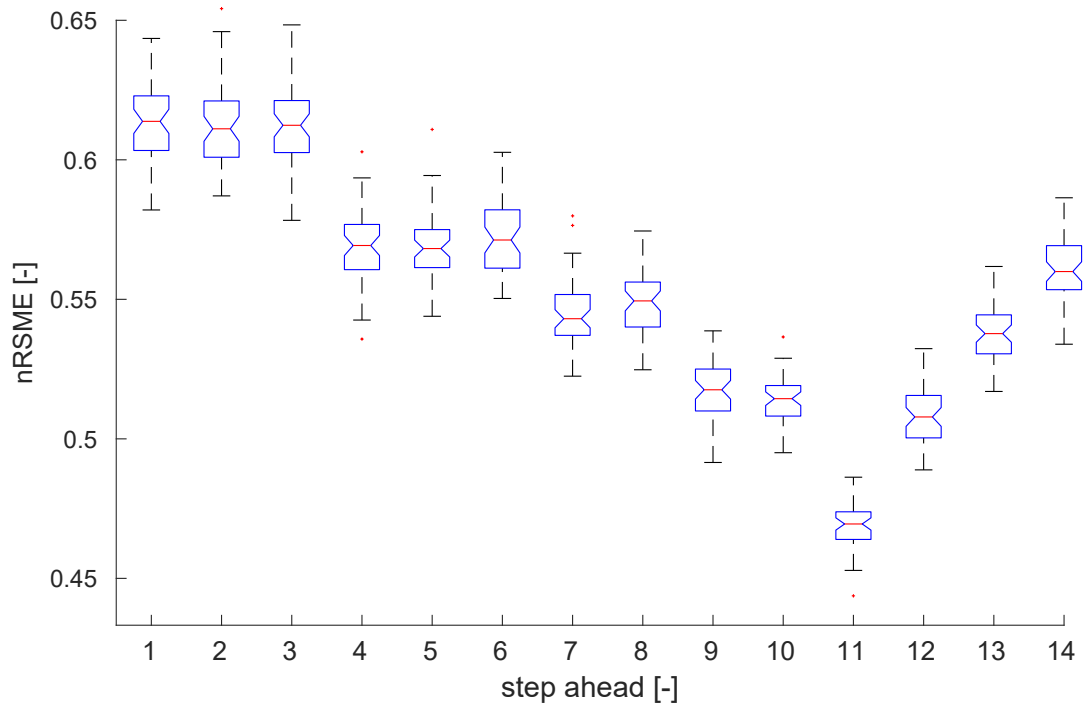
Figure 3.10 – RMSE of NWP forecasted GHI, as a function of the step ahead, normalized with the total observed energy per step.

The continuous lines refer to the whole dataset, while the dotted lines are the ECDFs referring to the clear sky dataset, which was obtained using only observations for which $s_t = 1$. It is clear that when the NWP for GHI for the next 24 hours are close to the extraterrestrial irradiance, the forecasting error for the whole horizon is significantly lower. Fig. 3.13 shows the same results when NWP forecasts for GHI and T are used. Also in the case of clear days, modeling the PV does not significantly increase the accuracy of the forecasts, since when no clouds are present, splitting GHI in its direct and diffuse components is easier, and the map which links PV production to GHI is much easier to learn.
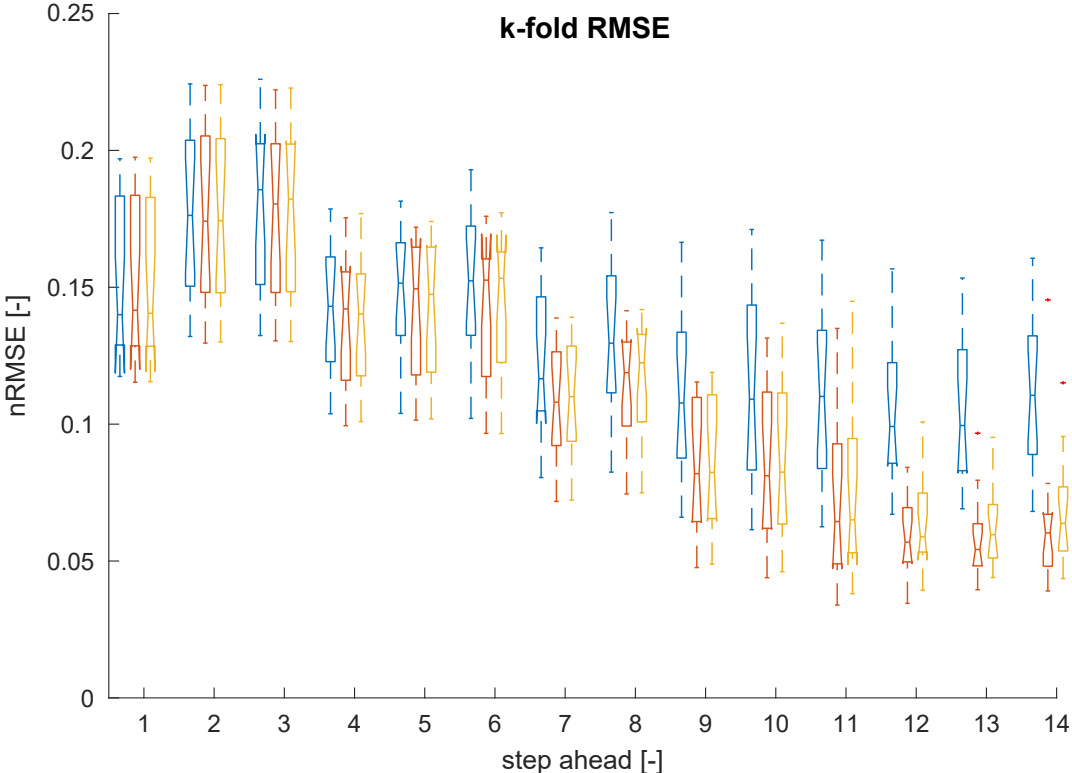
Figure 3.11 – nRMSE as a function of step ahead for 1 hour downsampled perfect forecasts. Blue: base case. Red: with PV model. Yellow: with PV model estimated without *GHI*. The effect of modeling PV is negligible for the first 3 steps ahead.
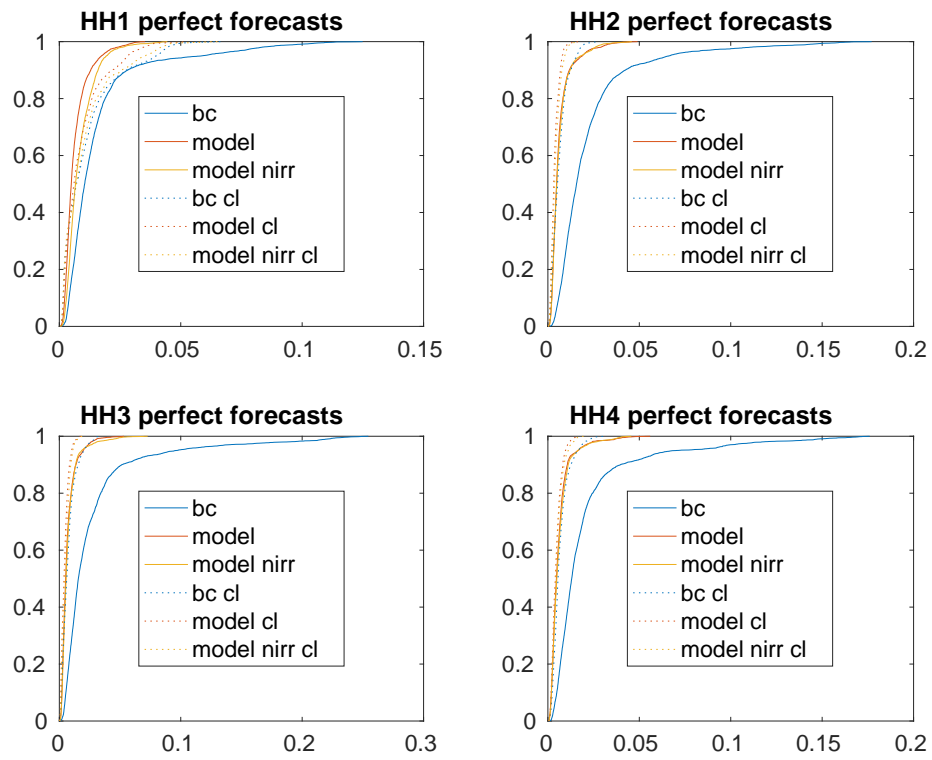
Figure 3.12 – ECDF of the horizon nRMSE for the base forecast and the two PV model forecasts, for each household. The dotted lines are referred to the clear sky day dataset.
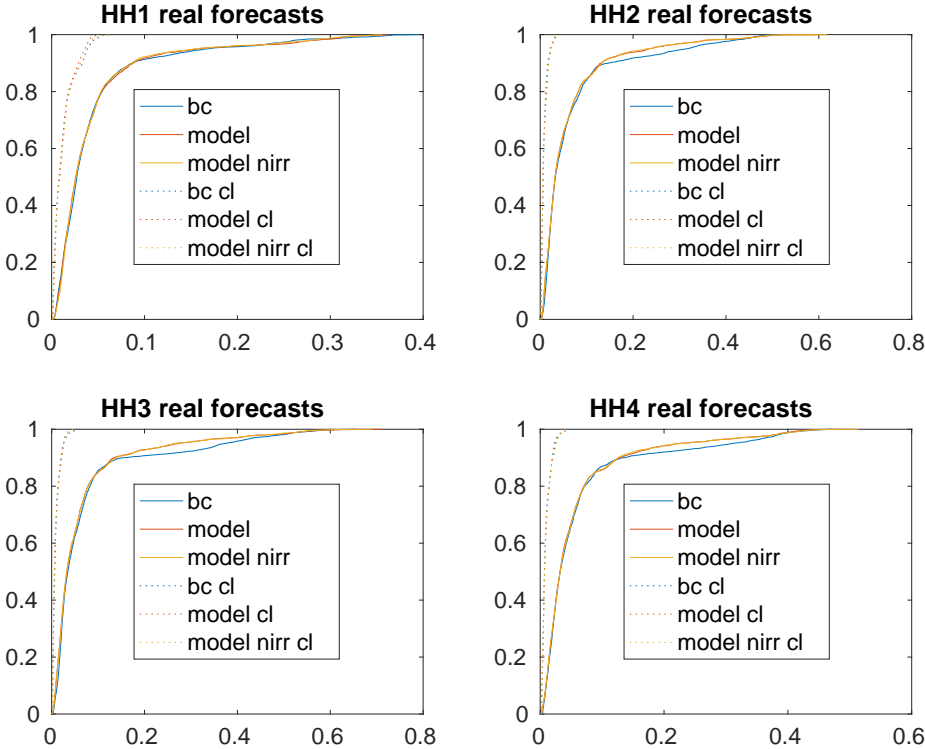
Figure 3.13 – ECDF of the horizon nRMSE for the base forecast and the two PV model forecasts, for each household. The dotted lines are referred to the clear sky day dataset.

4 Hierarchical forecasts for distributed control algorithms

When forecasting power consumption in the electrical grid, we are typically interested in different levels of aggregation. For example, when we aim at controlling an HVAC system of a residential building for minimizing future expenses, we are interested in single household level forecasts. On the other hand, when we want to provide a peak-shaving service to a local DSO, we are most interested in the forecast of its entire power consumption. When we try to optimize for both a local (ar household level) and a global (e.g. at the PCC with the MV grid) objective functions, we must require consistency in the forecasted power profile. This means that forecasts at households level should sum up to the forecast of the aggregated power profile, which is not guaranteed when forecasting them separately. Moreover, single house electrical consumption is usually difficult to predict, due to the high variability of the signal. Overall forecast accuracy could be increased performing a forecast of the aggregated signal and using hierarchical forecasting techniques to reconcile the measurements. In the second part of the chapter we focus on the reconciliation of probabilistic hierarchical forecasts, in the presence of PV.

The main outcomes of this section are the following:

- In section 4.1, a new distributed method to reconcile forecasters at different levels of a hierarchical structure is presented. This method can be used to make aggregate-consistent forecasts, thus usable in distributed control. The main advantage in redistributing the reconciliation is that private information, which could be used by the base forecasters, is not disclosed. Furthermore, informations at upper levels of the hierarchical structure is only available by means of aggregated power profiles.

- In section 4.2 a new method to obtain aggregated consistent pdfs for hierarchical power forecasts is presented. We show that nontrivial methods for summing the bottom level forecasts' pdf are needed especially in the case of high penetration of PV. In this case forecasting errors becomes dependent, due to imperfect NWP.

4.1 Distributed hierarchical forecasting

4.1.1 Introduction

Consistency can be enforced by encoding the hierarchical structure in a learning algorithm. One way of achieving this, is to firstly obtain forecasters for all the levels of the hierarchy and then reconcile them based on the hierarchical structure. Following this approach, in [90] the authors used ordinary least squares (OLS) regression to reconcile the forecasts in the hierarchy. Elaborating on this approach, [125] proposed a trace minimization method in which the covariance matrix of the forecasters error is estimated to perform a weighted least squares regression. In [126], an elastic net penalization was proposed in order to induce sparseness in the forecasters adjustments, and benefit was shown on the reconciliation of the forecasts for the power consumption of residential consumers. We included this variant in our analysis. In the following we present a distributed algorithm to obtain hierarchical reconciliation of the different time series. Being able to reconcile time series through a distributed algorithm allows to not fully disclose informations about individual time series, respecting the privacy of individual prosumers. In fact, this information would be available only in an aggregated form to the upper level of the structure. Apart from the basic case in which hierarchical reconciliation techniques are useful due to geographical smoothing of the power consumption, the distributed mechanism can be also beneficial in the case in which prosumers possess additional information about their future consumption, for example the internal scheduling of their HP or EV chargers. Moreover this technique can be applied to temporal hierarchy [92], but in this case decomposing the problem is less interesting, since typically a single entity would be in possess of the information needed to apply the technique in this case.

4.1.2 Problem formulation

We consider a hierarchical structure which can be described by a rooted tree, which is a unidirected acyclic graph, with every node having exactly one parent, except for the root node. Each node is identified by a tuple $(d_1...d_i...d_l)$ where l is the level to which the node belongs, and each entry represents the enumeration of its i_{th} level ancestor. Formally, we indicate with τ the set of all the nodes in the tree. Given the forecasts for the next t timesteps of all the n nodes of the rooted tree, called the base forecasters, we can collect them in the matrix $T \in \mathbb{R}^{t \times n}$. Reconciling the forecasts is then equal to the task of finding the set of bottom level forecasts $X \in \mathbb{R}^{t \times n_b}$ which minimizes the residual ϵ

$$T = [T_u, T_b] = XS^T + \epsilon \tag{4.1}$$

where $T_u \in \mathbb{R}^{t \times n - n_b}$ and $T_b \in \mathbb{R}^{t \times n_b}$ are the matrices of the upper level and bottom level base forecasters. The intuition behind this is that we are seeking for a set of latent variables, X, which generate the forecasts at all the levels, through the summation matrix S. For example,

for a 3 levels hierarchy with two nodes in the second level, with 4 bottom forecasters, the matrix S would be the following:

$$S = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.2}$$

If the covariance matrix of ϵ, W, is known, the optimal solution of the reconciliation is then given by solving the generalized least squares problem [90]

$$X^* = \underset{X}{\operatorname{argmin}} \frac{1}{2} \| T - X S^T \|_W \tag{4.3}$$

where X^* is the set of reconciled bottom forecasts, with the analytical solution

$$X^* = (S^T W^\dagger S)^{-1} S^T W^\dagger T^T \tag{4.4}$$

where $W^\dagger$ is the pseudoinverse of W. In the following we limit ourselves to the case in which W is diagonal, the simplest case being the one proposed in [90], in which W is the identity matrix. We start splitting problem (4.3) among the nodes of the rooted tree:

$$\underset{X}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{w_i}{2} \| \hat{t}_i - X s_i^T \|_2^2 \tag{4.5}$$

where $s_i \in \mathrm{IR}^{n_b}$ are the i_{th} rows of the summation matrix S and $\hat{t}_i$ are the base forecasts. We can then decompose the problem introducing additional variables y_i:

$$\begin{aligned} \underset{X,Y}{\operatorname{argmin}} \quad & \sum_{i=1}^{n} \frac{w_i}{2} \| \hat{t}_i - y_i \|_2^2 \\ s.t \quad & y_i = X s_i^T \quad \forall\, i \in \{n\} \end{aligned} \tag{4.6}$$

we proceed with an augmented Lagrangian relaxation to turn the problem in a set of unconstrained optimizations, and solve it with an ADMM [127] strategy. The overall problem becomes:

$$\underset{X,Y}{\operatorname{argmin}} \sum_{i \in \mathscr{P}} \frac{w_i}{2} \| \hat{t}_i - y_i \|_2^2 + \frac{1}{2\rho} \| y_i - X s_i^T + \lambda_i \|_2^2 + \sum_{i \in \mathscr{B}} \frac{w_i}{2} \| \hat{t}_i - x_i \|_2^2 \tag{4.7}$$

where $\lambda_i \in \mathrm{IR}^T$ are the Lagrangian multipliers associated to the constraints in (4.6), $\mathscr{P}$ is the set of nodes not belonging to the set of terminal nodes $\mathscr{B}$, formally $\mathscr{P} = \tau \setminus \mathscr{B}(\tau)$. Problem 4.7 can

be interpreted as a sharing problem, in which each node of the hierarchy tries to minimize the distance of its decision variable (the latent variables, that is the reconciled forecasts) from its target (the original base forecasts $\hat{t}_i$), while being subject to the structural constrained encoded by the summation matrix S. We then use a parallelized formulation of the sharing problem, which makes use of ADMM [127]. The resulting formulation is the following:

$$
\begin{aligned}
x_i^{k+1} &= \underset{x_i}{\operatorname{argmin}} \frac{w_i}{2} \| \hat{t}_i - x_i \|_2^2 + \frac{1}{2\rho} \| x_i - r_i \|_2^2 \quad \forall\, i \in \mathcal{B} \\
y_i^{k+1} &= \underset{y_i}{\operatorname{argmin}} \frac{w_i}{2} \| \hat{t}_i - y_i \|_2^2 + \frac{1}{2\rho} \| X^{k+1} s_i^T - y_i + \lambda_i \|_2^2 \quad \forall\, i \in \mathcal{P} \\
\lambda_i^{k+1} &= \lambda_i^k + X^{k+1} s_i^T - y_i^{k+1} \quad \forall\, i \in \mathcal{P}
\end{aligned}
\tag{4.8}
$$

where r_i is a reference signal coming from the parent node of node i:

$$
r_i = \sum_{a \in \mathcal{A}_i} \left(y_i - X^k s_a^T \right) / n_a + x_i^k - \lambda_a
\tag{4.9}
$$

where $\mathcal{A}_i$ is the set of ancestors of node i and n_a is the number of the children of the ancestor a. Intuitively, (4.8) and (4.9) divides equally (division by n_a in (4.9)) the quadratic loss needed to respect the consistency constraints when moving away from the target (the original forecasts). Note that the minimizations in (4.8) have analytical solutions, so that the final algorithm can be rewritten as:

$$
\begin{aligned}
x_i^{k+1} &= \frac{r_i + \hat{t}_i \rho w_i}{n_l + \rho w_i} \quad \forall\, i \in \mathcal{B} \\
y_i^{k+1} &= \frac{\rho w_i \hat{t}_i + X^{k+1} s_i^T + \lambda_i}{1 + \rho w_i} \quad \forall\, i \in \mathcal{P} \\
\lambda_i^{k+1} &= \lambda_i^k + X^{k+1} s_i^T - y_i \quad \forall\, i \in \mathcal{P}
\end{aligned}
\tag{4.10}
$$

where n_l is the number of levels in the hierarchy. The algorithm can thus be computed only using summation and multiplication; furthermore, it can be solved following a forward-backward strategy. The forward passage consists in each parent node sending the updated Lagrangian multipliers λ_i downward through the hierarchy. When the Lagrangian is received by a non-terminal node, this will send it, together with its own Lagrangian, to its children. This allows terminal nodes to compute x_i, since these depend on the the sum of the λ_i coming from all of their ancestors, as described in (4.9). In the backward passage, the terminal nodes compute their update for x_i as in (4.10), and send it upward to their ancestors. Note that each ancestor only needs information from its own children to compute its minimization, since $X^{k+1} s_i^T$ filters out all the other optimization variables in X^{k+1}. As soon as the ancestors computes their optimization, they send the solution, and so on, up to the root node. Lagrangian

multipliers are updated in a similar fashion.

4.1.3 Inducing regularization

It is easy to see that when W is the identity matrix, the solution of (4.3) just retrieves a set of bottom level forecasts which mimimize the distance of aggregate consistent forecasts with the original base forecasts T. This would totally ignore the historical accuracy of the base forecasters. In fact, in the case in which some base forecasters present a higher accuracy with respect to the others, we should include it in the reconciliation. This is possible through matrix W. Anyway, estimating W is difficult and was avoided in [90], where it was replaced with the identity matrix. In [125] W is directly estimated for historical error covariance matrix. We follow the approach reported there, also used in [126]:

$$
\begin{aligned}
W &= \theta W_d + (1 - \theta) W_1 \\
W_d &= diag(W_1) \\
W_1 &= \mathbb{E}\left(ee'\right) \\
\theta &= \frac{\sum_{i \neq j} Var\left(\hat{r}_{i,j}\right)}{\sum_{i \neq j} \hat{r}_{i,j}^2}
\end{aligned}
\tag{4.11}
$$

where $\hat{r}_{i,j}$ are the elements of the one step ahead sampled covariance matrix. More details on the computation of θ can be found in [128]. In practice, though, this method alone can induce too large adjustments in the bottom level forecasts, since it does not allow the base forecasters to be unchanged, and could lead to reconciled forecasts with poor prediction accuracy. A regularization technique can be applied to the reconciliation problem with favorable results, as shown in [126], where large displacement of X^* from the bottom level base forecasters are punished, inducing a sparsity structure in the forecasts corrections. We refer to this method as the minT strategy. This method can be readily included in our distributed algorithm, adding to the second line of (4.8) a punishment for the deviation of the upper level forecasts from the one generated by the original bottom level forecasts T_b:

$$
k\left(\frac{1-\alpha}{2}\|y_i - T_b s_i^T\|_2^2 + \alpha\|y_i - T_b s_i^T\|_1\right)
\tag{4.12}
$$

Similarly, the same punishment can be added for the bottom level forecasts (first line of (4.8)). In this case, the minimization in (4.8) has still a closed form, which is equal to the proximal operator of the L1 norm, also known as the soft threshold operator [129]. For completeness, the final equations are reported below:

$$
x_i^{k+1} = \begin{cases} \frac{a_{x,i} - \rho\alpha}{b_{x,i}} & \textbf{if} \quad \frac{a_{x,i} - \rho\alpha}{b_{x,i}} > 0 \\ \frac{a_{x,i} + \rho\alpha}{b_{x,i}} & \textbf{if} \quad \frac{a_{x,i} + \rho\alpha}{b_{x,i}} < 0 \\ t_i & otherwise \end{cases} \quad y_i^{k+1} = \begin{cases} \frac{a_{y,i} - \rho\alpha}{b_{y,i}} & \textbf{if} \quad \frac{a_{y,i} - \rho\alpha}{b_{y,i}} > 0 \\ \frac{a_{y,i} + \rho\alpha}{b_{y,i}} & \textbf{if} \quad \frac{a_{y,i} + \rho\alpha}{b_{y,i}} < 0 \\ t_i & otherwise \end{cases}
\tag{4.13}
$$

where t_i are the original base forecasters of the i_{th} node, and

$$
\begin{aligned}
a_{x,i} &= t_i \left(\rho \, w_i + \rho (1 - \alpha) \right) + r_i \\
b_{x,i} &= \rho \, w_i + \rho (1 - \alpha) + n_l \\
a_{y,i} &= \rho \, w_i t_i + X s_i^T + \lambda_i + \rho (1 - \alpha) T_b s_i^T \\
b_{y,i} &= \rho \, w_i + \rho (1 - \alpha) + 1
\end{aligned}
\tag{4.14}
$$

and the λ update is the same as in (4.8).

4.1.4   Results

In order to estimate the effect of forecast reconciliation, we used 2200 power profiles coming from a smart meter trial measurement campaign in Great Britain provided by AECOM, which is available at `https://www.ukdataservice.ac.uk/`. In Fig. 4.1, the error in terms of $RMSE$ is reported for different levels of the hierarchy. The first column refers to the mean RMSE in the whole hierarchy, while the second and third columns refer to the top and bottom level, respectively. We can see how the bottom-up approach, which is consistent by construction, introduces an error with respect to the original forecasts for the top level. The use of reconciliation solving (4.3) when W is the identity matrix, doesn't have a significant effect on the top layer prediction, while reducing the mean RMSE for the bottom level forecasters. The use of the minT strategy, on the other hand, helped to reduce RMSE for both the top and the bottom levels. We report no significant differences in the results obtained with the centralized solution and the distributed approach. Finally, we stress out that the increase of accuracy in the bottom level forecasters when applying hierarchical reconciliation is due to the smoothing effect during the aggregation process, and uncorrelated errors of the bottom level hierarchy. In the case in which forecast errors of the bottom levels are highly correlated, hierarchical reconciliation can still be used to obtain aggregate consistent forecasts, but the increase of accuracy in the bottom level forecasts will be reduced. Furthermore, retrieving an aggregate consistent pdf is a more complex problem with respect to considering point forecasts. In the next sections these problems are addressed, with a particular attention to the forecast of PV power production.

4.2   Forecasting sums of random variables

As anticipated in the last section, consistency among forecasts at different level of aggregation is essential when we are trying to apply DSM in the distribution grid. This is mainly due to the presence of grid constraints, which link the objective functions of the agents across different portions of the distribution grid. Since we are interested in applying DSM in a probabilistic setting, achieving aggregate-consistent point forecasts is not sufficient. We are rather interested in retrieving aggregate-consistent pdfs, which is a much harder task in general. In particular, in the presence of high penetration of rooftop installed PV systems, as
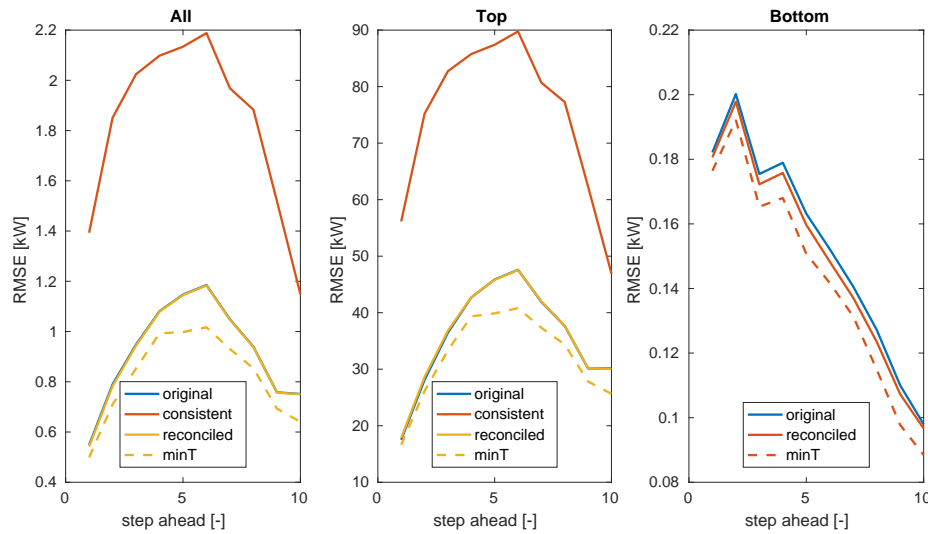
Figure 4.1 – RMSE for the base forecasters, the reconciled profiles with W equal to the identity matrix, and the reconciliation using the minT strategy. The first column refers to the mean RMSE across the whole hierarchy, the second to the error of the top level, and the third to the error in the bottom forecasters.

we will see, the errors become mutually dependent, making the summation of the forecasted pdfs hard.

A second motivation for having aggregate-consistent forecasts is when we are trying to optimize both local and global objective function, since in this case, contrasting interests of agents could lead to suboptimal solutions. This issue is further explained in chapter 5, where the case of energy sharing communities is investigated form a game theoretic point of view. In this case, prosumers can be described as selfish non-coopeartive agents, which are willing to minimize their own total energy costs. If the pdf of their own power profiles is not aggregate consistent with the pdf of the power at the PCC of the self consumption community, the problem decomposition will become intractable.

In this section we present a method which generates aggregate consistent probability distributions for power forecasts, without the need of estimating a multidimensional empirical copulas. We report that in [130], a copula-based hierarchical risk aggregation method [59] for the estimation of aggregate-consistent pdfs is used. However, these techniques, which still requires to estimate a big number of copulas, only model child-parent probabilistic relations, and thus can not be used for stochastic coordination. The method proposed in this section, although being less generic, guarantees to generate samples from all the nodes of the tree, which are consistent by construction. The main outcomes from this section are summarized in the following:

- We show how errors of forecasted PV generated power are mutually dependent, due to
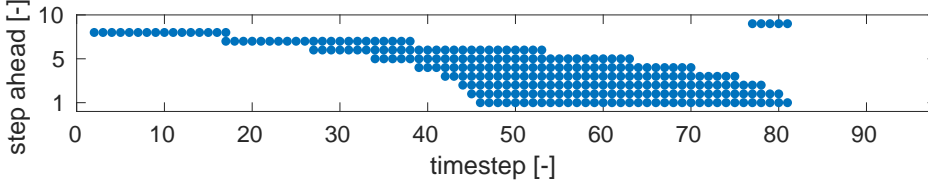
Figure 4.2 – PV error selection mask for one day of observations, with 96 timesteps of 15 minutes each. The mask entries are 1 if in all the time window of that step ahead $\theta_{az} > 0$.

the error on meteorological forecasts.

- Common modeling assumptions for estimating sum of random variables from the marginal probabilities are too stringent in the case, and would result in inaccurate pdfs.

- A method for retrieving aggregate-consistent pdfs of sums of random variables for power profiles in the presence of PV is presented, with experimental results form measurements of 8 PV power plants.

4.2.1 Correlation in PV forecast errors due to imperfect NWP

We used the synthetic power profiles described in section 2.1 to investigate up to which extent the hypothesis of correlated forecasts errors for the PV power forecasting holds. In particular, we want to understand if, in the case of correlated errors, this correlation is due to imperfect NWP forecasts of GHI and T. We generated 3 sets of 100 profiles, each of which containing only PV, HP or uncontrolled profiles. Additionally, we used 2200 power profiles from the UK dataset described in section 4.1. For each of the 4 datasets we retrieved forecasts for the 24 hours ahead, using the same logarithmically spaced bins reported in table 2.2. Since for this analysis we are not interested in the accuracy of the forecasters, but rather in seeing the effect of considering perfect or real NWP forecasts for the meteorological data, we used only one class of forecaster. We chose to use the QRF model described in section 2.3, since was the one with the best a-priori scores for all the considered KPIs.

In order to investigate the correlation of the forecast errors, we computed the expectation of the Pearson correlation coefficient over the test dataset:

$$\rho_{i,j,k} = \left| \frac{cov(\epsilon_{i,k}, \epsilon_{j,k})}{\sigma_{i,k}\sigma_{j,k}} \right|$$
$$\rho_{c,i,j,k} = \mathbb{E}_{D_{te}}\left[\rho_{i,j,k}\right] \tag{4.15}$$

where $\mathbb{E}_{D_{te}}$ is the expectation operator with respect to the testing dataset, and $\epsilon_{i,k}$ is the forecast error of the i_{th} profile for the k_{th} step ahead. The training and testing datasets are divided in 6 folds, for each of which the testing and training days are divided as in Fig. 3.2.
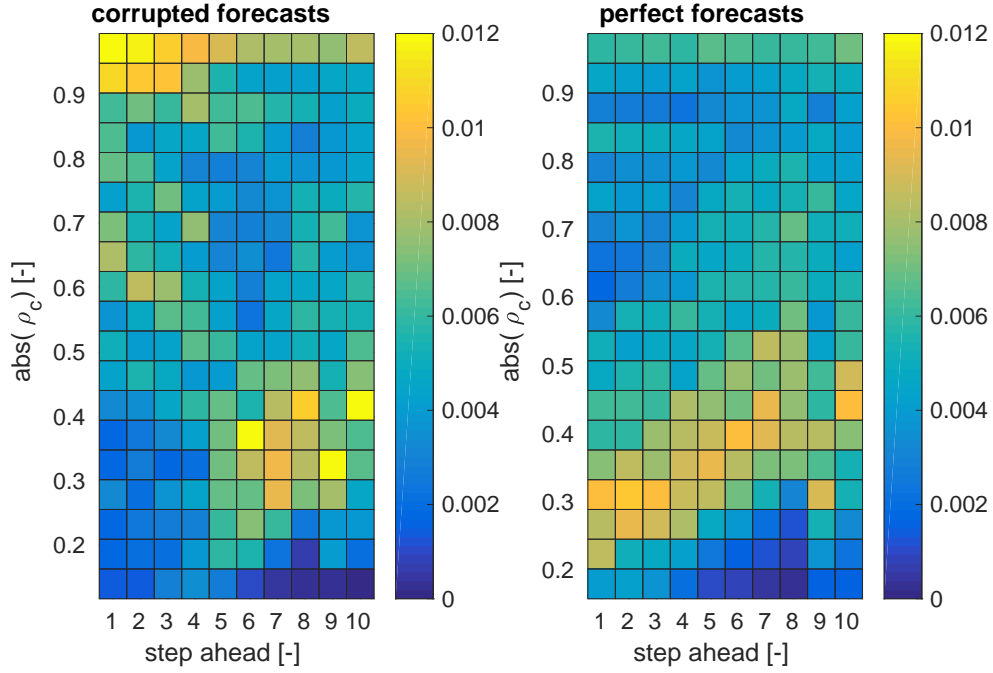
Figure 4.3 – Probability density of the absolute value of the correlation coefficient, as a function of step ahead, for 100 simulated PV profiles. Left: with corrupted forecasts. Right: with perfect forecasts.

Since power signals coming from PV panels are strongly correlated, due to the circadian cycle, we risk to overestimate $\rho_{c,i,j,k}$. To remove this effect, we discard all the forecasts belonging to prediction steps k, if there is at least one observation with a negative elevation of the sun. This means that we discarded prediction steps if they refer to a period of time during which the sun is below the horizon.

$$s_t = \{k|\theta_{el,i} \geq 0 \quad \forall i \in \mathscr{I}_k\} \tag{4.16}$$

where s_t is the selection set at timestep t, with length n_a (in this case $n_a = 10$), $\theta_{el,i}$ is the sun elevation at time i, and $\mathscr{I}_k$ is the set of observations at the original resolution, which are averaged in the k_{th} step. An example of the resulting mask for one day is shown in Fig. 4.2.

The variables used for the forecasts of the signals are shown in table 2.3. Since we do not possess real forecasts for the TMW data which was used for the generation of the synthetic dataset, we corrupted perfect forecasts of GHI and T applying a one hour down-sampling of the signal. As shown in 3.8 this approach is conservative in terms of NWP forecast error.

The result of the correlation analysis for PVs are shown in Fig. 4.3. The left panels shows a density map of the $\rho_{c,i,j,k}$ coefficient, as a function of the step ahead. It can be seen that for the first steps ahead, the forecast errors are highly correlated. Beside knowing that the error are correlated, we would like to know if they are mutually dependent, since in this case, we
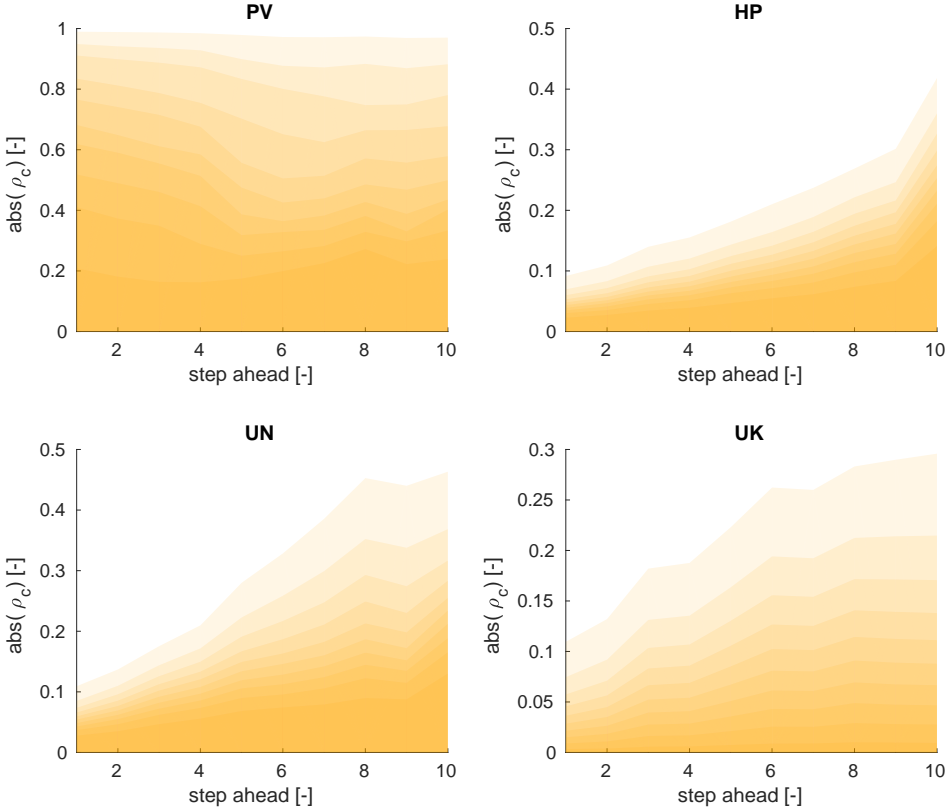
Figure 4.4 – Distribution of the absolute value of the correlation coefficient, by means of quantiles (form 0.05 to 0.95), as a function of the step ahead. Left: 100 simulated heat pump profiles. Center: 100 simulated uncontrolled profiles. Right: 2200 measured mixed power profiles from the UK dataset.

cannot easily sum their pdfs together, as it's explained later. Since correlation doesn't imply dependence in general, we investigated the change of correlation when perfect forecasts for GHI and T are used. The results are shown in the right panel of Fig. 4.3. We can conclude that when we have perfect information about the future GHI and T signals, the forecast errors for the PVs decrease their correlation. This means that in the case of non perfect NWP, the PV forecasts errors are mutually dependent, and the dependence arises due to inaccurate NWP forecasts.

In Fig. 4.4 the same analysis is repeated for the errors coming from the forecasts of the PV profiles, the HP profiles, the uncontrolled ones (UN) and the forecasts' errors for the 2200 signals from the UK dataset, described in the previous section 4.1. Each quadrant shows the density map of ρ_c as a function of step ahead, in terms of quantiles ranging from 0.05 to 0.95. We can see that the correlation of the forecasts' errors for the HP and UN are similar, with increasing correlation for increasing step ahead. The correlation of the errors for the UK dataset are even lower, never exceeding the value of 0.3. This means that for this class of dataset, hierarchical forecasts techniques are expected to increase the accuracy of the bottom level forecasters, as was confirmed in 4.1 and in [126]. However, the same technique will fail in the presence of a high penetration of PV, since the error of the single timeseries won't be smoothed out by aggregation, but on the contrary, the will add up due to their correlation and dependence.

4.2.2 Modeling assumptions for sum of random variables

Different techniques do exist to model sum of random variables starting from marginal pdfs, in the case of uncorrelated and independent random variables. Here we apply them to the forecasts of the single time series $\hat{P}_i$, retrieved in the previous section, and visually compare the resulting a-priori pdf with the a-priori empirical distribution of the forecast error of $\hat{P}_{bu}$, that is, the point forecast of the aggregated profile, obtained with the bottom-up approach. We remark that these a-priori distributions are not the ones we are interested in (we want indeed to estimate the conditional pdfs), but we can use them to quickly assess if the modeling assumptions for the pdf of $\hat{P}_{bu}$ are valid.

In the following the investigated methods for retrieving the pdf of $\hat{P}_{bu}$ are summarized, based on the modeling assumptions.

- **Gaussian uncorrelated profiles**: if we assume that the errors of the forecasters follow Gaussian distributions and are uncorrelated, we can model them as:

$$\epsilon = \mathcal{N}\left(\sum_{i=1}^{N} \mu_i, \left(\sum_{i=1}^{N} \sigma_i^2\right)^{1/2}\right) \tag{4.17}$$

  where μ_i and σ_i are the mean and standard deviation of the forecast errors of the single profiles, respectively, and N is the number of summed profiles. That is, if we group the

prediction errors of each forecaster e_i, in the matrix $E = [e_i]_{i \in \mathbb{N}_{\lfloor 1:N \rfloor}}$, μ_i and σ_i are the mean and standard deviation of the columns of E.

- **Gaussian correlated profiles**: when we further assume that the errors are correlated, we can model ϵ as:

$$\epsilon = \mathcal{N}\left(\sum_{i=1}^{N} \mu_i, \left(\mathbb{1}^T W \mathbb{1} \right)^{1/2} \right) \tag{4.18}$$

where $W = cov(E)$ is the covariance matrix of the forecasts error, and $\mathbb{1}$ is the unitary vector with appropriate dimension. Equation (4.18) can be found by induction, starting from the formula for the sum of two correlated Gaussian distributions.

- **Independent distributions**: if the errors are mutually independent, the pdf of ϵ can be estimated by means of multidimensional convolution of the pdfs of the single signals. As the number of signals N grows, computing the convolution by means of multidimensional integration becomes computationally intense. Instead of computing it directly, we applied the convolution theorem to the Fourier transform of the single signals, and then recovered the final pdf by means of inverse Fourier transform. This method relies on the fact that the convolution of two signals is the product of their Fourier transform:

$$f * g = \mathscr{F}^{-1}\left[\mathscr{F}\left[f \right] \mathscr{F}\left[g \right] \right] \tag{4.19}$$

where $\mathscr{F}$ denotes the Fourier transform and $\mathscr{F}^{-1}$ its inverse. Using (4.19) in combination with FFT results in an efficient algorithm, which can be used for computing the convolution of hundreds of pdfs, as in our case. This method involves padding the single pdfs with zero vectors of appropriate dimensions. More details on this approach can be found in [131].

The results of applying these three methods to the single a-priori pdfs of the errors are reported in Fig. 4.5 for the uncontrolled profiles and in Fig. 4.6 for the PV forecasts. As we can see, even in the case of uncontrolled profile, the assumption of uncorrelated distributions is not realistic. However, the modeling of the a-priori forecast error as a sum of correlated Gaussians seems reasonable, as the pdf obtained with this method is similar to the empirical a-priori pdf of the error of $\hat{P}_{bu}$. On the other hand, when considering $\hat{P}_{bu}$ for 100 PV forecasts, none of the aforementioned modeling approaches results in a pdf similar to the empirical one. As we can see, considering the pdfs as uncorrelated (using both the Gaussian or generic distribution assumptions), results in a way too optimistic pdf for the error of $\hat{P}_{bu}$. On the contrary, considering Gaussian correlated errors results in a too pessimistic pdf, since the PV errors do not follow a Gaussian distribution.
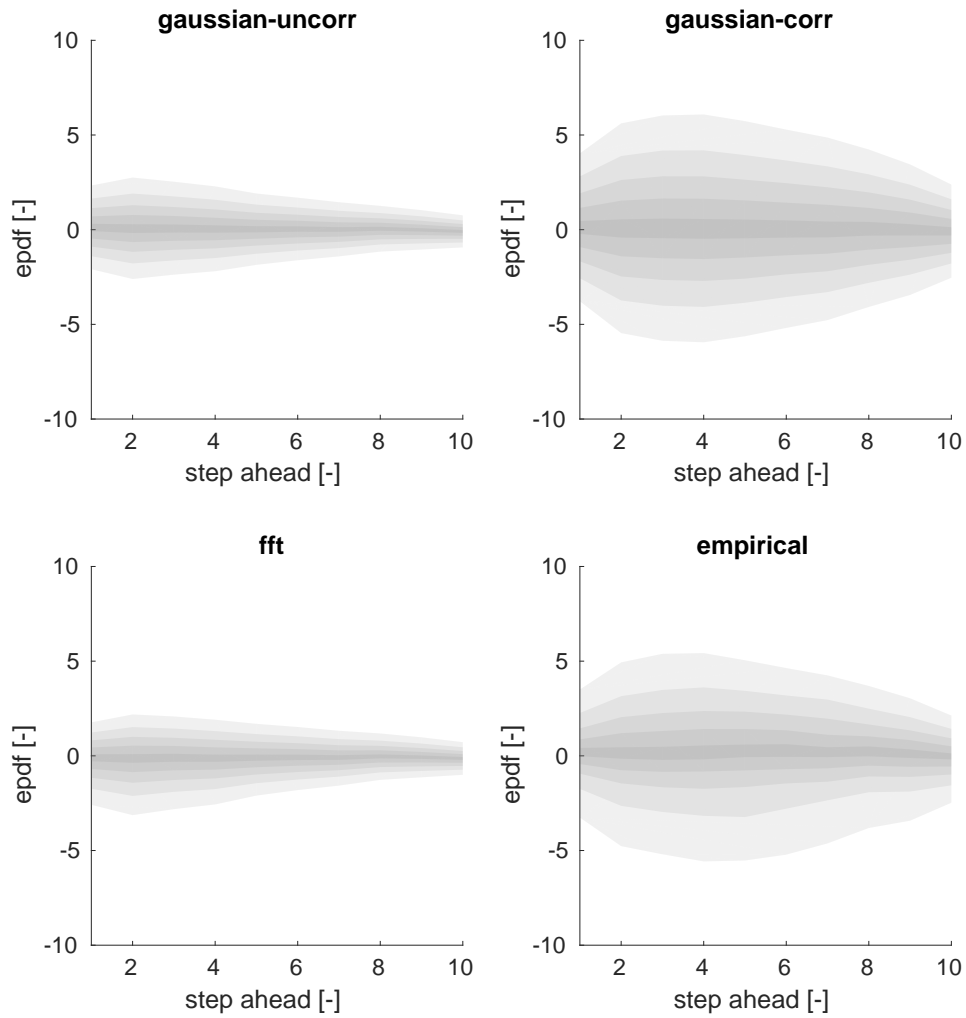
Figure 4.5 – Empirical a-priori pdf of the forecast error of the sum of 100 uncontrolled power profiles, by means of quantile intervals. Upper left: assuming uncorrelated Gaussian errors. Upper right: assuming correlated Gaussian errors. Lower left: convolution of single pdf. Lower right: empirical distribution.
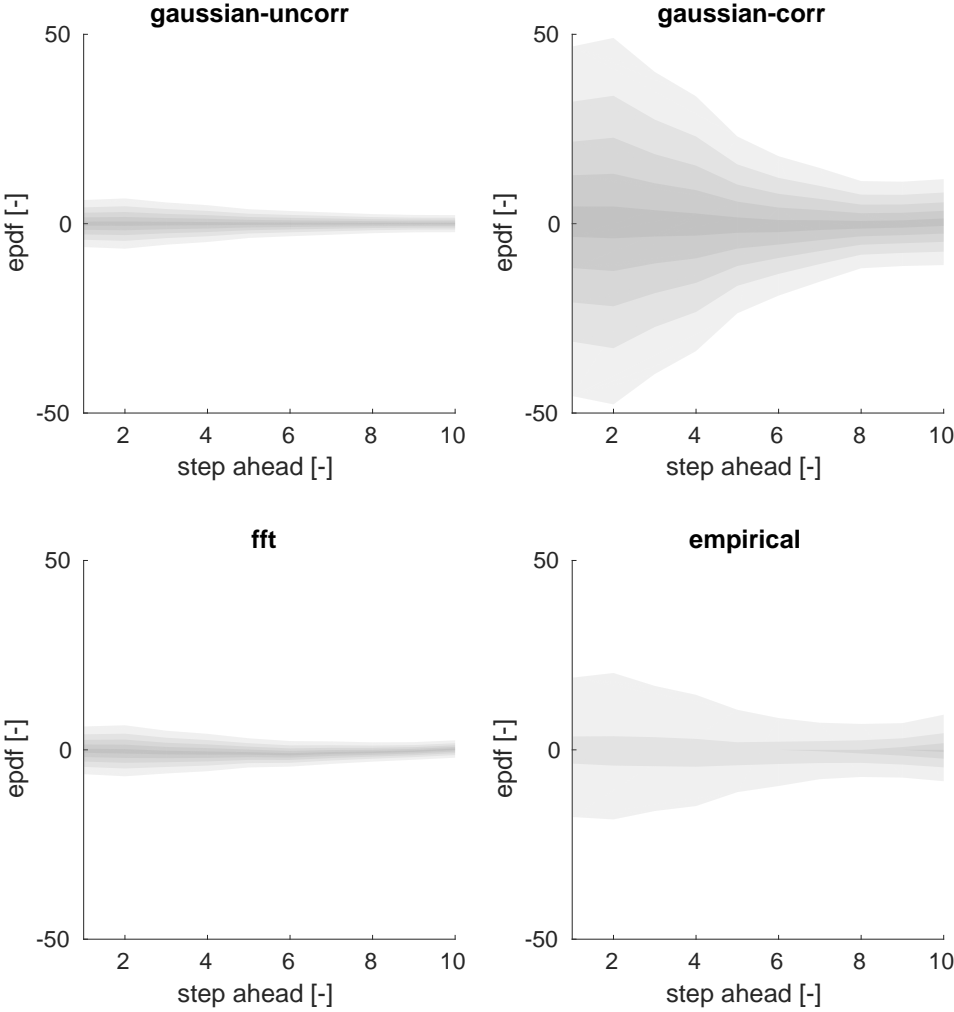
Figure 4.6 – Empirical a-priori pdf of the forecast error of the sum of 100 PV power profiles, by means of quantile intervals. Upper left: assuming uncorrelated Gaussian errors. Upper right: assuming correlated Gaussian errors. Lower left: convolution of single pdf. Lower right: empirical distribution.

4.2.3 Probabilistic hierarchical forecasts avoiding copulas

We now introduce a method for recovering a set of aggregate consistent pdfs without the need of modeling multidimensional (or a great number of) copulas. The rationale of the method, is that we can think of forecasters in terms of deterministic maps $f_i(x_{i,t}, z_t) \rightarrow p\left[y_i | x_{i,t}, z_t\right]$, linking a set of regressors $x_{i,t}$, which only affects the i_{th} base forecaster, and some global variables z_t affecting all the forecasts, to the marginal pdf of the predicted variable y. It is reasonable to think that the correlation (end dependence) of forecast errors of different power profiles, is due to inaccuracy of z_t, which is used by all the forecasters, and not of $x_{i,t}$ which is private information, only exploited by the i_{th} forecaster. If we can model the pdfs of the global variable z_t, we can broadcast this information to all the forecasters f_i, in order to achieve aggregate consistent probabilistic pdfs. In other words, if we know the existence of a variable z which is used by all the forecasters, it is likely that if we provide a consistent description of the pdf of z to all the forecasters, their forecast errors will become independent.

Power forecasts usually make use of few exogenous variables, and in most of the cases they are NWP or historical values of meteorological conditions, for both demand forecasting [132] and PV and wind power forecasting [121]. It is reasonable to think that GHI and the external temperature T, are the only variables which influence all the loads, apart from time of the day and day of the week, which are known in advance without uncertainties. Since as we have seen in the previous section, the forecasts' errors correlation vanishes when the forecasters are provided with exact weather prediction, we assume that for power forecasts the global variable which induce mutual dependence in the forecast errors is the tuple $z = (GHI, T)$. We suggest that, rather than retrieving all the conditional distributions $p(y_i | x_{i,t}, z_t)$ and then sum them by means of empirical copulas, pdfs of power forecasts for arbitrary aggregation structures, consistent by construction, can be obtained passing enumerated scenarios of z drawn from the cdf of the meteorological conditions at current time, $F_{m,t}$, to all the forecasters. This approach seems feasible in practice, since to obtain accurate power forecasts, single users already need to retrieve predictions of meteorological conditions from a NWP provider. This provider could, instead of responding with a point forecast, provide several scenarios of GHI and T. In the experimental trials we have used enumerated scenarios, but the amount of data can be drastically reduced, passing a tree of prediction, as described in section 1.3, and logarithmically spaced steps, since the performance of the controller are negligibly affected by this kind of aggregation, as shown in section 2.2.

Problem formulation

We want to retrieve a tuple of conditional pdf, $\left(p\left[P_i | x_{i,t}, z_t\right], p\left[\sum_{i=1}^{N}(P_i | x_{i,t}, z_t)\right]\right)$. The pseudo-code of the proposed method is provided in Algorithm (1). Line 6 in 1 extracts the s_{th} scenario from the joint conditional pdf $p[z_t] = p[z|t] = p[(GHI, T)|t]$. This is expressed more generally by means of a deterministic function $f_m(x_t)$ which takes as input a set of regressors, and outputs the a set of n_s scenario. In the numerical study, we have used a QRF to generate scenarios, which are linked using the multivariate Gaussian copula coupling method described

in section 1.4. In line 7 of (1), the generated scenario $z_{s,t}$ is then passed as an input to all the power profile forecasters f_i. In general, f_i could return the whole conditional pdf for P_i, but once the uncertainty on GHI and T is removed, this pdf would be very narrow. Moreover, since we expect the errors of f_i to be uncorrelated, the summation process will tend to smooth them out. Finally, in line 8, the output of all the forecasters f_i is summed into $\hat{P}_{bu,t,s}$. Obviously, this bottom-up approach produces aggregate consistent forecasts by construction. The various scenarios for the prediction of the aggregated power profiles are then used in line 9 to retrieve the cdf of $\hat{P}_{bu,t}$, by means of quantiles. Fig. 4.7 visualizes the conceptual steps of the proposed method.

Algorithm 1 Aggregate consistent probabilistic forecasts through scenarios

---

1: n_s : number of scenarios, N: number of power profiles.
2: $x_{m,t}$: set of regressors for meteorological forecaster at current time.
3: $x_{i,t}$: set of regressors for profile i at current time.
4:
5: **for** $s \in \mathbb{N}_{[1:n_s]}$ **do**
6: $\quad f_m(x_{m,t}) \rightarrow \left(GHI_{t,s}, T_{t,s}\right) = z_{t,s}$ $\qquad\qquad\qquad$ ▷ generate GHI,T scenarios
7: $\quad f_i(x_{i,t}, z_{t,s}) \rightarrow \hat{P}_{i,t,s} = \mathbb{E}\left[P_i | x_t, z_{t,s}\right] \quad \forall i \in \mathbb{N}_{[1:N]}$ $\qquad$ ▷ forecast all profiles
8: $\quad \hat{P}_{bu,t,s} = \sum_{i=1}^{n_s} \hat{P}_{i,t,s}$ $\qquad\qquad\qquad\qquad$ ▷ sum forecasts by scenarios
9: $\quad \hat{P}_{bu,t,s} \rightarrow \hat{F}_{bu,t}$ $\qquad\qquad\qquad\qquad\qquad$ ▷ estimate cdf of the sum

---

Table 4.1 reports the regressors used by the meteorological forecaster and by the power forecasters. In particular, the regressors for the meteorological forecaster f_m consist of: the observations of the past 24 hours of GHI and T, reduced using K logarithmic bins; the future values of sun azimuth and elevation, $\theta_{s,az}$ and $\theta_{s,el}$ respectively, similarly reduced; the hour of the day at current time h_t. The power forecasters f_i, use the same set of regressors, plus the set of GHI and T scenarios generated by the meteorological forecaster f_m, and the past data of P_i.

Table 4.1 – Set of regressors for each type of forecaster

forecaster	type	regressors
f_m	QRF	$x_m = \left([GHI_k, T_k]_{k \in \mathbb{N}_{[t-K:t]}}, [\theta_{az,k}, \theta_{el,k}]_{k \in \mathbb{N}_{[t:t+K]}}, h_t\right)$
f_i	QRF	$x_i = \left([P_{i,k}, GHI_k, T_k]_{k \in \mathbb{N}_{[t-K:t]}}, [\theta_{az,k}, \theta_{el,k}]_{k \in \mathbb{N}_{[t:t+K]}}, h_t\right), z_i = [GHI_s, T_s]_{s \in \mathbb{N}_{[1:n_s]}}$

4.2.4 Numerical results

The performance of the proposed method was estimated using one year of data coming from 5 PV power plants located in Lugano, a hilly region in the alpine foothills. The power signals are related to 5 different industrial PV plants, with nominal power ranging from 126 to 275 kWp. Each PV plant has a number of PV fields ranging from 6 to 11. Since the PV field for each plant are similarly oriented, we took only two PV fields per power plant, for a total of 10 signals and 220 kWp. The acquired data refers to a period between 16 November 2016 and 13
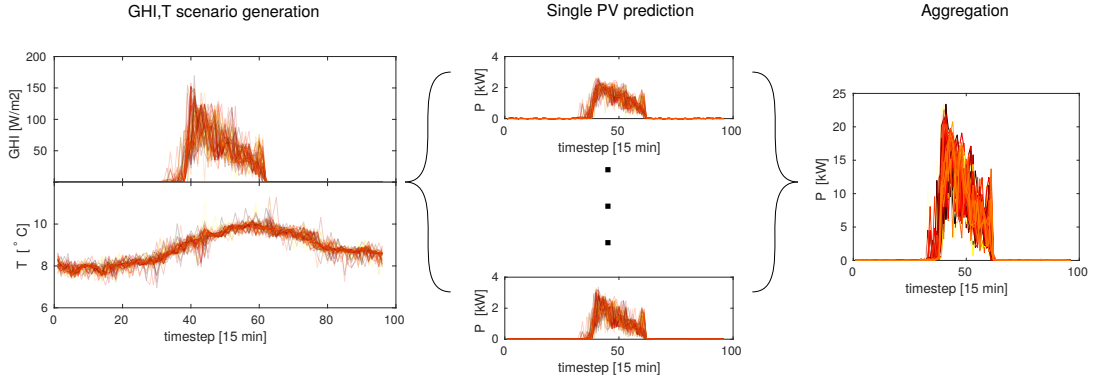
Figure 4.7 – Visualization of the steps of the proposed methods. On the left, the generation of meteorological scenarios by means of f_m is depicted (top:GHI,bottom:T); the central part represents the deterministic output of the power forecasters, based on the set of received scenarios; the right part depicts the deterministic summation of the f_i responses, producing aggregate consistent scenarios, which are then used to retrieve the final pdf.

November 2017, sampled with a 15 minutes resolution. Once again, the method was tested in cross validation using 6 folds.

We compare three methods for the description of $p\left[\sum_{i=1}^{N} P_i\right]$. The first one is the proposed method, described by algorithm 1, which we refer to as model $\mathcal{M}_s$. The second one is the bootstrapped pdf built using errors from the training set, to which we refer to as model $\mathcal{M}_p$. This model uses the method described in 1.4, for which sets of scenarios for the forecast errors are bootstrapped from the errors of the training set, based on the hour of the day.

$$P_{t,s} = \hat{y}_t + [\hat{\epsilon}_h]_{s \in \mathbb{N}_{[0,n_s]}} \quad t \in \left[t_{start}^h, t_{end}^h\right] \tag{4.20}$$

The final bootstrapped scenarios $P_{t,s}$ are then used to estimate $p\left[\sum_{i=1}^{N} P_i\right]$ by means of empirical quantiles.

The third method is the summation of the single conditional pdfs of the base forecasters, by means of convolution. We refer to this method as model $\mathcal{M}_c$. Note that this approach is different from the one used to obtain 4.5 and 4.6, which were retrieved from the convolution of the empirical a-priori pdf of the error. On the contrary, model $\mathcal{M}_c$ performs the multidimensional convolution of the conditional pdfs $p\left[P_i|x_{i,t}, z_t\right]$ at each prediction time t.

In Fig. 4.8 the reliability plot for 10 empirical quantiles, linearly spaced in the interval $[0.05, 0.95]$ is shown, based on the step ahead. The results are mediated over the 6 folds. The reliability of the α quantile for the k_{th} step ahead is defined as:

$$r_{\alpha,k} = \frac{1}{n_{obs}} \sum_{t=1}^{n_{obs}} \mathbb{I}_{[y_{t,k} < \hat{q}_{\alpha,t,k}]} \tag{4.21}$$

which is, the average number of times the observed signal was actually below the predicted α
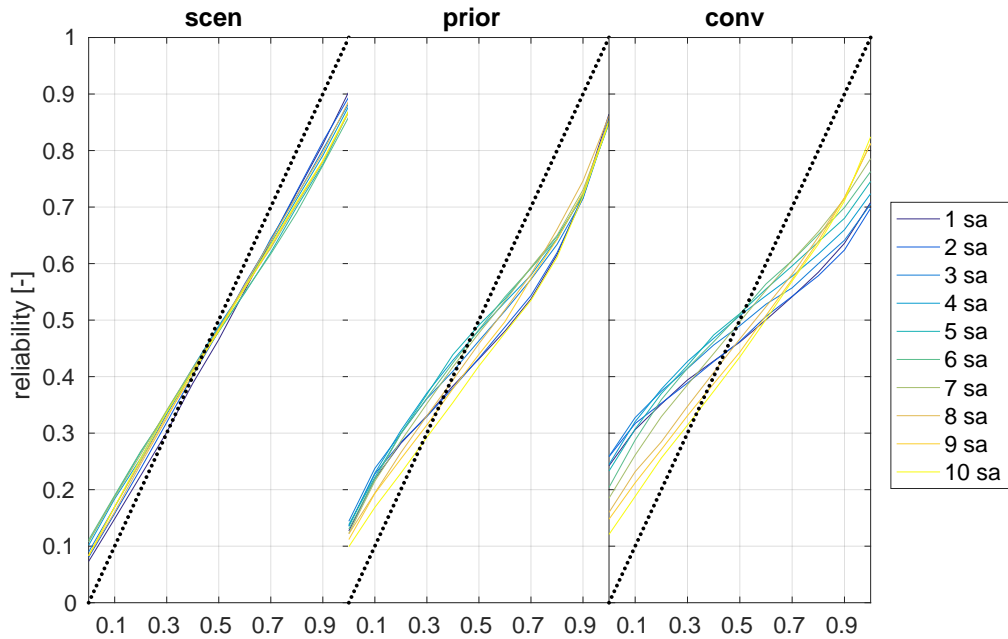
Figure 4.8 – Reliability plots for the tree different methods of forecasting. First column: $\mathcal{M}_s$, second column: $\mathcal{M}_p$, third column: $\mathcal{M}_c$

quantile. We can see that all the methods return a too narrow distribution of the quantiles, overestimating the value of the quantiles below $\alpha = 0.5$, and underestimating for higher α values. However, model $\mathcal{M}_s$ clearly provides the better results. The reliability plot for model $\mathcal{M}_c$, using the convolution of the conditional pdfs, confirms the results obtained using the a-priori pdfs: the hypothesis of error independence results in a too optimistic pdf for the error of the aggregated profile.

In Fig. 4.9, the fold boxplot of different KPIs for $\mathcal{M}_p$ and $\mathcal{M}_c$, normalized with the same KPIs for $\mathcal{M}_s$, are plotted, based on the step ahead. The KPIs are the RMSE, the MAE and the quantile skill score, as introduced in section 2.3.2. We can see how the ratio for the quantile skill score is always positive for all the steps ahead, meaning that the pdf for the aggregated forecasts returned by $\mathcal{M}_s$, are not only aggregate consistent, but also more accurate than the ones returned by the other two models. From the second and third row of Fig.4.9, we can see how the $\mathcal{M}_p$ and $\mathcal{M}_c$ models return a slightly lower RMSE for the first 3 timesteps (corresponding to the first hour ahead of the prediction horizon), while the MAE becomes strictly better starting from 30 min ahead.
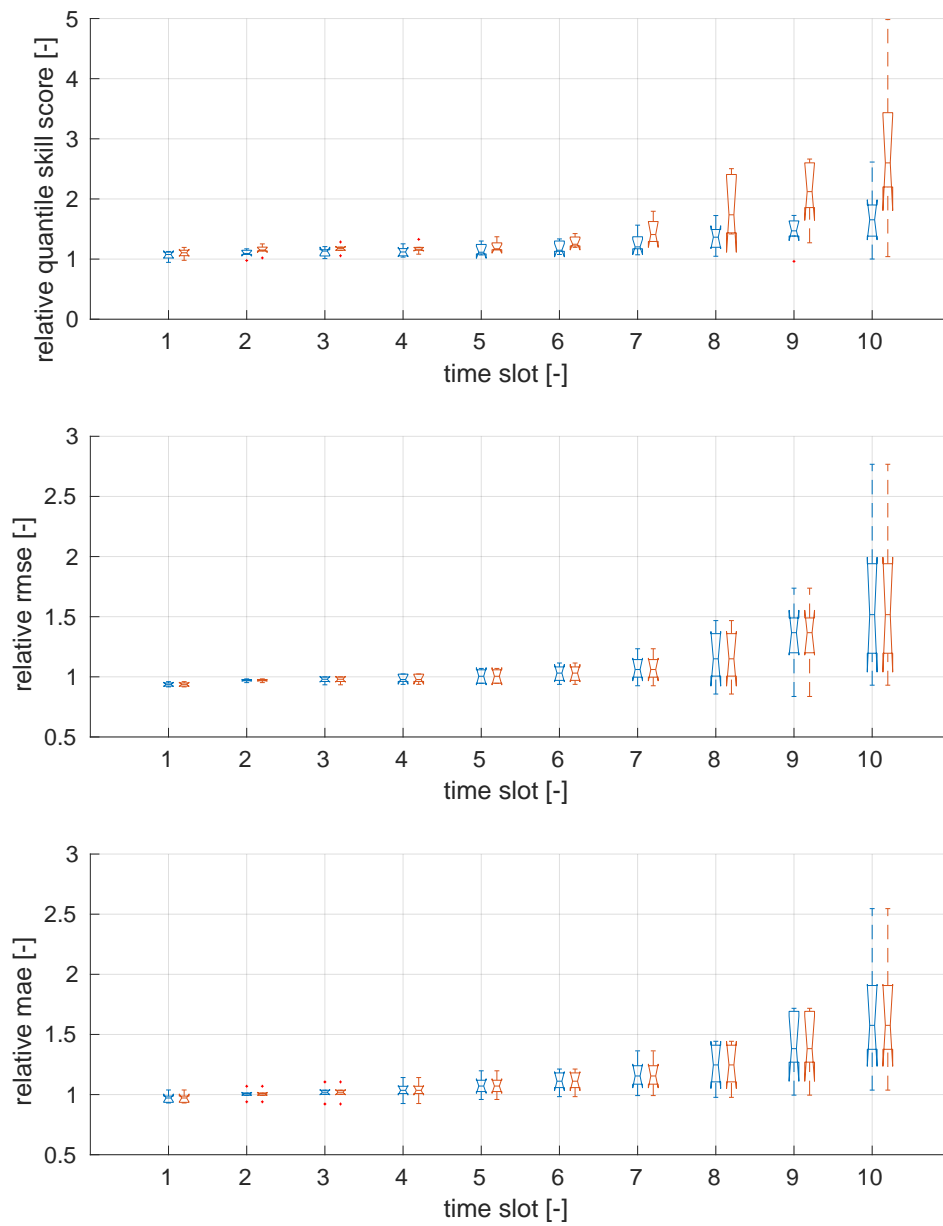
Figure 4.9 – In-folds distributions of the quantile skill scores, RMSE and MAE, for model $\mathcal{M}_p$ (blue) and $\mathcal{M}_c$ (red), normalized with the same values from $\mathcal{M}_q$

.

5 Distributed energy markets

Distributed DSM implies the implementation of a method for multi-agent coordination and poses certain fundamental questions that must be investigated in detail: scalability, user privacy, fault tolerance and resilience against malicious attacks. Scalability ensures that the computational time of coordination scales near-linearly with increasing number of agents, allowing for fast control. Prosumers' privacy is inherently guaranteed if they do not need to share their private information (e.g. size of batteries, desired set-point temperatures in their homes), or their forecasted power profile. Fault tolerance mitigates the effect of any kind of agent misbehaviour, from missing data, to computational errors, to malicious attacks.

In this chapter, I introduce two promising methods for grid-aware multi-agent and multilevel coordination, relying on distributed control and game theory, which can be used to achieve these desirable properties. Both methods rely on a hierarchical structure of aggregators, which reflects the voltage level separations of the electrical grid. In this approach, upper levels of the hierarchy have only access to aggregated information from the underlying levels, allowing scalability and privacy preservation. Moreover, this scheme considers only communication regarding power consumption/production forecasts for the agent, and does not imply sharing private information about one's own system parameters, not even to the trusted party.

The problem formulation and proposed solution strategies are briefly introduced in sections 5.2 and 5.3, while the detailed explanation of the methods and numerical simulations can be found in the annexed papers C and D, respectively. The main outcomes of this chapter are summarized in the following.

- We propose a distributed control algorithm for the coordination of electric batteries, which takes into account a multilevel hierarchical setting, able to cope with competing objective functions.

- We turn the proposed model into a market, which is then analyzed using game theory. We demonstrate that, when the original problem belongs to the class of so called sharing problem, the resulting game has a unique Nash equilibrium, which can be reached

through a trivial modification of the algorithm used for solving the original problem. We then compare this algorithm with the recently presented preconditioned Forward Backward (pFB) algorithm [133], and show that the proposed one converges in a lower number of iterations.

5.1 From distributed OPF to distributed energy markets - state of the art

The mathematical problem of coordinating distributed agents in order to achieve a desired global effect is pretty general, and different approaches can be found in the literature for its application to the electrical grid. We will generally refer to this class of problem as distributed optimal power flow (D-OPF). This can be described as a networked optimization problem, in which each agent can only communicate with their neighbors, which are defined by a communication graph [134]. When dealing with D-OPF, grid constraints must be decomposed among the different agents, and we refer to them as coupling constraints. Various approaches exist to solve them in a distributed way [135; 136; 137]. In [25] a non-exhaustive review on D-OPF methods can be found. A non-comprehensive review on architectures for distributed MPC can be found in [138]. In [139] several papers on distributed MPC, including deterministic and robust approaches, are cataloged based on process commonalities, control architecture and theoretical properties. When applied to power systems, most of the studies are focused on a single level decomposition [140; 141; 142], where agents can communicate with a coordinator. Fewer works are focused on completely decentralized protocols. Notably, in [143], a method to solve the D-OPF through ADMM on arbitrary graphs is presented, which is referred by the authors as proximal message passing. Methods found in the D-OPF literature for coordination of distributed agents are usually based on some decomposition techniques, such as ADMM, primal-dual Douglas-Rachford splitting [140], which is equivalent to ADMM (cfr. [61] §4.5), Dantzig-Wolfe decomposition [144; 145], proximal minimization [146]. These methods have been historically developed in different contexts, and multiple interpretations can be given for each algorithm (see for example [61] for interpretations of proximal algorithms), but they can all be derived using monotone operator theory. See for example [62] on this topic.

Despite the abundant literature on distributed control, the ongoing paradigm shift from centralized to decentralized generation poses new challenges. When power generators belong to a single entity, which is usually the assumption in D-OPF, the redistribution of revenues among competing entities does not have to be taken into account in the problem decomposition by the independent system operator (ISO). This is not the case for DSM in general, in which prosumers' actions are motivated by their own utility. In this setting, proper regulatory mechanisms defining rules for remunerations and compensations need to be set up to allow the integration of DSM services in the wholesale market. Such rules need to address the problem of conflicting interests between market participants.

Additionally, a growing number of publications [13; 15; 14] highlights the possibility for some

agents to alter the distributed optimization mechanism in their favour, to the detriment of the welfare of the other agents, by just lying about their predictions. This setting requires to treat the problem decomposition in a game theoretic way, shifting the focus from distributed control to market design. Solving the D-OPF can be described in terms of welfare maximization (WM) problem, in which the sum of costs and revenues faced by generators is maximized. On the contrary, mechanism design literature focuses on the Nash equilibrium (NE) concept, which analyzes the distributed problem from the point of view of the single prosumers. The solution under NE concepts will in general differ from the un-weighted WM equilibrium, leading to the so called price of anarchy [134], which is the overall loss in welfare (social cost) while passing from the WM equilibrium to NE. See for example [147] on this topic in the case of DSM.

In Table 5.1 and 5.2 selected papers on distributed control literature are presented, based on the adopted concept of equilibrium and uncertainty, and on the decomposition technique, respectively.

Table 5.1 – Distributed control methods for OPF and DSM based on uncertainty formulation and equilibrium formulation.

			Uncertainty			
			Deterministic	Robust	CC	ST/S
Equilibrium	WM		[143; 140; 148; 149; 150; 151; 152; 141; 153; 154]	[155] [142]	[156]	[146]
	NE		[157; 158; 159]	[160]	-	-
	GNE/VNE		[133; 161; 162; 63; 64]	-	-	-

Table 5.2 – Distributed control methods for OPF and DSM based on decomposition method.

Decomposition	ADMM	[143; 140; 142; 148; 149; 141; 153; 154]
	Douglas-Rachford splitting	[140]
	Saddle point dynamics	[155]
	Nikaido-Isoda funcion	[157]
	Proximal minimization	[146]
	Lagrangian relaxation	[150; 151]
	C+I	[152; 163]
	pFB/APA	[64; 133; 161; 162]
	FBF	[63]
	Gradient descent	[159]
	Dantzig–Wolfe Decomposition	[145]

5.1.1 Trustless coordination mechanisms

Different authors have designed market rules using a game theoretic framework, describing the problem as a non-cooperative game. The underling concept in game theory and mechanism design is that, if the right amount and type of incentives are given, rational prosumers will decide to cooperate. In the following, we briefly introduce fundamentals concepts of mechanism design, which are needed to understand the ideas behind different kinds of Nash equilibriums. Usually, mechanism design implies the redistribution of a scarce resource among the agents. We slightly simplify the presented concepts, considering that in our case, we assume energy needs of agents are always satisfied, that is, we do not consider islanded portions of the grid.

In a centralized setting, the ISO is assumed to know all the relevant information in order to optimally solve the intended problem. This consists in minimizing a function $f : \Theta \to U$, where $\Theta = \prod_{i=1}^{N} \Theta_i$ is the Cartesian product of the space of private information of all the N controlled agents, and $U = \prod_{i=1}^{N} U_i$ is the space of economic outcomes for all the agents, usually referred to as utility. Function f is also called the social choice function.
On the other hand, when we decentralize the problem, each agent will follow its own strategy $\sigma : \Theta_i \to X_i$, which is a function of his own private information only, and will return its decisions $x_i \in X_i$ to the ISO, where X_i is the feasible space of actions for agent i. These are then retrieved by the ISO, and mapped into agents' costs by means of the so called outcome function $g : X \to U$, where $X \in \prod_{i=1}^{N} X_i$ is the joint action space of all the agents.
The purpose of mechanism design is to build an outcome function g such that a strategy σ^* exists, which, if followed by the agents, provides them the maximum economic benefit, and at the same time, minimizes the original social function f. In this way, there will be no reasons for the agents for not following the suggested strategy $\sigma*$, since doing so will guarantee them the maximum profits.

If agents think that the outcome function g is such that true telling is a *dominant strategy* (i.e., the strategy that maximizes an agent's utility is independent from other's agent strategies), they will chose to truthfully report their types θ_i to the aggregator, without any strategic manipulations. In this case the mechanism is called direct, since the aggregator will be able to directly implement the social choice function f, solving the associated problems. If true telling is a dominant strategy, the mechanism is said to be dominant strategy *incentive compatible* or *strategy-proof*. Formally, a strategy-proof mechanism ensures that:

$$u_i(f(\theta_i, \theta_{-i})) \geq u_i(f(\hat{\theta}_i, \theta_{-i})) \quad \forall i \in N \tag{5.1}$$

where $u_i(\lambda_i)$ is the utility of the i_{th} agent, as a function of the price chosen by the mechanism, θ_i denotes the true private information, while $\hat{\theta}_i$ denotes a manipulated declaration of θ_i. Note that f is used instead of $g \circ \sigma$ is used, since the mechanism can be directly implemented by the ISO. Most of the studies historically focused on the class of direct mechanisms implementable in a centralized way, due to the fact that one of the most celebrated outcomes of mechanism

design, the *revelation principle*, states that any implementable social choice function can be implemented by a direct mechanism. Belonging to this class are the well known Vickrey-Clarke-Groves (VCG) mechanisms [164; 165; 166] . The VCG has many teoretical theoretical properties, among which being weakly budget-balanced and strategy-proof. The strategy-proofness is achieved by the mechanism paying each agent the sum of the value of all other agents. In this way, the social choice coincides with the interest of each agent, but at the same time it makes the mechanism extremely expensive for the aggregator, which should pay each agent the total amount of money that it would have paid in a trusted setting. This problem is alleviated using the Clarke pivot rule, with a resulting total value for the i_{th} agent equal to:

$$u'_i(x) = \max_{x \in \mathscr{X}_i} \sum_{i=1}^{N} u_i(x) - \max_{x_{-i} \in \mathscr{X}_{-i}} \sum_{j \neq i} u_j(x) \tag{5.2}$$

It is clear that since the second term does not directly nor *indirectly* depend on any decision of i, the only way to influence its total value is influencing the first term. Since maximizing the first term is equal to maximizing the social welfare, the mechanism implements social welfare maximization and is strategy-proof. Anyway, the requirement that the second term has to be independent from agent i implies that N optimization problems must be solved, each of which is performed without considering a given agent, which is a major drawback when the total number of agents is big. Since the computational cost of solving a single problem is expected to scale linearly with the number of users, the overall computational cost will scale as $\mathcal{O}(n^2)$. Moreover, when trying to distribute strategy-proof mechanisms as VCG, dominant strategy equilibrium can not be guaranteed anymore [167]. This effect is also known as the cost of decentralization. The strongest notion of equilibrium when decentralizing a mechanism is known as ex-post Nash equilibrium. Informally, this equilibrium guarantees that the suggested strategy $\sigma^*(\theta_i)$ is the best strategy when all the other agents follow it. Formally:

$$u_i\left(g(\sigma^*(\theta_i), \sigma^*_{-i}(\theta_{-i}))\right) \geq u_i\left(g(\sigma'(\theta_i), \sigma^*_{-i}(\theta_{-i}))\right) \tag{5.3}$$

Authoritative work on decentralized mechanism have been carried out in the contest of algorithmic game theory [167]. The ex-post implementations of VCG has been theoretically studied in [168], where design principles are presented. Notable work has been done in [169] trying to reuse parts of the redundant computation of distributed VCGs mechanisms. In [170] a Cournot model is considered for dynamic bidding in the electricity market. In [171] a VCG mechanism is adopted to model a virtual inertia market.

None of the studies that have been identified in our literature review covers the design of a non-cooperative game, taking into account coupling constraints (e.g. grid constraints) applied to energy markets [172; 173; 63; 174]. A way of considering coupling constraints in games comes from the seminal work of [175] on n-person non cooperative games. In this work, conditions for the existence and uniqueness of the NE are given in terms of convexity of the considered game map. For instance, this work has been used in [176], applied to the problem f bidirectional energy trading between residential loads ad electric vehicles. The

concept has been lately reconsidered, showing the existence of a new class of equilibrium refinement, called Variational NE (VNE) [177], which guarantees the same marginal prices for all the users at equilibrium. In [64] conditions are given for the uniqueness of a VNE in terms of monotonicity of the game map, in presence of coupling constraints. A similar description using monotone operators is provided in [133]. We have used this solution concept in D, in order to demonstrate convergence to a VNE using ADMM.

5.2 Multilevel hierarchical control

As previously stated, the flexibility of prosumers can be exploited providing multiple services to the electrical grid, for example secondary control and energy management. Due to the electrical grid interconnections, in order to provide these services safely, an optimization method taking into account their effects on multiple voltage levels should be used. Moreover, coordinating groups of prosumers is always better than having prosumers to follow individual policies, in terms of aggregated effect. For example, let us consider the case in which we want to minimize a given objective function in terms of the aggregated power profile of a group of prosumers connected at the same LV transformer. Additionally, suppose that we require the power at the PCC to be symmetrically bounded, that is, $Sx \in \left[-\overline{P}, \overline{P}\right]$, where with $S = \mathbb{1}_N^T \otimes I_T$ we denoted the time summation matrix of all the prosumers' actions. In order to guarantee that this constraint is never violated using individual policies, we must replace it with the set of prosumers' constraints $x_i \in \left[-\overline{P}/N, \overline{P}/N\right]$. On the other hand, the feasible space of prosumers' actions would increase when allowing communication, thus leading to a better solution. Formally:

Lemma 5.2.1. *The n-dimensional hypercube intersecting each axis in $\{-\overline{P}/N, \overline{P}/N\}$ is always included in the convex hull defined by the points $\left[e_i(-\overline{P}, \overline{P})\right]_{i=1}^{N}$, where e_i is the versor of the ith dimension.*

Proof. The maximum distance of the center from the n-dimensional hypercube defined in 5.2.1, equal to half of its diagonal, is $\overline{P}/\sqrt{N}$, which equal to the distance of the hyperplanes defined in 5.2.1 form the center, in each orthant. $\square$

This means that constraining the aggregate instead of the single users will enlarge the feasible set $\mathscr{X}$, always resulting in a (not strictly) better solution.

The concept of hierarchical control of power distribution system is not new [178], as well for DSM [179]. In the literature, different methods have been proposed to coordinate groups of prosumers by means of an aggregator. Most of them refer to a single level hierarchy, as in [142; 145]. Both [141; 180] consider distributed MPC for a single level hierarchy, without grid constraints. In [181] a three level hierarchy has been proposed for the control of reactive power from PV inverters. The adopted sequential strategy does not explicitly take into account grid constraints and is solved through particle swarm optimization heuristic. In paper C,
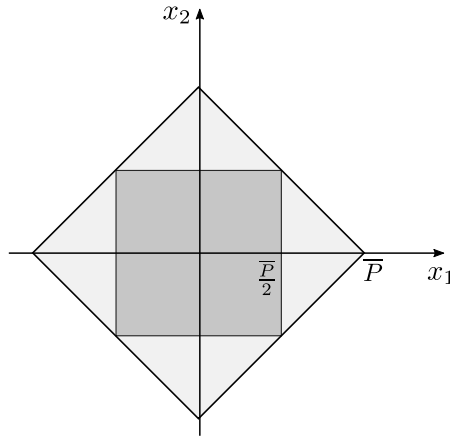
Figure 5.1 – Feasible sets for the space of two prosumers' actions, for the constraint $|x_1 + x_2| < \overline{P}$ in the case of individual policies (dark gray) and in the case of communication (light gray). Communication enlarges the feasible set, thus potentially improving the solution.

we present a multilevel hierarchical algorithm for the coordination of prosumers located in different voltage levels of the electrical grid. The hierarchical structure of the grid is described by means of a rooted tree. The coordination of prosumers is made possible by aggregators located at branching nodes of the tree. It is not necessary that the levels of the hierarchical structure coincide with the grid voltage levels. For instance, the first level of aggregation could be done at building level, while the second level could be placed at the LV cabinets. Formally, given N prosumers with a controllable battery, located in a hierarchical structure, we solve the so called sharing problem [182]:

$$\underset{x}{\text{argmin}}\, e(S_\emptyset x) + \sum_{i=1}^{N} f_{c,i}(x_i)$$

$$s.t.: S_B x \le v_B \quad \forall\, B \in \mathcal{B}(\tau) \tag{5.4}$$

where $x_i \in \mathbb{R}^T$ are the overall charging or discharging power from the i_{th} prosumer, $x = [x_i]_{i=1}^{N}$ is the vector containing the overall battery operations from all the prosumers, $S_\emptyset = \mathbb{1}_N^T \otimes \mathbb{I}_T$ is a summation matrix summing all the x_i. Here S_B are similar summation matrix, each of which encodes a constraint in the set of branching points of the hierarchical structure, $\mathcal{B}(\tau)$. The private functions $f_{c,i}(x_i) : \mathbb{R}^T \to \overline{\mathbb{R}}$ are extended real value functions, encoding both prosumers' objectives and constraints, including battery dynamics. As explained in the next section, the inequalities in (5.4), can be used to encode grid constraints. Despite having a simple formulation, problem (5.4) is very flexible, allowing having multiple objectives in different branches, or appliances providing possibly competing services to the grid, like primary regulation and dispatchable operations.

5.2.1 Respecting grid constraints in grids with unknown topology

Different relaxations of power flow equation exist, which can be used when solving D-OPF [25]. Usually, knowledge of phasors' angles (e.g. DC approximation), or knowledge of the lines' parameters and topology (e.g. the DistFlow model) are required inputs to these approximation. However, these information are not always available. Residential users, for example, are located in the LV grid, for which the grid topology is usually unknown or hard to retrieve. In this case we should opt for an approximated formulation of the power flow, whose parameters can be estimated by means of smart meter data. One of these formulation consists in the first order linearization of the power flow equations. The linear coefficients of this formulation are known as the voltage sensitivity coefficients:

$$
\begin{aligned}
k^p_{i,j} &= \frac{\partial |V_j|}{\partial P_k} \\
k^q_{i,j} &= \frac{\partial |V_j|}{\partial Q_k}
\end{aligned}
\tag{5.5}
$$

where P and Q are the active and reactive power, respectively, and $k^p_{i,j}, k^q_{i,j}$ are the sensitivity coefficients between node i and node j. The analytical expression of voltage sensitivity coefficients, and an efficient method to compute them based on the state of the grid and admittance matrix, is provided in [183]. In [184], it has been shown that the voltage sensitivity coefficients can be estimated by least squares regression of the time derivatives of voltage magnitudes, P and Q, such that:

$$
\Delta V = \Delta (P, Q) K + \epsilon
\tag{5.6}
$$

where $\Delta V \in \mathbb{R}^{T-1 \times 3N}$ is the matrix containing the time derivatives of the voltage magnitudes for the N nodes of the (three-phase) grid, over T timesteps, $K \in \mathbb{R}^{6N \times 3N}$ is the matrix containing the voltage sensitivity coefficients, and ϵ is a random variable with the same dimension of ΔV representing the model error. Moreover, in [185] it has been reported how estimating voltages using (5.6) could lead to lower errors with respect to the case in which the grid is modeled using a white-box approach, but having inaccurate parameters information (e.g. uncertainty in the length of the cables). As previously reported in chapter 1, voltage sensitivity coefficients can be used to optimize a relaxation of the OPF:

$$
\begin{aligned}
&\min_{P_c, Q_c} f(P_c) \\
&s.t. \ \ |V| = V_0 + K_p \Delta P + K_q \Delta Q \\
&\qquad |V| \in \mathcal{V}, \quad P \in \mathcal{P}, \quad Q \in \mathcal{Q}
\end{aligned}
\tag{5.7}
$$

where $K_p \in \mathbb{R}^{3N \times 3N}$ and $K_q \in \mathbb{R}^{3N \times 3N}$ are the voltage sensitivity matrices related to active and reactive power and P_c and Q_c are the vectors of active and reactive powers in the controlled
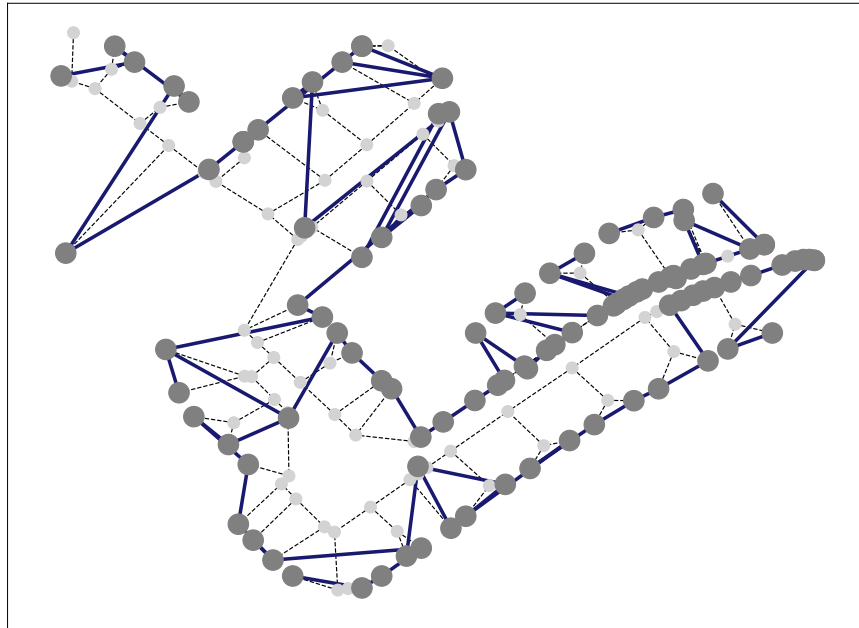
Figure 5.2 – Example of topology identification using the IEEE European Low Voltage Test Feeder. Small light gray dots represents the actual nodes in the grid. Big dark gray dots represents PQ nodes where measurements are available. Dashed lines are the actual physical connections, while dark blue lines are the identified connections.

nodes. It is worth mentioning that, since the voltage sensitivity coefficients are a linearization of the power flow equations, K depends on the state of the grid V. For a particular grid, this dependency can be estimated empirically, testing if the approximation of $|V|$ in (5.7) is accurate enough under changing values of V. Alternatively, different sets of the rows of matrices K_q and K_p can be identified, based on the voltage magnitude range of the associated node, leading to the control of a piecewise affine linear system. In order to define the constraint set $\mathcal{P}$ and $\mathcal{Q}$, it could be helpful to retrieve an approximated topology (connectivity) of the various nodes. In this way we can know which group of nodes can influence the total power in a given point of the distribution grid. In the case of LV grids, which are typically radial, this task can be carried out using only smart meter measurements. In particular, only covariance matrix of the voltages' magnitude deviations is needed to infer the structure of the grid, through a minimum spanning tree [186]. The identification is also possible in the case of missing nodes, as described in [187]. In Fig. 5.2, an example of topology identification for the IEEE European Low Voltage Test Feeder [188] with the algorithm presented in [186] is shown. Each PQ node of the Test Feeder (dark gray dots in the figure) has been populated with synthetically generated residential power profiles. The identification is carried out using one day data of the three

phases voltage magnitudes of the PQ nodes. It is clear that the accuracy of the identified topology strictly depends on the available points of measurements in the grid. In particular, in the case in which we are only in possess of smart meter data as in the case shown in Fig. 5.2, the topology identification can just be used to divide users in groups and to assess their approximated distance between each other and from the PCC. On the other hand, more useful results can be retrieved if measurements from cabinets and transformers can be used.

5.2.2   Main results from paper C

Paper C introduces an algorithm for the coordination of prosumers organized in hierarchical structure. The aim of the paper is to exploit a hierarchical communication topology in order to obtain a scalable algorithm, while preserving prosumers' privacy and increase resiliency through system redundancy. The main results from paper C are summarized in the following.

- System redundancy. The proposed decomposition alleviates the problem of having a single point of failure, typical of one-level hierarchical control. Since all branching nodes perform the same kind of computations, if the root node goes offline, its children can inherit its role or revert to a predefined objective function (e.g. peak shaving). This introduces redundancy and resiliency in the system.

- Prosumers' privacy. In addition to the level of privacy guaranteed by decomposition methods, that allow prosumers' to keep their own parameters and constraints private, the suggested method prevents higher levels of the hierarchy to directly know prosumers' power forecasts and realizations. In fact this information is only used by each prosumers' parent, and is available to higher branches only by means of aggregations.

- Computational time. A trivial way of decomposing the problem would consist in reaching convergence in the lower level, then sending the results to the upper level of the hierarchy. In this way, the whole number of iteration before convergence in the lower level would represent a single iteration in the upper level. This solution concept would clearly result in an exponential computational time with respect to the number of considered levels, namely:

$$N_{tot} \sim tnN_i^L \prod_{l=1}^{L} N_b^l \sim N_i^L N_b^{\frac{L+L^2}{2}} \tag{5.8}$$

where t is the computational time for each agent for solving its local problem, n is the number of agents per level, N_i is the number of iterations before convergence for a single level, L is the number of levels, and N_b is the number of branches per level. Instead of following this strategy, we decompose the monolithic formulation of the problem to obtain nearly linear convergence, with respect to the number of levels. This is shown in Fig. 5.3, where the computational time before convergence from 500 simulations is shown as a function of the number of levels. The boxplots refer to randomly generated
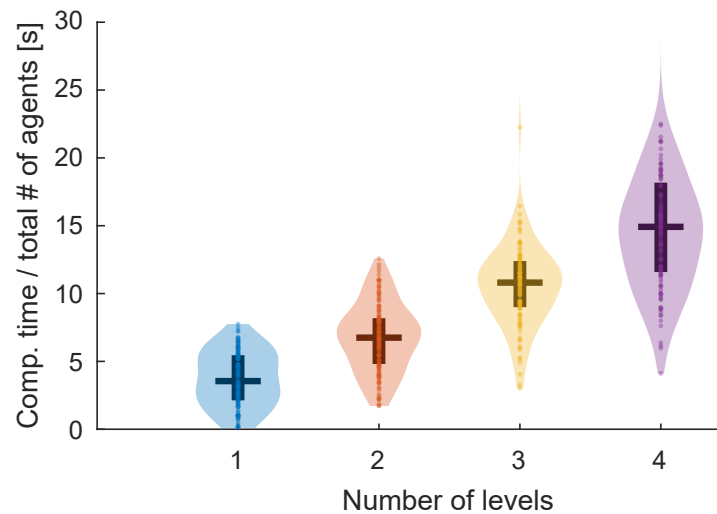
Figure 5.3 – Estimated pdfs of the computational time divided by the total number of agents, as a function of the number of levels in the hierarchy. The vertical bar is the interquartile range, the horizontal line is the median.

tree structures with at most 4 aggregation levels, a branching factor of 2 and a maximum of ten agents for each branching node, and agents are represented by controllable batteries.

5.3 Coordination in a trustless setting

As anticipated in section 5.1, the exploitation of DSM requires to provide users with rules for remuneration and compensation of their flexibility. This means that a market for flexibility must be created. This is a fundamentally different scenario, as opposed to the D-OPF or UC problems, in which the units to be coordinated belong to a single entity. In the case of coordination of prosumers, we must consider the case in which single agent will try to exploit weaknesses in the market design or directly try to hack their smart meters. In the following, the reasons that make these behaviours likely are summarized.

- High number of agents with conflicting economical interests

- Possible interactions of different technologies (smart meters/ algorithms)

- Cheating only requires to trivially modify the distributed control algorithm [13; 14; 15].

For these reasons, I decided to treat the distributed coordination of prosumers using algorithmic game theory [167] and distributed mechanism design [189]. In paper D, the sharing problem 5.4, for a single level hierarchy is considered. I focused on a single use case, in which

the energy market is deployed for an energy sharing community (also known as self consumption community). These communities, which are made possible by law in Switzerland from January 2018 [190], are part of the grid, which is rented or privately built, which are billed only at the point of common coupling with the main grid. This means that members of these communities do not pay individually, but they are billed as a single entity. This pushes the members of the community to increase their self consumption through coordination. The possibility of aggregating consumption and production behind the meter represents an improvement in counteracting the volatility of locally installed DERs, with respect to the existing indirect control embodied by asymmetric buying and selling prices. For this reason, in this case, we considered the following system-level objective $e(x)$:

$$e(x) = c\left(\sum_{i=1}^{N} x_i\right) - \sum_{i=1}^{N} c(x_i) \tag{5.9}$$

where $x_i \in \mathrm{IR}^{\mathrm{T}}$ is the vector of total power of the ith agent, $c(\cdot)$ is the energy cost function defined in 2.37. Function $e(x)$ represents the difference between the energy cost paid by the community at the PCC and the sum of the cost that the prosumers would have paid individually in the base case. This sum is always negative in the presence of an heterogeneous mix of producers and consumers, and vanishes in case of an homogeneous mix (i.e. all the users are consuming or all users are producing). As such, 5.9 represents the economic (negative) surplus generated by the community with respect to the base case. Minimizing $e(x)$ indirectly minimizes the total energy exchanged at the PCC. In a welfare maximization setting (WM), problem 5.4 can be decomposed, and the generated surplus redistributed among the users using a metric of choice. In a trustless setting, the redistribution rule must be declared in advance, so that prosumers can decide their actions based on their projected economic benefits. The rule of repartition, along with the cost function $c(\cdot)$, constitutes the outcome function g, introduced in 5.1.1.

5.3.1 Main results from paper D

Paper D introduces an energy market concept, enabling to optimally exploit the flexiblity of prosumers, while respecting grid constraints. Users are assumed not to have any active decision nor belief in the formation of the market equilibrium price. Instead, the well known concept of Lagrangian relaxation form optimization theory is used to automatically identify prices for the grid's constraints. I further avoided to model users' utility as intended in standard auction and game theory, and replaced it with costs and users' constraints sets. In fact, the latter can be interpreted as a binary and non differentiable utility function, and prevent us from making any assumption on users' marginal satisfaction with respect to consumed energy. The main outcomes from paper D are summarized in the following.

- Existence and uniqueness of a VGNE for the class of sharing problems. In D, it is proven that decomposing 5.4 using different repartition weights for the surplus, induces a

game with unique generalized variational equilibrium, which can be reached jointly minimizing the utility function of the agents, given by:

$$u(x_i, x_{-i}) = c_i(x_i) + \alpha_i e(x) + A_i^T \lambda \qquad (5.10)$$

where α is a repartition coefficient for prosumer i, $A_i \in \mathbb{R}^{T \times T}$, T being the optimization horizon, with $S_B = \left[A_i^T \right]_{i=1}^{N}$, N is the number of prosumers, and λ is a vector of Lagrangian multipliers associated with grid constraints.

- Enforcement of individual rationality. Lagrangian multipliers are not bounded a-priori; this could lead to a situation in which the prosumers pay more than in the base case. In paper D, I proposed to limit the increment of the Lagrangian multipliers to guarantee the individual rationality (IR) constraint. IR requires that prosumers won't face higher energy bills when opting in the proposed energy market. Limiting the Lagrangian reflects a natural economic reasoning, which prevents prosumers from providing their flexibility for free at the only scope of balancing voltage and power fluctuations in the grid. In Fig. 5.4, an example of solution is shown in terms of prosumer's battery state of charge, aggregated power profiles and voltage at the PCC. The results for the WM problem and for the associated game are reported, using the proposed formulation and the pBF algorithm. In this case, grid constraints, represented by the dashed red lines, are respected. This could not always be the case, for two reasons. The first is that the prosumers' flexibility could be in any case not enough to keep power and voltages inside the constraints set. For instance, this could depend on the total number of agents which are willing to participate to the proposed energy sharing community, which could be low with respect to the total number of users. Secondly, as anticipated, the hard cap on the Lagrangian multipliers enforcing individual rationality could prevent prosumers from using all their batteries' flexibility for grid regulation.

- Comparison with the preconditioned forward backward (pFB) algorithm. We compared the proposed algorithm with the pFB. The work presented in [133] states that all the projected-gradient algorithm class, when applied to aggregative games, can be interpreted as pFB. The comparison was done by means of iterations before convergence, and showed that the ADMM based formulation requires on average fewer iterations to converge.
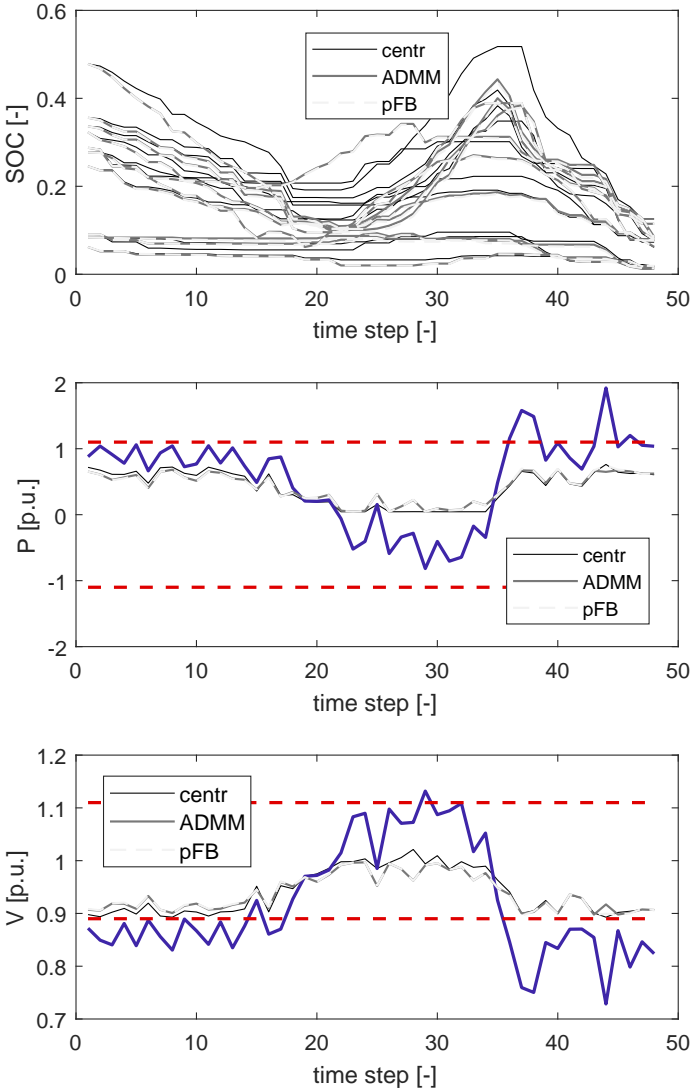
Figure 5.4 – Time series example, N = 10. Blue: forecasted profiles. Red: constraints. Grays: solutions of the centralized and decentralized approaches. Top: state of charge for each battery. Middle: power profiles. Bottom: voltage profiles.

6 Conclusions

In this thesis I have proposed new methods for the distributed control of prosumers located among a multilevel hierarchical structure, and their coordination in a trustless setting. In this context, I have proposed novel methods for the unsupervised identification of PV power plants and the forecast of their power production.

The thesis focuses on the three main topics:

- Probabilistic forecasts for the 24 hours ahead residential power production and consumption, and their performance assessment by means of a-priori and closed loop tree-based stochastic MPC evaluation.

- Unsupervised methods for retrieving physics-based models for PV power plants, and their influence on the forecasting accuracy of PV power signals.

- Deterministic and probabilistic forecasting of hierarchical time series, in the case of a high penetration of PV power plants.

- Multilevel hierarchical control of prosumers in the distribution grid, and their coordination in a trustless setting.

6.1   Discussion

Probabilistic forecasts of residential power profiles   In the context of distributed control, in which residential forecasts and operation planning should be computed on a smart meter, methods with low computational requirements and memory usage must be taken into account. As remarked in [191], the yearly shiftable energy through DSM is strictly dependent on the penetration scenario of smart meters. The cost factor is pivotal to the roll out of these devices, and how to reduce smart meters cost is an active area of research [192]. Smart meters' cost is crucial for the deployment of DSM and for the economic benefits it can provide with respect to grid refurbishment. In this view, I have suggested new probabilistic forecasters using a

non-uniform stepsize for the training and for the prediction horizon. The benefit of having non-uniform stepsize is twofold, lowering in general the computational time for forecasting and control. In section 2.2 of chapter 2, I have numerically verified that for the class of control problems with linear-quadratic objective functions, using disturbance averages obtained with logarithmically spaced stepsizes provides a good approximation of the original problem. This led us to use a similar formulation for the forecasting algorithms. In fact, a non-uniformly sampled training set reduces the dimension of the historical values to be stored in the smart meter in order to train a forecaster. For a general purpose interpolator, such as the QRF, the training set is usually composed by Hankel matrices of some significant signals, such as the historical values of the power, temperature and irradiance. In chapter 2 it's shown how reducing a 15 minutes sampled training dataset using 10 logarithmically spaced bins does not affect the prediction accuracy. This reduces the training set dimension by a factor of 9.6, impacting on both the memory requirement and computational training time. On the other hand, other existing forecasting methods, such as the Holt-Winter forecaster, require a far lower computational time and memory usage. The drawback of the HW, which belongs to the class of the exponential smoothing forecasters, is that it cannot exploit exogenous signals. We propose to previously detrend the power forecasts by means of linear regressions taking into account both temperature and GHI, in order to overcome this limitation. We have additionally proposed a bagging of extreme learning machine (ELM) interpolators. ELMs are general purpose MIMO interpolators, and thus can directly output a consistent set of scenarios, without the need of modeling the temporal dependence of the predictions. They also present a favorable computational time with respect to QRF. In section 2.3 of chapter 2, the performance of the forecasters were assessed by means of a-priori KPIs and closed loop MPC control. The best forecasters on both evaluations was the QRF. In particular, this forecaster, coupled with a multivariate Gaussian copula for the modeling of the scenarios, provided a better description of the conditional pdf. This was confirmed by both the a-priori evaluation and by the superior performance gap reduction obtained with the QRF forecasts, when passing from a deterministic control to a stochastic one.

Unsupervised methods for PV modeling and their influence on forecasting   In chapter 3, unsupervised and robust methods to identify physics-based models for PV power plant were introduced. The methods are able to retrieve a model for the prediction of PV power output, starting from AC power measurements only. The resulting models map the meteorological conditions, in terms of GHI and T, to the power output of the PV. In section 3.2.1, was shown that the prediction performance of the identified models can be improved when coupling them with a RF, outperforming also the predictions of a RF without physics based information. This result shows that including physical knowledge about the geometric projections, low angle of incidence efficiencies, and temperature influence, helps increasing the accuracy of the estimator. In paper A, similar methodologies are used to disaggregate composite residential power measurements starting from the total power signal and from GHI observations. The different proposed methods rely on the hypothesis that the residential load is uncorrelated for

a given range of sampling times or frequencies with the PV power production. The methods have been tested on four residential buildings, in which electrical batteries were installed in order to increase the self consumption of the households. As expected, when the battery operations are not included in the overall power signal, the disaggregation techniques shown higher performances. In fact, when trying to maximize the self consumption, the battery operations get highly correlated with the PV power production, leading to an underestimation of its nominal power. The best performing method was the one based on a robust regression of the low-pass filtered power and GHI signals, whose only drawback the need for high frequency input data (> 1 sample/min). This would require the installation of a pyranometer, or to estimate GHI from a nearby monitored PV plant. This possibility was investigated in paper B, where a new unsupervised algorithm for estimating the GHI signal seen by a PV panel is presented. The method is able to fuse observations from different nearby power plants, thanks to its ability of identifying and mapping persistent shadings seen by each panel. Since the method is unsupervised, it can be used to turn already available power production data of residential PV power plants into GHI pseudo-measurements. Additionally, as remarked in B, the proposed method showed significant increase in the accuracy, with respect to commercially available satellite-based methods. Finally, the influence of PV modeling on forecasting the power production up to 24 hours ahead is investigated. In section 3.2.2, the PV models are used in combination with QRF, and their accuracy is evaluated using experimental data. An increase of performance is seen in the range going from one hour ahead up to 12 hours ahead. This can be explained considering the temporal resolution of the NWP and the frequency at which updates are available. The used NWP service provided forecasts with 1 hour temporal resolution. This resolution is not enough to provide an increase of accuracy for the first hour ahead when using a PV model, since at 1 hour resolution, the local measurements of PV production, that are already available, are more informative. This due to the low spatial resolution of NWP services, and is likely the same reason for which the method presented in paper B outperformed satellite-based irradiance estimation services. On the other hand, for forecasts beyond 1 hour ahead, modeling the PV provided higher accuracy, up to 12 hours. This can be explained considering that the forecasts are updated each 12 hours. It is highly likely that if the weather forecasts were updated more frequently, the PV modeling would perform better even in the time range between 12 and 24 hours.

Forecasting hierarchical time series   In chapter 4 I focused on the problem of forecasting hierarchical time series in distributed DMS problems. In this case, having local and global competing objective functions, or simply the inclusion of grid constraints in a multilevel setting, requires consistency of the forecasts at the bottom and at the aggregated level.
In the deterministic case, this is usually achieved through a bottom up approach, in which the bottom level forecasts are simply summed to estimate the forecast of the aggregated power. This has been shown to be suboptimal in the case in which the bottom level forecasters present a low signal to noise, as for residential power production/consumption. Following the work done in [126], in which a reconciliation techniques are used to fuse forecasts at different levels

of aggregation for the power retrieved by smart meter data, I proposed the decomposition of the problem through ADMM. The main advantage of decomposing the reconciliation relies in the fact that individual prosumers forecasts get aggregated by the branching nodes, and their private information is never disclosed to higher levels of the grid.

In a probabilistic setting, even the simple bottom up approach becomes non-trivial. In the simplest case in which forecasts errors are not correlated, the summation of random variables involves a multidimensional convolution. The sum of random variables is a subject of interest for risk aggregation, and from this field the use of hierarchical copulas [59] has been suggested. However, this technique requires to empirically estimate the copulas linking different time series in the hierarchy, while the one proposed in this thesis guarantees aggregate-consistent pdfs by construction, without the need of a reconciliation step. The approach I proposed requires a NWP provider to broadcast ordered scenarios for the forecasts of the meteorological conditions (GHI and T), for the next prediction horizon. As remarked in section 4.2, this seems feasible in practice, since to obtain accurate power forecasts, single users already need to retrieve predictions of meteorological conditions from a NWP provider. Moreover, the amount of data can be reduced if instead of a set of scenarios, the provider pass a tree of prediction, previously obtained as described in section 1.3.

Multilevel hierarchical control and trustless coordination   In paper C, I proposed a method for systematically taking into account multiple possibly contrasting objective functions of agents located at different voltage levels in the distribution grid. This was done by decomposing a monolithic convex problem in the presence of separable coupling constraints for the agents, using techniques of networked optimization and using an ADMM based convergence scheme. The convergence of both primal and dual variables is imputed to agents based on their location in the hierarchical structure. This resembles the physical structure of the distribution grid, for which agents located under different transformers can influence each other's actions or constraints only indirectly. The same concept is retained in paper D, where I proposed a energy market for energy sharing communities. In this case I have used a generalized Nash equilibrium refinement, called variational equilibrium [177], which ensures equal marginal prices for agents responsible for the same sets of grid constraints. The main outcomes of paper D are provided by theorem II.1, which demonstrates existence and uniqueness of the VNE for the class of games generated by the class of sharing problems, and the demonstration in chapter III that this equilibrium can be reached re-weighting the quadratic terms of proximal operators generated by the standard ADMM decomposition. A superior convergence in terms of iterations has been shown with respect to the pFB algorithm for solving the same class of problems [133; 64]. Additionally, in order to ensure the individual rationality condition, requiring that prosumers never incur in higher costs with respect to the base case, I proposed to cap the Lagrangian multipliers associated to their grid constraints. This procedure reflects a natural economic reasoning, which prevents prosumers from providing their flexibility for free at the only scope of balancing voltage and power fluctuations in the grid. A system operator willing to increase the exploitation of prosumers' owned assets for this

scope, should provide them with a discount in terms of energy tariffs.

6.2 Future development

In the following, suggestions for future works are presented based on treated topics.

Probabilistic forecasts of residential power profiles

- Experimental validation of forecasters performances for control. The a-priori and a-posteriori evaluation of the forecasters in this thesis did not include a detailed nonlinear model for the simulation of the operated battery. This was avoided to fairly compare the forecasters performance without dealing with goodness of fit for the identified battery model. For real life operation, the state of charge of the battery would be provided by a proprietary API, while the control of the battery could be done through a piecewise linear time-varying model as in [97]. However, in this way, the performance assessment of the forecasters would have been influenced by the goodness of fit of the model identification, undermining the validity of the results. However this additional step is needed in order to estimate the profitability of using different forecasters.

Unsupervised methods for PV disaggregation

- Use monitored PV panels to estimate GHI and disaggregate composite power flows. The most accurate disaggregation technique from A, required a sampling time of 1 minute for the GHI signal. This temporal resolution is not currently available from satellite services. This would require the installation of a pyranometer. However, this kind of instrument is expensive and usually installed only at utility-scale PV power plants for monitoring reasons. As an alternative, a monitored PV panel could be used to estimate the local GHI, with the technique described in paper B, in order to disaggregate composite power flow of nearby buildings.

Forecasting hierarchical time series

- Use of aggregate-consistent pdf in hierarchical control. The distributed stochastic control in electrical grid is a challenging topic, since requires to consider coupled probabilistic constraints. This class of problems is considered in [156], where under the assumption of independent disturbance, coupling constraints are treated computing discrete scalar convolutions of the disturbance pdfs of single agents. However, as seen in section 4.2, independence of the forecasts errors does not hold in the presence of high penetration of PVs. The method proposed in the same section to obtain aggregate-consistent forecasts can be coupled with TBSMPC and decomposition techniques of

chapter 5 to solve the the problem in a stochastic setting. This will require, in addition to a disturbance and a decision variables tree, a tree of Lagrangian multipliers for the coordination of the prosumers' actions.

Multilevel hierarchical control and trustless coordination

- Closed loop simulations including electric grid feedback. Although the accuracy of sensitivity coefficients has been investigated, a closed loop evaluation of the proposed strategy including feedback from a simulated grid, is necessary.

- Include mesh grid formulations. The algorithm presented in paper C is suited to take into account linear formulations of the PF. However it was not tested using a combination of PF approximations for radial and mesh grids. This is necessary when willing to control loads located both in the LV (typically radial) and MV (typically meshed) grids.

- Comparison with other trustless methods. The method proposed in paper D for the trustless coordination of prosumers is based on the concept of Nash equilibrium. One of the advantages of modeling the optimization problem as a game is that the theoretical properties are independent on the specific hardware and software implementation. The prosumers will simply incur in a loss if they do not follow the suggested strategy. On the other hand, the total achievable welfare under this equilibrium concept is lower than the one achievable by a centralized (and trusted) controller. Distributed computing techniques, such as secure multiparty function evaluation (SMFE), could represent an alternative solution for achieving trustless coordination. However, the number of messages required by SMFE could present quadratic grow when applied to distributed control [189]. Other techniques, such as probabilistic checkable proofs (PCP), are able to guarantee faithful execution of code on a hosting machine and to keep personal information private. However, the computational burden increment with respect to native code performances of various protocol implementing PCP, seems to be incompatible with distributed control applications [193]. On the other hand, the reported computational and communication complexity of the recently proposed ZK-STARK protocol [194] seems to allow its use for distributed control and trustless coordination.

A Appendix A

Paper A: *F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised Disaggregation of Pho-toVoltaic Production from Composite Power Flow Measurements of Heterogeneous Prosumers," IEEE Transactions on Industrial Informatics, 2018.* List of authors:

- Sossan, Fabrizio: Distributed Electrical Systems Laboratory at the Swiss Federal Institute of Technology of Lausanne (DESL, EPFL).

- Nespoli, Lorenzo: Swiss Federal Institute of Technology of Lausanne (EPFL), University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

- Medici, Vasco: University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

- Paolone, Mario: Distributed Electrical Systems Laboratory at the Swiss Federal Institute of Technology of Lausanne (DESL, EPFL),

Personal contributions:

- Methods A and B for the disaggregation of composite power flows

- Performance assessment of the different methods

1

Unsupervised Disaggregation of PhotoVoltaic Production from Composite Power Flow Measurements of Heterogeneous Prosumers

Fabrizio Sossan, *Member, IEEE,* Lorenzo Nespoli, *Member, IEEE,* Vasco Medici, *Member, IEEE,* and Mario Paolone, *Senior Member, IEEE,*

Abstract—We consider the problem of estimating the unobserved amount of photovoltaic (PV) generation and demand in a power distribution network starting from measurements of the aggregated power flow at the point of common coupling (PCC) and local global horizontal irradiance (GHI). The estimation principle relies on modeling the PV generation as a function of the measured GHI, enabling the identification of PV production patterns in the aggregated power flow measurements. Four estimation algorithms are proposed: the first assumes that variability in the aggregated PV generation is given by variations of PV generation, the next two use a model of the demand to improve estimation performance, and the fourth assumes that, in a certain frequency range, the aggregated power flow is dominated by PV generation dynamics. These algorithms leverage irradiance transposition models to explore several azimuth/tilt configurations and explain PV generation patterns from multiple plants with non-uniform installation characteristics. Their estimation performance is compared and validated with measurements from a real-life setup including 4 houses with rooftop PV installations and battery systems for PV self-consumption.

Index Terms—PV generation, Demand, Disaggregation, Optimization problems, Algorithms, Unsupervised learning.

Nomenclature and Acronyms

PV	Photovoltaic.
GHI	Global Horizontal Irradiance.
GNI	Global Normal Irradiance.
nRMS	Normalized Root Mean Square Error.
PCC	Point of Common Coupling.
MPPT	Maximum Power Point Tracking.
k	Discrete time index.
P_k	Active power flow at the PCC at time interval k.
$\mathcal{I}_k^-$	GHI observation at time k.
j	Index for panel tilt/azimuth configuration.
$I_{jk}^{\searrow}$	Estimated GNI corrected for temperature for configuration j at time k.
$\widehat{G}_k$	Estimated PV production at time interval k.
$\widehat{L}_k$	Estimated demand at time interval k.
α_j	PV nominal capacity at configuration j.

F. Sossan and M. Paolone are with the Distributed Electrical Systems Laboratory at the Swiss Federal Institute of Technology of Lausanne (DESL, EPFL), L. Nespoli and V. Medici are with the ISAAC at the University of Applied Sciences and Arts of Italian Switzerland (SUPSI), CH. Emails: {fabrizio.sossan,mario.paolone}@epfl.ch, {vasco.medici, lorenzo.nespoli}.supsi.ch

I. Introduction

Incresed levels of distributed photovoltaic (PV) generation determine higher reserve requirements at the system level and violations of voltage and line ampacity constraints in distribution systems during peak production hours [1], [2]. Technical solutions envisaged to mitigate PV generation drawbacks are curtailment strategies, control of converters active/reactive power, PV self-consumption schemes and dispatch of local power flows according to network-safe power consumption trajectories (e.g. [3]–[11]). A requirement for the implementation of those strategies is the availability of real-time production measurements from PV facilities. Incidentally, these are also useful to train data-driven prediction models (e.g. [12], [13]). However, such a precondition is not always met in real-life conditions because installations are not always monitored, and, even when they are, factors such as *i)* privacy concerns, *ii)* conflicts due to the different owners of the metering infrastructures, and *iii)* lack of standards for monitoring and aggregation of measurements, and their communication, play against the possibility of collecting real-time PV production measurements.

As an alternative to direct monitoring of PV systems, we consider in this paper the problem of disaggregating PV generation from the aggregated active power measurements of a group of prosumers. The estimation principle relies on modelling PV generation as a function of the global horizontal irradiance (GHI), assumed known from local measurements. Four estimation algorithms are proposed and compared: the first assumes that the variability in the aggregated power flow measurements are mostly given by variations of the PV generation, the second and third leverage a model of the demand to improve estimation performance, and the fourth assumes that there is a certain frequency range in which the aggregated power flow measurement is dominated by PV generation components. All four algorithms use a transposition model to project GHI into a number of pre-defined differently oriented tilted planes to explain production from sites with different configurations. The algorithms are designed to be unsupervised, i.e., they do not require measurements of the PV power profiles to be trained. The algorithms are tested with measurements from a real-life setup of four houses with monitored rooftop PV plants and grid-connected battery systems, enabling the testing of estimation performance even when the demand is correlated with PV generation.

Even if in the existing literature the problem of disaggregation has been extensively investigated for nonintrusive load monitoring (e.g., [14], [15]), its application to PV disaggregation was considered in [16], which develops estimators of the total PV generation using the active power profile of a nearby installation and GHI proxy measurements as explanatory signals. With respect to [16], we leverage PV transposition models to identify PV production patterns from installations with different tilt/azimuth configurations, a key factor in urban contexts where PV generation is from rooftop PV plants and tilt and azimuth configurations are dictated by roofs characteristics and might not be uniform.

The paper is organized as follows: Section II states the problem, III presents the disaggregation algorithms, IV discusses procedures and measurements for validation, V presents and discusses performance, VI summarizes key results and contributions.

II. NOTATION AND PROBLEM STATEMENT

A. Configuration of the system

We consider a feeder with distributed PV production, possibly from installations with different tilt and azimuth, and demand (e.g., Fig. 1). The power injections at the single buses are not measured, however the total prosumption (PV generation + demand) is known thanks to sensing the active power flow[1] at the point of common coupling (PCC). Local GHI values are from a pyranometer (although other methods could be considered, e.g. leveraging information available from nearby monitored PV installations [17]). We do the modeling assumption that the PV installations in the area are subject to the same GHI. Local GHI measurements are known to be accurate in a range of 50 meters [18]. Therefore, the proposed algorithms are expected to perform adequately when PV plants are spread over a small area, and their performance to decrease when considering larger areas. Due to the small size of the networks that these methods target, grid transmission losses are neglected at this stage due to the short length of the cables.
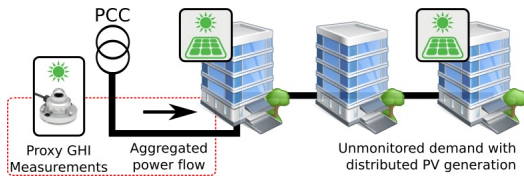
Fig. 1. A network topology with unmonitored demand and PV generation from multiple production sites with different azimuth/tilt configurations. The active power flow at the PCC and GHI are known from measurements. The problem is estimating the raw PV generation.

B. Notation

The active power flow measured at the PCC at the discrete time interval k is denoted as P_k (kW). Positive flows denote consumption and vice-versa (i.e., passive sign notation).

[1]Reactive power is not of special interest since PV plants normally operate at unitary power factor and, more in general, it is not possible to do assumptions on the kind of reactive power control implemented.

GHI measurements are denoted by $\mathcal{I}_k^-$ in $kW\,m^{-2}$, while $I_{jk}^{\searrow}$ ($kW\,m^{-2}$) denotes the estimated global normal irradiance (GNI) to a certain tilted plane j corrected for temperature (as described in III-E), where $j = 1, \ldots, J$ denotes the plane tilt/azimuth configuration. We consider $J = 21$ tilted planes with tilt and azimuth values equally spaced on a south-facing semi-sphere, chosen to have a reasonably representative set of potential configurations of PV installations in the northern hemisphere. GNI estimations are from the Hay-Davies transposition model [19], [20]. The quantity $\widehat{G}_k$ and $\widehat{L}_k$ (kW) respectively denote the estimated PV production and estimated demand, which are to be determined. A practical example of the disaggregation process is described in the following paragraph.

C. Problem statement

The problem is estimating the trajectories of the demand and total PV generation from measurements of the active power flow at the PCC and local GHI observations. This is exemplified in Fig. 2a (night time observations are omitted) which shows the inputs, intermediate results and outputs of one among the proposed algorithms. The inputs are the aggregated power flow P_k at the PCC (top panel) and GHI (middle panel, solid fill). The middle panel plot in Fig. 2b also shows the GNI trajectories $I_{jk}^{\searrow}$, used to explore the potential PV production from plants with various tilt/azimuth configurations, as typical in urban feeders where panels are installed according to roof characteristics. Finally, the lower panel plot in Fig. 2c shows the output of one of the proposed algorithms[2], with the estimated demand and estimated PV generation (orange line), the latter close to the measured ground truth PV generation (solid gray fill).

III. DISAGGREGATION ALGORITHMS

The estimated global PV generation $\widehat{G}_k$ at the PCC is modelled as the sum over all the tilt/azimuth configurations $j = 1, \cdots, J$ of the transposed irradiance $I_{jk}^{\searrow}$ times J nonnegative coefficients $\alpha_j \in \mathbf{R}_+$:

$$\widehat{G}_k(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \alpha_j \cdot I_{jk}^{\searrow}, \qquad k = 1 \ldots, K, \qquad (1)$$

where $\boldsymbol{\alpha} = \{\alpha_j, j = 1, \cdots, J\} \in \mathbf{R}_+^J$ denotes the set of α_j. Physically, α_j is the PV generation capacity installed at configuration j. It is measured in kWp, kilowatt peak, and denotes the amount of power produced in standard test conditions (STC, $kW\,m^{-2}$ GNI at $20\,°C$). The temperature effect on the PV conversion efficiency is accounted for by preprocessing the input GNI time series as described in Section III-E. By modelling the PV generation as in (1), we assume that PV plants operate in the *maximum power point tracking* (MPPT) mode; in case the output of a PV plant is controlled (i.e. curtailed), it is likely to be monitored and its contribution can be removed from the aggregated power flow, still allowing to apply the algorithms. Partial shading effects is not explicitly

[2]Method C, parameters $T_s = 30$ s $c = 10$, mean nRMSE 5%.

Appendix A. Appendix A

(a) **Observed active power flow at the PCC.**

(b) **Exploratory GNI profiles derived from GHI observations (solid).**

(c) **Estimated demand and PV generation (*vs.* ground truth PV, solid, Method C).**
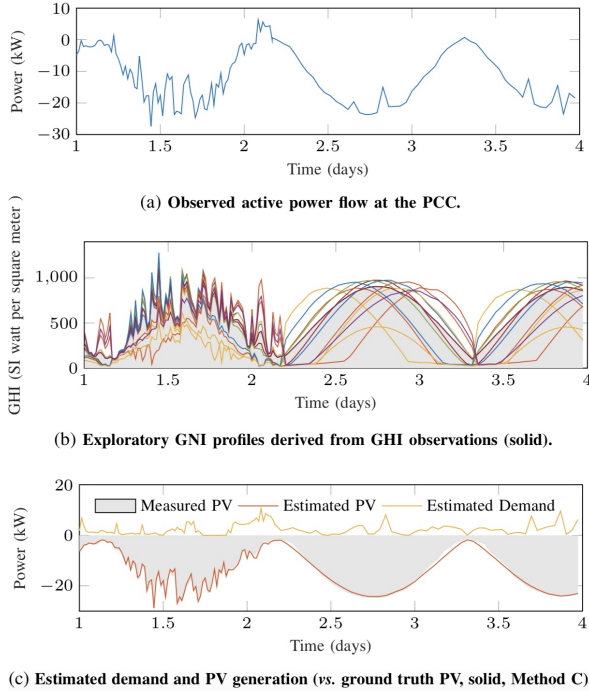
Fig. 2. Input, intermediate results and output of the proposed disaggregation algorithms (night hours not shown).

modeled, even if, as discussed in the following, some of the proposed methods are robust against it.

As detailed in the following, the proposed disaggregation algorithms estimate α (thus PV generation $\widehat{G}_k$) and require a training phase. Four algorithms, denoted as Method A, B, C and D, are discussed. They attempt to estimate PV generation by exploiting different modeling principles inspired by the following empirical considerations:

- variations of PV generation dominate the variations of the power flow at the PPC. Method A estimates PV generation by seeking for a trajectory with variations as close as possible to those observed at the PCC. The drawback is that variations are also due to demand changes (e.g. load inrushes);

- the power flow at the PCC is modelled as the sum between PV generation and demand, where the latter is described by using a load model. Method B and Method C exploit two different load models, as explained later;

- in a certain frequency range, the dynamics of the power flow at the PCC are dominated by those of PV generation. Method D exploits the fact that demand and PV generation have different time dynamics: filtering the power flow measurements at the PCC makes possible to estimated the PV generation.

The validity of these empirical modelling considerations are tested in the results section by assessing and comparing the algorithms performance.

A. Method A

The unknowns $\boldsymbol{\alpha}$ are determined by assuming that the variability in the observed aggregated power flow are due to variations of the aggregated PV power. This is modeled by minimizing the norm-1 of the difference between the once differentiated time series P_k and $\widehat{G}_k$ while subject to the estimated total PV production model (1):

$$\boldsymbol{\alpha}^o = \arg\min_{\boldsymbol{\alpha} \in \mathbf{R}_+^J} \left\{ \sum_{k=1}^{K} \left| (P_k - P_{k-1}) + \right. \right. \left. \left. - \left(\widehat{G}_k(\boldsymbol{\alpha}) - \widehat{G}_{k-1}(\boldsymbol{\alpha}) \right) \right| \right\} \quad (2)$$

subject to:

$$\widehat{G}_k(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \alpha_j \cdot I_{jk}^{\searrow}, \qquad k = 1 \ldots, K. \quad (3)$$

The problem in (2)-(3) is linear, thus convex and tractable. The resolution of the input time series is a parameter of the Method A, and its importance is discussed in the performance assessment. Method A does not allow to model demand dynamics, e.g. load inrushes would be considered as variations of PV generation. The next two methods use a model of the demand to work around this limitation.

B. Method B

Let $\widehat{\boldsymbol{L}} = [\widehat{L}_1, \ldots, \widehat{L}_K]$ be the estimated demand trajectory. The estimated active power flow at the PCC $\widehat{P}_k$ is now written as:

$$\widehat{P}_k(\boldsymbol{\alpha}, \widehat{\boldsymbol{L}}) = \widehat{L}_k - \widehat{G}_k(\boldsymbol{\alpha}), \qquad k = 1 \ldots, K, \quad (4)$$

i.e. the sum between the estimated total PV generation and demand, the former with negative sign because corresponding to generation. Method B attempts to determine $\widehat{L}_k$ and $\boldsymbol{\alpha}$ by minimizing the norm-2 of the estimation error $P_k - \widehat{P}_k$. However, this problem is under-determined since the $K + J$ free variables are more than the number of observations K. Therefore, we augment the cost function and consider the sum of the least square and norm-1 of the once differentiated $\widehat{L}_k$ series (i.e., a combined linear regression and trend filtering problem, as for example in [21] [22]):

$$\begin{bmatrix} \boldsymbol{\alpha}^o \\ \widehat{\boldsymbol{L}}^o \end{bmatrix} = \arg\min_{\begin{bmatrix} \boldsymbol{\alpha} \in \mathbf{R}_+^J \\ \widehat{\boldsymbol{L}} \in \mathbf{R}_+^K \end{bmatrix}} \left\{ \sum_{k=1}^{K} \left(P_k - \widehat{P}_k(\boldsymbol{\alpha}, \widehat{\boldsymbol{L}}) \right)^2 + \right. \left. + \lambda \sum_{k=1}^{K} \left| \widehat{L}_k - \widehat{L}_{k-1} \right| \right\} \quad (5)$$

subject to:

$$\widehat{P}_k(\boldsymbol{\alpha}, \widehat{\boldsymbol{L}}) = \widehat{L}_k - \sum_{j=1}^{J} \alpha_j \cdot I_{jk}^{\searrow}, \qquad k = 1, \ldots, K. \quad (6)$$

The cost function is the sum of a vector norm-1 and a quadratic cost function and convex if the latter term is convex. As shown in Appendix A, the convexity of the quadratic term cannot depends on the input data and can be verified a-priori.

C. Method C

In Method B, under-determination was solved by minimizing the norm-1 of the demand trajectory $\widehat{L}$. As an alternative, we apply here a piecewise constant model of the demand, i.e. we require the unknown sequence $\widehat{L}$ to be piecewise constant for c consecutive samples, where c is a parameter, by enforcing the following $c-1$ equality constraints:

$$\widehat{L}_1 = \widehat{L}_2 = \cdots = \widehat{L}_c, \qquad (7)$$

for the case of the first c samples. Extending to the set of K measurements (K multiple of c) yields:

$$\widehat{L}_{c(i-1)+1} = \cdots = \widehat{L}_{c(i-1)+c} \qquad i = 1, \ldots, K/c. \qquad (8)$$

Modelling the demand as piecewise constant is a reasonable assumption when the length of the constant segment does not overlap with typical intra-day dynamics of the demand, i.e. for small c values and densely sampled series. In other words, it is reasonable when considering short periods of time (e.g., seconds), when the persistence model of the demand has unbeaten performance, see, e.g., [23], [24]. When the demand has shorter variations than c (e.g., load inrushes), the estimated demand will have the average value of the waveform and the residuals will be the estimation error. The sampling time and c are design parameters: the sensitivity of the algorithm performance with respect to their values is assessed in Section V. Method C consists in minimizing the norm-2 of the estimation error $P_k - \widehat{P}_k$ subject to the estimated aggregated power flow and piecewise constant demand models:

$$\begin{bmatrix} \boldsymbol{\alpha}^o \\ \widehat{\boldsymbol{L}}^o \end{bmatrix} = \operatorname*{arg\,min}_{\substack{\boldsymbol{\alpha}\,\in\,\mathbf{R}_+^J \\ \widehat{\boldsymbol{L}}\,\in\,\mathbf{R}_+^K}} \left\{ \sum_{k=1}^{K} \left(P_k - \widehat{P}_k(\boldsymbol{\alpha}, \widehat{\boldsymbol{L}}) \right)^2 \right\} \qquad (9)$$

subject to:

$$\widehat{P}_k(\boldsymbol{\alpha}, \widehat{\boldsymbol{L}}) = \widehat{L}_k - \sum_{j=1}^{J} \alpha_j \cdot I_{jk}^{\searrow}, \qquad k = 1, \ldots, K \qquad (10)$$

$$\widehat{L}_{c(i-1)+1} = \cdots = \widehat{L}_{c(i-1)+c} \qquad i = 1, \ldots, K/c, \qquad (11)$$

The additional constraints (11) are linear and do not impact convexity, thus the same consideration as for Method B applies.

D. Method D

Method D splits the observed active power flow at the PCC by exploiting similarities between the signals P_k and $\widehat{G}_k(\boldsymbol{\alpha})$ in a certain frequency range. This approach is motivated by having verified similarities in the spectral density of the measured aggregated power flow and measured PV generation (available from the test site) with the Welch's periodogram method for coherence analysis [25]. Method D initially filters the input GNI $I_{jk}^{\searrow}$ and aggregated power flow P_k with a sixth order Butterworth band-pass filter, where the low and high cut-off frequencies are parameters that reflect the frequency range where the aggregated power profile and PV generation are similar. They are free parameters and the sensitivity of the algorithm performance to their values is investigated in

Section V. Let $\mathcal{P}_k$ and $\mathcal{I}_{jk}^{\searrow}$ respectively denote the above mentioned filtered version of the sequence P_k and transposed irradiance $I_{jk}^{\searrow}$. The vector of unknowns $\boldsymbol{\alpha}^o$ is computed by the following robust linear regression:

$$\boldsymbol{\alpha}^o = \operatorname*{arg\,min}_{\boldsymbol{\alpha}\,\in\,\mathbf{R}_+^J} \left\{ \sum_{k=1}^{K} \rho\left(\mathcal{P}_k - G_k(\boldsymbol{\alpha}) \right) \right\} \qquad (12)$$

subject to:

$$\widehat{G}_k(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \alpha_j \cdot \mathcal{I}_{jk}^{\searrow}, \qquad k = 1 \ldots, K. \qquad (13)$$

where $\rho(\cdot)$ is the bisquare loss function, see [26], a nonconvex relationship which gives less weight to extreme values in the cost function to be robust against outliers. The problem (12)-(13) is solved by applying an iterative least square approach with guaranteed covergence [27].

E. Temperature correction

To account for the dependency between PV conversion efficiency and temperature, GNI values are corrected to reflect temperature variations from the reference value T_{ref} (25 °C) by using the empirical model proposed in [28]:

$$I_{jk}^{\searrow} = \mathcal{I}_{jk}^{\searrow} \left[1 + \gamma(T_{\text{cell},k} - T_{\text{ref}}) \right] \qquad (14)$$

where $\mathcal{I}_{jk}^{\searrow}$ is the original irradiance value, and $T_{\text{cell},k}$ the cell temperature at time k, estimated as [28]

$$T_{cell,k} = T_{\text{air}} + \beta \mathcal{I}_{jk}^{\searrow}, \qquad (15)$$

where T_{air} is the air temperature, assumed known from local measurements, and $\beta = 3.78 \times 10^{-2}$ and $\gamma = -4.3 \times 10^{-3}$ are empirical and plant specific values here assigned considering typical average values[3].

IV. METHODS FOR PERFORMANCE EVALUATION

The performance assessment of the disaggregation algorithms is performed by evaluating their ability to reconstruct the PV generation time series starting from the measurements of the power flow at the PCC and GHI. Let $e_k = G_k - \widehat{G}_k$ (G_k and $\widehat{G}_k$ are the PV generation ground-truth value and estimation, respectively) be the estimation error. The performance metrics are: normalized root mean square of the estimation error $\left(\frac{1}{G} \frac{1}{K} \sum_{k=1}^{K} e_k^2 \right)^{1/2}$, normalized mean absolute error (nMAE) $\frac{1}{G} \frac{1}{K} \sum_{k=1}^{K} |e_k|$, and normalized mean error (nME) $\frac{1}{G} \frac{1}{K} \sum_{k=1}^{K} e_k$, where G is the total installed PV capacity (35.3 kWp) and K is the samples number in the testing data set.

---

[3] β is the average of values for the close roof mount and open rack configurations from [28], and γ the average of the values for polycrystalline modules from [29].

A. Data sets for training and testing

Time series are GHI, power flow at the PCC and PV generation measurements for 1 year from the real-life test facility described in the next paragraph. The first 2 are used for the training, while the last as the ground truth value to test the estimation performance. To preserve daily dynamics of the signals, time series are divided into daily sequences, then randomly shuffled to avoid to train and test the algorithms on different periods of the year. The series time resolutions are 10, 30, 60, 120, 300, 600 and 900 sec (downsampling by samples average) and are to assess algorithms performance with respect to sampling time. Selected resolutions include those normally implemented in existing metering systems, e.g., 900 s is the resolution of smart meters in Switzerland and Italy. Here, the intent is to verify whether such a sampling time is enough for the proposed application, or if performance would benefit from more densely sampled data. Each of the 7 datasets at different resolution is further split into 3 sub-sequences to perform a three-fold cross-validation, i.e., for each resolution, the algorithms are trained on the first fold and tested on the remaining 2; the process is repeated for all the 7 datasets, each time testing the algorithms on the part of the data which is not used for the training. In total, each algorithm is trained and tested 3 times for each resolution, for a total of 21 training and test runs. Measurements refer to days with a uniform mix of sky conditions: partly cloudy, clear sky and overcast. Algorithms performance is tested both when there are batteries performing PV self-consumption and when not, thus allowing to account for the case when the demand is correlated with PV generation.

B. Experimental Setup

Measurements are from a real-life setup in the region of Basel, Switzerland, with 4 private households, each equipped with a rooftop PV installation with different characteristics, as reported in Table I. PV converters operate in MPPT mode at unitary power factor. The total PV installed capacity is 35.3 kWp and the peak demand is 12 kWp. The households are also equipped with grid-connected battery systems with bidirectional power converter to implement PV energy self-consumption policies (actuated at 5 minutes resolution). Batteries specifications are summarized in Table I. Battery injections are monitored. They are removed from the power flow at the PCC by algebraic difference, such that, in the following analysis, it is possible to consider two cases: with and without battery action (self consumption).

TABLE I
PV AND BATTERY SYSTEMS IN THE EXPERIMENTAL SETUP

House ID	PV capacity (kWh)	Azimuth	Tilt (°)	Distance from House 1 (m)	Battery rating (kVA/kWh)
1	10.0	95	14	0	3/8.8
2(a)	7.2	187	36	100	3/4.4
2(b)	3.5	266	40	100	–
3	8.0	187	40	260	3/8.8
4	6.6	180	24	170	3/4.4

a) PV and power flow measurements: The power flow at the PCC is the sum of the four households flows, measured synchronously at 10 s resolution. Similarly, the global PV production (used as the ground truth value to validate the estimation performance of the algorithms) is the sum of the single PV facilities power injections, measured at the converter level.

b) Global horizontal irradiance measurements: GHI measurements are from a pyranometer installed on the roof of the household ID1. The line distance of the remaining households from the GHI observation point is shown in the second to last column of Table I. All the measurements are synchronized and timestamped, and logged in a time series database.

V. RESULTS AND DISCUSSION

For a visual exemplification of the disaggregation process, the reader is referred to Section II-C and Fig. 2. In this section, we first assess the performance of the proposed methods individually. In V-E, we perform a joint performance assessment to compare the quality of the estimations of the different algorithms and support the assertion that they can be considered unsupervised. Key results are discussed and summarized in V-F. In V-G, the algorithms are tested in scenarios with a lower penetration of PV generation to verify if less prominent patterns of PV generation are detrimental to estimation performance. Finally in V-H, we discuss the computational performance.

A. Method A

The estimation nRMSE as a function of the only parameter of Method A (i.e. sampling time) with and without PV self-consumption is shown in Fig. 3. With no self-consumption, the nRMSE stabilizes at around 2 kW for sampling times larger than 200 s; with self-consumption, performance is poorer and is best at around 150 s. In both cases, estimation performance when the input time series is densely sampled (large sampling time) is poor.
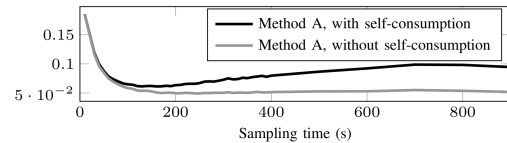
Fig. 3. Method A nRMSE as a function of the input time series sampling time.

B. Method B

On top of the series sampling time, Method B has a smoothing parameter λ in (5) to weight the demand time variations $\widehat{L}_k - \widehat{L}_{k-1}$ in the cost function. Performance as a function of the two parameters is shown in Fig. 4. With self-consumption, best performance happens in the middle right region of the parameter space. This region is larger in the case without self-consumption. Performance degradation patterns do not have a well identifiable trend.

C. Method C

The parameters of Method C are input time series sampling time and piecewise constant segment length c (in number of samples). Their influence on the nRMSE is shown in Fig. 5. Estimation performance decreases when moving away from the axis origin, denoting that using densely sampled input time series and small c values (the best performance is with 20 s resolution) are convenient. As mentioned in the formulation stage, this is to be expected because the choice of the two parameters affects the constant segment length of the demand piecewise constant model (i.e., the shorter it is, the better performance the persistence model has). Estimation performance is worse with self-consumption (a numerical quantification is given in the next paragraph). Contour lines of Fig. 5 denote that the performance degradation follows the same pattern when with and without self-consumption. Thus, even if estimation performance is different in the two cases, the optimal values of the parameters lay in the same parameters space area.

D. Method D

Method D parameters are the lower and upper cut-off frequencies of the bandpass filter, and a tuning constant of the bisquare loss function $\rho(\cdot)$. The last was found not to impact substantially on the algorithm performance and is therefore excluded from the current analysis. The sensitivity of algorithm performance to upper and lower cut-off frequencies is shown in Fig. 6. Best performance happens in a well identifiable region in the lower left part of the parameter space, which however tend to shrink in the case with self-consumption.

E. Joint Performance Comparison

The *min, max, mean, and median* statistics of the estimation nRMSE, nMAE and nME of the 4 algorithms are reported in Table II. For each algorithm, the reported statistics are calculated over all the combinations of the considered parameters values. They are to be interpreted in the following way:

- *min*: performance to be expected assuming to know a-priori the best performing set of parameters.
- *max*: performance to be expected when choosing the worst possible combination of parameters.
- *mean*: performance to be expected when choosing a random combination in the parameters space.
- *median*: to evaluate performance distribution skewness.

Table II shows that all the methods perform poorer under self-consumption regimes (e.g. mean nRMSE 4.6 to 5.6% and 5.3 to 7% for Method D and Method C, respectively), in other words when the demand includes a component anti-correlated with PV generation. In terms of *mean* and *median* statistics, Method D scores the best metrics, followed by C, B and A. In terms of *min* value, Method C outperforms the other, except for for the cases nMAE and nRMSE without self-consumption, where Method B is better, and nME with self-consumption, where Method D is the absolute best for all the metrics.

TABLE II
ESTIMATION PERFORMANCE STATISTICS (%)

Statistic	A	B	C	D
nRMSE with self-consumption				
min	5.22	5.2	**4.0**	4.3
max	20.6	17.8	10.4	**8.2**
mean	9.7	9.2	7.0	**5.5**
median	8.4	8.4	7.0	**5.4**
nRMSE without self-consumption				
min	3.5	4.4	**3.4**	4.1
max	17.9	17.7	9.5	**8.8**
mean	6.6	7.62	5.3	**4.6**
median	4.6	4.8	4.7	**4.4**
nMAE with self-consumption				
min	3.3	3.3	**2.5**	2.7
max	14.0	11.8	8.5	**5.1**
mean	6.2	5.9	4.7	**3.4**
median	5.3	5.3	4.4	**3.4**
nMAE without self-consumption				
min	**2.2**	2.8	2.3	2.5
max	12.1	11.8	**5.4**	5.6
mean	4.3	4.9	3.2	**2.9**
median	3.0	3.0	3.0	**2.7**
nME with self-consumption				
min	-13.9	-11.7	-8.1	**-4.3**
max	**-1.8**	-1.8	8	4.3
mean	-5.7	-5.2	2.3	**-1.0**
median	-4.7	-4.8	-2.8	**-1.3**
nME without self-consumption				
min	-11.9	**-1.2**	-4.8	-5.0
max	1.9	**-1.1**	4.6	4.0
mean	-2.2	-3.8	0.8	**0.5**
median	-1.2	-1.7	-1.0	**0.2**

F. Discussion

The previous results showed that the algorithms with the largest number of best scores is Method D, followed by C, B and A. If only sparsely sampled power flow observations are available (such as those from smart meters, typically at 15 minutes resolution), Method D should be selected because it keeps good performance even at low resolutions. If densely sampled observations are available, the performance of Method C and D are comparable. In this case, Method C has two advantages: *i*) parameters can be selected with an educated physical-based guess, *ii*) degradation patterns are the same for both with and without self-consumption.

When selecting the two best performing Methods (C and D), the mean nRMSE is 7% and 5.5% and nME is rather small, 0.5 and 0.8%. The latter metric is of importance because it denotes that estimators are almost unbiased, in other words, even if a single estimation in time is wrong, the estimated global PV production over a period is nearly correct.

It is worth noting that, in the proposed sensitivity analysis, PV measurements have been used to assess estimation performance. However, in practical applications when PV observations are not available, it is not possible to do so otherwise there would be no use for disaggregation algorithms. In case of Method C (min/max nRMSE in the range $4.0 \div 13.6\%$), parameters can be chosen with an educated guess in order to get closer to the best performance. As far as Method D is concerned, the min/max nRMSE gap is smaller ($4.2 \div 9.6\%$), and estimation performance is good over a wide

Appendix A. Appendix A

(a) nRMSE with self-consumption.

(b) nRMSE without self-consumption.
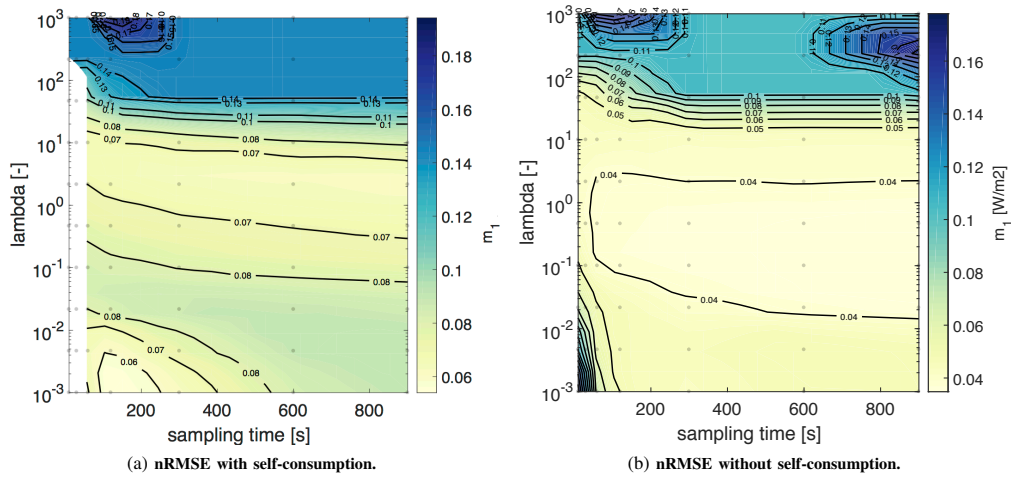
Fig. 4. Method B sensitivity analysis: nRMSE as a function of the input time series sampling time and scaling parameter λ.

(a) nRMSE with self-consumption.

(b) nRMSE without self-consumption.

Fig. 5. Method C sensitivity analysis: nRMSE as a function of the input time series sampling time and and length c of the piecewise constant segment.

(a) nRMSE with self-consumption.

(b) nRMSE without self-consumption.

Fig. 6. Method D sensitivity analysis: nRMSE as a function of the lower and upper cut-off frequencies.
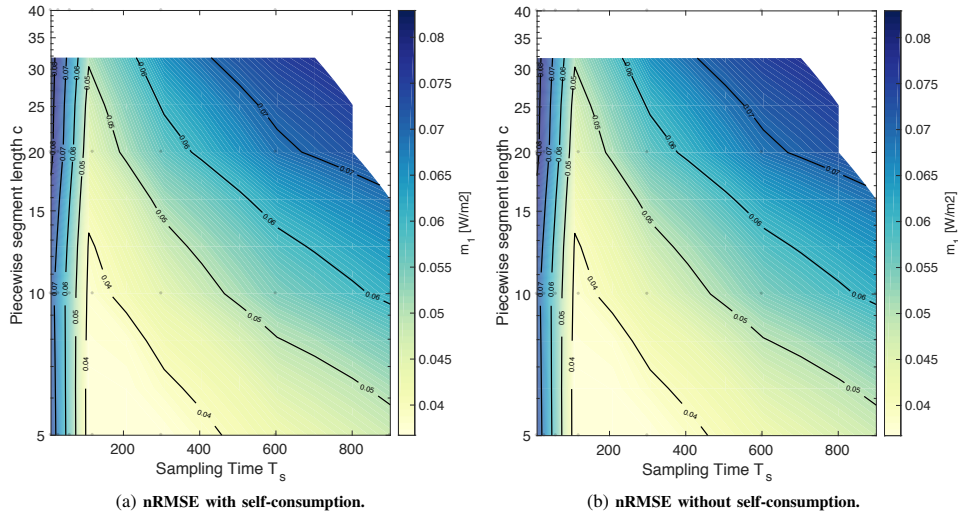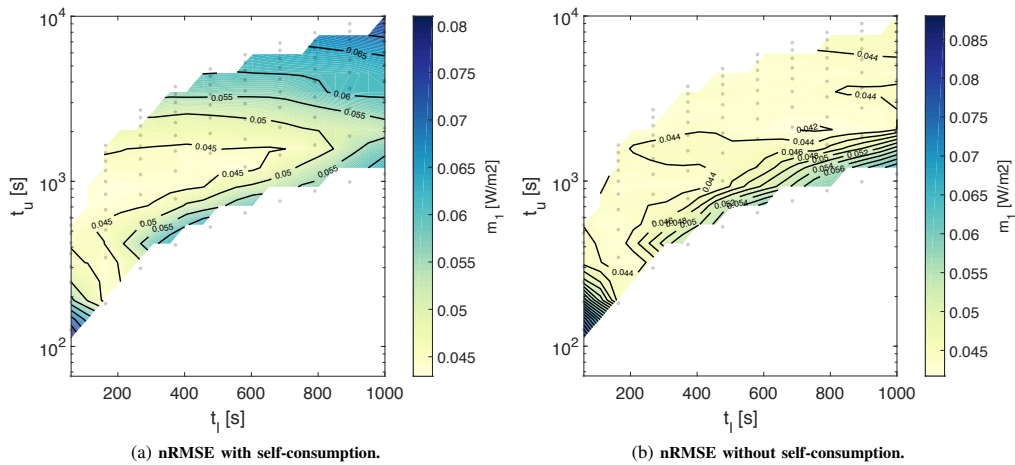
area of the parameter space. In other words, although it is not possible to derive analytical criteria for performing an a-priori assignment of the parameters, empirical results showed that, in the proposed case, parameters can be chosen in a wide range of values without a sensible deterioration of the performance.

G. Extension to cases with lower PV generation levels

In this section, the algorithms with the best performing parameters from the previous analysis are tested in scenarios with lower PV production capacity to verify how less prominent PV generation patterns are detrimental to their performance. In each scenario, synthetic time series of the power flow at the PCC are generated as the sum between the demand and a fraction of the original PV generation. Two additional scenarios are considered, with 50% and 25% of the original PV capacity (17.5 and 8.75 kW, respectively), which correspond to 146 and 76% PV penetration levels (i.e. installed PV capacity over the observed peak demand), respectively. Results are reported in Table III. Results show that lower PV penetration levels affects estimation performance of methods B and C (their performance worsens approximately by a factor 2 with a quarter of PV generation), whereas Method D is more robust and its performance is minimally affected.

TABLE III
ESTIMATION NRMSE (%) AT DIFFERENT LEVELS OF PV PENETRATION WITH SELF-CONSUMPTION.

Nominal PV Capacity (kWp)	PV Penetration (%)	A	B	C	D
35.3	294	5.2	5.2	**4.0**	4.3
17.6	146	9.4	6.9	9.7	**4.6**
8.8	73	18.7	10.4	11.8	**5.2**

H. Computational aspects

In the disaggregation and estimation process, the only required real-time operation is the computation of (1), a cheap task which involves algebraic and trigonometric relationships. Computing $\boldsymbol{\alpha}^o$ is a training process without real-time requirements which can be executed off-line with historical data. The computational time is 183 s for Method A, 709 s for B, 103 s for C and 67 s for D.[4]

VI. CONCLUSIONS

The problem of disaggregating a sequence of active power flow measurements composed of unobserved PV generation and demand into the respective trajectories was considered. Four disaggregation algorithms were discussed. They attempt to explain similarities between the time series of the aggregated and estimated PV generation, three in the time domain and one in the frequency domain. Estimation algorithms

[4]Computational times refer to a workstation equipped with an Intel Xeon processor running at 2.70 GHz with Matlab on a virtualized operating system. Method C and Method A,B and D were executed on two different machines, machine 1 and 2. The computation time of Method C was adjusted by a factor t_2/t_1, where t_1 and t_2 is the computation time of a reference problem executed on machine 1 and 2, respectively.

leverage GHI measurements transposed onto a number of tilted planes with the objective of explaining PV production patterns from sites with potentially different configurations (a key feature in urban/suburban context where PV generation is mostly from rooftop PV facilities with tilt/azimuth configurations dictated by roof characteristics). The effect of the air temperature was modelled by preprocessing GHI values with a model-based approach. Algorithms require an offline optimization problem-based training phase with historical data. For three algorithms, the convexity of the underlying optimization problem, important to assure tractability, is verifiable a-priori by inspecting the input data. Reconstructing the PV power output requires computing on-line an algebraic relationship and is suitable for implementation with deterministic deadlines and low processing power, in real-time target devices. Algorithm performance was tested with data from a real-life setup, with PV generation from multiple sites with different configurations, different demand profiles, and battery systems for PV self-consumption. Results show that the best performing algorithms estimate PV generation with a root mean square and mean estimation errors in the ranges $3.4 \div 8.8\%$ and $0.5 \div 2.3\%$, respectively, and that performance is minimally affected by the level of PV penetration in the prosumption mix. The practical utility of the proposed algorithms is envisaged in the context of power and energy management of distributed energy resources and data-driven PV generation forecasting in those situations where information from PV plants is not available due to issues such as privacy concerns or lack of adequate communication infrastructures.

APPENDIX A
ON THE CONVEXITY OF METHOD B AND C

We discuss on the convexity of the problem (9)-(11). Let $\boldsymbol{P} = [P_1, P_2, \cdots, P_K]$, $\boldsymbol{I}_j^{\searrow} = [I_{j,1}^{\searrow}, \cdots, I_{jK}^{\searrow}]$, $j = 1, \ldots, J$, $M \in \mathbf{R}^{K \times J} = \left(\boldsymbol{I}_1^{\searrow}, \boldsymbol{I}_2^{\searrow}, \ldots, \boldsymbol{I}_J^{\searrow} \right)$ the matrix obtained by stacking horizontally the GNI columns. The estimated total PV production (1) is (matrix product) $\widehat{\boldsymbol{G}} = M\boldsymbol{\alpha}$, which replaced into (4) yields:

$$\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{L}} - M\boldsymbol{\alpha} = \left(\mathbb{1}_{K \times K}, \quad -M \right) \begin{pmatrix} \widehat{\boldsymbol{L}} \\ \boldsymbol{\alpha} \end{pmatrix} = S\boldsymbol{x} \qquad (16)$$

where $\mathbb{1}$ is the $K \times K$ identity matrix, $S = \left(\mathbb{1}_{K \times K}, -M \right) \in \mathbf{R}^{K \times (K+J)}$ and $\boldsymbol{x} = (\widehat{\boldsymbol{L}}, \boldsymbol{\alpha})^T$. The least square cost (9) is:

$$J = (\boldsymbol{P} - \widehat{\boldsymbol{P}})^T (\boldsymbol{P} - \widehat{\boldsymbol{P}}) = \boldsymbol{P}^T \boldsymbol{P} + \widehat{\boldsymbol{P}}^T \widehat{\boldsymbol{P}} - 2\boldsymbol{P}^T \widehat{\boldsymbol{P}}, \quad (17)$$

Minimizing (17) is the same as minimizing (minimization is invariant under sum with constants and scale factors):

$$J = \widehat{\boldsymbol{P}}^T \widehat{\boldsymbol{P}} - 2\boldsymbol{P}^T \widehat{\boldsymbol{P}} = (S\boldsymbol{x})^T S\boldsymbol{x} - 2\boldsymbol{P}^T S\boldsymbol{x} = \qquad (18)$$

$$= \boldsymbol{x}^T S^T S\boldsymbol{x} - 2\boldsymbol{P}^T S\boldsymbol{x} = \frac{1}{2}\boldsymbol{x}^T H\boldsymbol{x} - \boldsymbol{f}^T \boldsymbol{x}, \qquad (19)$$

where (16) is used, $H = S^T S$ and $\boldsymbol{f} = S^T \boldsymbol{P}$. Eq. 19 is convex if H is semidefinite positive. Since H depends on input data, convexity cannot be enforced by construction, but it can be checked numerically. It was noted that adding a regularization term to the matrix quadratic matrix $H' = H + \beta \cdot \mathbb{1}$ ($\beta = 1 \times 10^{-4}$) helps to achieve convexity while not impacting substantially on algorithms performance.

Appendix A. Appendix A

REFERENCES

[1] M. Delfanti, M. Merlo, G. Monfredini, V. Olivieri, M. Pozzi, and A. Silvestri, "Hosting dispersed generation on italian mv networks: Towards smart grids," in *Proceedings of 14th International Conference on Harmonics and Quality of Power - ICHQP 2010*, Sept 2010, pp. 1–6.

[2] S. Hashemi, J. Østergaard, T. Degner, R. Brandl, and W. Heckmann, "Efficient control of active transformers for increasing the pv hosting capacity of lv grids," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 270–277, Feb 2017.

[3] O. Gagrica, P. H. Nguyen, W. L. Kling, and T. Uhl, "Microinverter curtailment strategy for increasing photovoltaic penetration in low-voltage networks," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 369–379, April 2015.

[4] M. N. Kabir, Y. Mishra, G. Ledwich, Z. Y. Dong, and K. P. Wong, "Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 967–977, May 2014.

[5] D. Vanhoudt, D. Geysen, B. Claessens, F. Leemans, L. Jespers, and J. Van Bael, "An actively controlled residential heat pump: Potential on peak shaving and maximization of self-consumption of renewable energy," *Renewable Energy*, vol. 63, pp. 531–543, 2014.

[6] F. Sossan, A. M. Kosek, M. Marinelli, and H. W. Bindner, "Scheduling of domestic water heater power demand for maximizing PV self-consumption using model predictive control," in *IEEE International Conference on Innovative Smart Grid Technologies (ISGT)*, 2013.

[7] R. Luthander, J. Widén, D. Nilsson, and J. Palm, "Photovoltaic self-consumption in buildings: A review," *Applied Energy*, vol. 142, 2015.

[8] S. Hashemi and J. Ostergaard, "Efficient control of energy storage for increasing the pv hosting capacity of lv grids," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[9] A. Molina-Garcia, R. A. Mastromauro, T. Garcia-Sanchez, S. Pugliese, M. Liserre, and S. Stasi, "Reactive power flow control for pv inverters voltage support in lv distribution networks," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, Jan 2017.

[10] M. Brenna, F. Foiadelli, M. Longo, and D. Zaninelli, "Energy storage control for dispatching photovoltaic power," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[11] F. Sossan, E. Namor, R. Cherkaoui, and M. Paolone, "Achieving the dispatchability of distribution feeders through prosumers data driven forecasting and model predictive control of electrochemical storage," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, Oct 2016.

[12] A. Bracale, G. Carpinelli, and P. D. Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, April 2017.

[13] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation – with application to solar energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, Sept 2016.

[14] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Advances in Neural Information Processing Systems*, 2010, pp. 1153–1161.

[15] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 2011, pp. 747–758.

[16] E. C. Kara, C. M. Roberts, M. D. Tabone, L. Alvarez, D. S. Callaway, and E. M. Stewart, "Towards real-time estimation of solar generation from micro-synchrophasor measurements," *CoRR*, vol. abs/1607.02919, 2016. [Online]. Available: http://arxiv.org/abs/1607.02919

[17] E. Scolari, F. Sossan, and M. Paolone, "Photovoltaic-model-based solar irradiance estimators: Performance comparison and application to maximum power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, Jan 2018.

[18] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and solar forecasting: state of the art," *IEA PVPS, Task*, vol. 14, 2013.

[19] J. E. Hay and J. A. Davies, "Calculation of the solar radiation incident on an inclined surface," in *Proc. of First Canadian Solar Radiation Data Workshop (Eds: JE Hay and TK Won), Ministry of Supply and Services Canada*, vol. 59, 1980.

[20] E. L. Maxwell, "A quasi-physical model for converting hourly global horizontal to direct normal insolation," Solar Energy Research Inst., Golden, CO (USA), Tech. Rep., 1987.

[21] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "Trend Filtering." *SIAM Review*, vol. 51, no. 2, pp. 339–360, 2009.

[22] R. J. Tibshirani, "Adaptive Piecewise Polynomial Estimation via Trend Filtering," *arXiv*, 2013.

[23] N. Hatziargyriou, *Microgrids: Architectures and Control*, ser. Wiley - IEEE. Wiley, 2014.

[24] D. Piga, A. Cominola, M. Giuliani, A. Castelletti, and A. E. Rizzoli, "Sparse Optimization for Automated Energy End Use Disaggregation," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, 2016.

[25] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Trans. Audio and electroacoustic*, vol. 15, pp. 70–73, 1967.

[26] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics - Theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[27] P. J. Huber and E. M. Ronchetti, *Robust Statistics 2e*, 2009, vol. 523, no. 3.

[28] J. A. Kratochvil, W. E. Boyson, and D. L. King, "Photovoltaic array performance model." Sandia National Laboratories, Tech. Rep., 2004.

[29] Photon. (2017) Photon databases of PV modules. https://www.photon.info/en/photon-databases.

Fabrizio Sossan is an Italian citizen and was born in Genova in 1985. He got his M.Sc. in Computer Engineering from the University of Genova in 2010, and, in 2014, the Ph.D. in Electrical Engineering from DTU, Denmark. Since 2015, he is a postdoctoral fellow at the Distributed Electrical Systems Laboratory at EPFL, Switzerland. In 2017, he has been a visiting scholar at NREL, Colorado, US. His main research interest are modeling and optimization applied to power system.

Lorenzo Nespoli received the M.Sc. degree in Energy Engineering from Politecnico di Milano in 2013. Since 2014 he works on thermodynamic simulations and electric grid optimization at SUPSI, as a scientific assistant. He is a Ph.D. candidate at the Ecole polytechnique federale de Lausanne at the IPESE department.

Vasco Medici received a M.Sc. in Micro-Engineering from the Ecole polytechnique federale de Lausanne and a Ph.D. from the Institute of Neuroinformatics at the University of Zurich and ETH. He actually leads the Intelligent Energy Systems Team at the Institute for Sustainability Applied to the Built Environment at SUPSI. He has experience in project management, system identification, algorithmics, modeling and simulation. He is the coordinator at SUPSI for the Swiss Competence Center for Energy Research on Future Swiss Electrical Infrastructure SCCER FURIES. His team runs a number of pilot projects in the field of smart grid, in close collaboration with industrial partners.

Mario Paolone was born in Italy in 1973. He received the M.Sc. (Hons.) degree in electrical engineering and Ph.D. degree from the University of Bologna, Bologna, Italy in 1998 and 2002, respectively. He was a Researcher of Electric Power Systems at the University of Bologna in 2005, where he was with the Power Systems Laboratory until 2011. In 2010, he was an Associate Professor with Politecnico di Milano, Milano, Italy. He is now a Full Professor with the Swiss Federal Institute of Technology, Lausanne, Switzerland, where he was the EOS Holding Chair of the Distributed Electrical Systems Laboratory. He has authored and co-authored more than 170 scientific papers published in reviewed journals and presented at international conferences. His current research interests include power systems with particular reference to real-time monitoring and operation, power system protections, power systems dynamics, and power system transients. He is the Secretary and a member of several IEEE and Cigré Working Groups. He was a recipient of the IEEE EMC Society Technical Achievement Award in 2013.

B Appendix B

Paper B: *L. Nespoli and V. Medici, "An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements," Sol. Energy, 2017.* List of authors:

- Nespoli, Lorenzo: Swiss Federal Institute of Technology of Lausanne (EPFL), University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

- Medici, Vasco: University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

Personal contributions:

- State of the art, theory, experimental results.

An Unsupervised Method for Estimating the Global Horizontal Irradiance from Photovoltaic Power Measurements

Lorenzo Nespoli[a,b,*], Vasco Medici[b]

[a]Swiss Federal Institute of Technology in Lausanne, Switzerland
[b]University of Applied Sciences and Arts of Southern Switzerland, Lugano, Switzerland

Abstract

The precise calculation of solar irradiance is pivotal for forecasting the electric power generated by PV plants. However, on-ground measurements are expensive and are generally not performed for small and medium-sized PV plants. Satellite-based services represent a valid alternative to on-site measurements, but their space-time resolution is limited. In this paper we present a method for estimating the global horizontal irradiance (GHI) from the power measurements of one or more photovoltaic (PV) systems located in the same neighborhood. The method is completely unsupervised and is based on a physical model of a PV plant. It can estimate the nominal power and orientation of multiple PV fields, using only the aggregated power signal from their PV power plant. Moreover, if more than one PV power plant is available, the different signals are reconciled using outliers detection and assessing shading patterns for each PV plant. Results from two case studies located in Switzerland are presented here. The performance of the proposed method at estimating GHI is compared with that of free and commercial satellite services. Our results show that the method presented here is generally better than satellite-based services, especially at high temporal resolutions.

Keywords: solar radiation, signal estimation, photovoltaic power systems, numerical optimization

1. Introduction

1.1. Motivation

Solar power generation, both at utility and residential level, will play a central role in the future of the electric power industry, with a predicted installed power ranging from 4.3 to 14.8TW by 2050 [1]. Although this trend is certainly to be welcomed, unless countermeasures are taken the intermittent nature of solar generation could lead to stability issues in the electrical grid [2]. In the distribution grid, these problems will be further emphasized by the increase of electricity consumption driven by the electrification of heat generation and mobility [3], which will further increase the amplitude of the power and voltage fluctuations [4].

Fortunately, in the meanwhile smart grid solutions that help to overcome the above-mentioned issues by modulating generation and demand are becoming available and affordable. For example, distributed energy storage systems [5] and demand side management [6, 7] can be exploited for the active control of distribution networks.

Nomenclature

α	pv field tilt [rad]
α_{pr}	proxy tilt [rad]
β	pv field azimuth [rad]
β_{pr}	proxy azimuth [rad]
δ	fixed GHI increment for the computation of derivatives [W/m^2]
η	combined module and inverter efficiency [$-$]
γ	PV power temperature coefficient [1/K]
γ_s	sun azimuth [rad]
λ	stepsize for the optimization algorithm [$-$]
ω	coefficients related to **Pr** [m^2]
E	estimation error matrix [W]
e	PV estimation error [W]
Pr	proxy matrix [W/m^2]
$\mathcal{F}$	trust function
$\mathcal{I}$	outlier detection function
$\mathcal{L}$	robust loss function
$\mathcal{N}$	normal distribution
AOI	angle of incidence [rad]

*Corresponding author
Email address:* lorenzo.nespoli@epfl.ch (Lorenzo Nespoli)

IAM	incident angle modifier [−]
I	irradiance on an oriented surface [$W/m2$]
I^{AOIT}	irradiance corrected with the angle of incidence and temperature [W/m^2]
I^{AOI}	irradiance corrected with the angle of incidence [W/m^2]
I_b	beam component of the irradiance on an oriented surface [$W/m2$]
I_d	diffuse irradiance on an oriented surface [$W/m2$]
I_g	ground reflected irradiance on an oriented surface [$W/m2$]
I_{STC}	reference irradiance [W/m^2]
Pr	proxy [W/m^2]
re_{clear}	relative PV estimation error with respect to clear sky condition[−]
μ	mean of distribution
ν	current iteration [−]
ω	coefficient associated to Pr [m^2]
ϕ	PV cell temperature correction coefficient [Km^2/W]
ρ	albedo [−]
σ	standard deviation of the distribution
θ_z	sun zenith [rad]
$\|\cdot\|_{fr}$	Frobenius norm
$\widehat{P}$	estimated power (W)
g	vector of geographical coordinates
h	objective function
n_{pv}	number of PV power signals [−]
P^n	normalized power signal [−]
P_{nom}	PV field nominal power [W]
T	number of time steps (-)
T_a	ambient temperature [°C]
T_{cell}	PV cell temperature [°C]
T_{ref}	reference temperature [°C]
DHI	diffuse horizontal irradiance [W/m^2]
DNI	direct normal irradiance [W/m^2]
GHI	global horizontal irradiance [W/m^2]
GHI*	optimized value of GHI [W/m^2]

To optimize the control and guarantee the quality-of-service in the electricity grid, it is important to predict the power flows accurately, as this favors a sensible management of the available flexibility. It has been shown that forecasting accuracy can be improved when the production and consumption in the grid are disaggregated and predicted separately [8, 9]. The disaggregation of solar generation from the total grid load can be achieved by using on ground irradiance measurements [10]. In general, global horizontal irradiance (GHI) measurements are often used as exogenous input when performing both long term and short term PV production forecasts [11, 12]. Unfortunately, accurate on ground irradiance measurements are often not available. Although irradiance measurements can also be used for the online estimation of PV power production and for fault detection, sensors are usually not installed for small and medium-sized plants, due to their high cost. If the irradiance is not measured directly by means of on-ground measurements, satellite estimations can be exploited. Satellite-based radiation assessment services provide an estimate of the time course of GHI for a given location, but their spatio-temporal resolution is constrained by technical limitations. Most of these services are based on the images acquired by the Meteosat 2nd generation satellites, which have a spatial resolution of 3 km at the nadir and a temporal resolution of 15 minutes [13]. These coarse resolutions limit the performance of satellite-based nowcasting methods. Moreover, the limited spatial resolution has a smoothing effect that can result in reduced accuracy levels for GHI estimation at a specific location, especially in the presence of local clouds. The active control of distribution networks, of which some of the critical sections can take up a small area, requires a more accurate and fast estimate of GHI. Satellite-based methods could also profit from an increased availability of on-ground GHI measurements, as they could be used for calibration [14, 15, 16], a technique also known as site adaptation.

In this study, we investigate the possibility of using local PV power measurements to estimate GHI with a high temporal and spatial accuracy. Being able to estimate GHI directly from PV power measurements will allow to better estimate and forecast the PV production of an entire neighborhood by monitoring the power output of only a small fraction of the PV plants, without the need to install additional irradiance sensors.

For the development of the proposed method, we focused on accessibility and simplicity. Indeed, the method is fully unsupervised and the only inputs it requires are the measurements of the AC power output of the monitored PV plants, the ambient temperature and the geographical coordinates of the neighborhood.

1.2. Previous work

The idea of using PV plants as surrogated irradiation sensors has already been researched in the past. In [17], voltage and current measurements from a PV module were used to calculate the incoming radiation on the plane of array. In [18], it is suggested that a nearby PV plant can be used as a proxy to estimate the power generated by another PV installation, even if only a linear relationship is considered between the proxy and the estimated output. In [19], proximate PV plants are used to

2

predict the PV outputs of other PV installations, for the purposes of automatic fault detection. In [20], a clear sky index, K_{pv}, is introduced: this is the ratio between the AC power of a simulated PV plant under clear sky conditions and the actual power measurements. The authors use a clear-sky radiation model, a transposition model and an inverter and PV module model. This method relies on an accurate, technical description of the PV system, which includes the PV module orientation. In [21], a methodology is proposed to project power generation between different PV systems. The PV system power output is modeled as a quadratic function of the solar irradiance on the plane of array (POA) and the ambient temperature. The five coefficients of the curve must be fitted for each type of PV module technology. The POA irradiance is obtained by inverting the quadratic expression. The GHI is then estimated from the POA irradiance, using an iterative procedure. In their discussion, the authors suggest that simultaneously considering PV systems with multiple orientations could increase the accuracy of the GHI estimation. In [22], a similar methodology is presented for obtaining GHI from PV power measurements. Building on the work of [21], a correction for low angle of incidence and wind speed is considered. The AOI correction is based on 18 coefficients, specific to the type of PV module coating considered. In all the aforementioned studies, namely [20, 21, 22], it is assumed that the inverter type, PV module type and PV module orientation are known. Furthermore, it is assumed that all the modules of a given PV plant have the same orientation.

1.3. Outline and objective

Despite the increasing number of PV installations and the abundance of available monitoring data, it is difficult to use them to estimate the GHI signal. Considering the works previously cited in subsection 1.2, we can identify three main causes that make this task particularly challenging:

1. Most of the estimation methods in the literature require a detailed description of the PV systems, including PV power plant nominal power, fields orientations, module and inverter types. Gathering all these metadata is difficult and time consuming. Moreover, when available, the data contained in databases could be imprecise or flawed [23]. This could lead to erroneous estimations.
2. Occasionally, PV power plants can be composed of different PV fields, each field having different nominal power and orientations. A typical case is a PV power plant with an east-west configuration, but more complex configurations are possible, as presented later in the paper. In this cases, if we possess only a single power signal, we should be able to retrieve nominal powers and orientations of an arbitrary number of PV fields to correctly estimate GHI.
3. If more than a power signal is available for a given geographical location, an automated procedure is needed to reconcile all the measurements and efficiently make use of all the signals.

We present here a fully unsupervised method for estimating the local GHI using only the power measurements from one or more PV plants, without the need to know their nominal power and module orientations. The problem of identifying PV plants with differently oriented modules from a single power signal is addressed by means of a robust regression. In order to increase the accuracy of the GHI estimation, the method can exploit multiple power signals from different PV plants. The different signals are reconciled by means of outlier detection and by determining the shading patterns of each PV plant. The code related to the GHI estimation method, including the PV system identification methodology, is freely available online (see Section 7).

The paper is structured as follows: Section 2 describes the methodology used for estimating the GHI. Section 3 discusses the issue of how to identify the PV plant orientations without knowing the actual GHI. Section 4 discusses the models used to obtain the PV power production proxies. Section 5 briefly discusses the numerical methods for solving the GHI estimation problem. In Section 6, the accuracy of the method is assessed for two case studies, and compared to satellite-based GHI estimations and secondary standard pyranometer measurements. Finally, conclusions are presented in Section 7.

2. Methodology

The combined effect of irradiance and ambient temperature on the PV power production is well-known. Accurate empirical models that assess the total incoming irradiation on an oriented surface, given the GHI, are also available [24, 25]. We can therefore build a function that links the GHI to a given PV plant power output:

$$\widehat{\mathbf{P}} = f(\mathbf{GHI}, \mathbf{t}, \alpha, \beta, \mathbf{g}, \mathbf{T}_a, \mathbf{P_{nom}}) \tag{1}$$

where $\widehat{\mathbf{P}} \in \mathbb{R}^{T \times 1}$ is the estimate of the power generated by a given PV plant, where T is the number of time steps in the data, $\mathbf{GHI} \in \mathbb{R}^{T \times 1}$ and $\mathbf{T}_a \in \mathbb{R}^{T \times 1}$ are the vectors of the observed GHI and temperatures at times $\mathbf{t} \in \mathbb{R}^{T \times 1}$, α and β are the vectors containing the tilts and azimuths of the modules, $\mathbf{g}$ is a vector containing the geographical coordinates of the plant, namely latitude, longitude and elevation, and $\mathbf{P_{nom}}$ is the vector of the nominal powers of the modules. Function f is described later by equations 7-16 and by the empirical disc model, as stated in Section 4. If the module orientations and nominal powers were known, f could be inverted in order to estimate GHI. Unfortunately, f is not always invertible, especially when $\widehat{\mathbf{P}}$ comes from a PV plant with differently oriented PV modules. So, for different values of GHI, the function f could return the same output P. In this case, a method is required in order to choose the correct GHI value from a range of possible choices. This problem is solved by following two steps. First, we estimate the panel orientation from the measured power of the given PV plant. We then use the calculated orientations to build function f and solve

$$\mathbf{GHI}^{\star} = \underset{\mathbf{GHI}}{\arg\min} \|\mathbf{P} - \widehat{\mathbf{P}}\|_2 \tag{2}$$

without inverting f. Here $\mathbf{GHI}^{\star}$ refers to the optimized values of GHI.

3

Equation 2 can be solved by using one or more PV power signals and can therefore be easily reformulated as:

$$\mathbf{GHI}^{\star} = \arg\min_{\mathbf{GHI}} \left\| \frac{1}{n_{pv}} \sum_{i=1}^{n_{pv}} \mathbf{P}_i^n - \widehat{\mathbf{P}}_i^n \right\|_2 \tag{3}$$

where n_{pv} is the total number of PV power signals and $\mathbf{P}_i^n$ and $\widehat{\mathbf{P}}_i^n$ are the observed and estimated power signals normalized with the estimated nominal power (see Section 3 for how this is estimated). The main drawback to this formula is that, when estimating GHI, all the PV signals are equally weighted. This solution is not robust in the event of shadows or faulty signals. Equation 3 can be improved in two ways:

1. Faulty signals can be statistically detected, to avoid using them for estimating GHI
2. When the modules are partially or completely shaded, the GHI estimation is not accurate. This effect can be mitigated by building a map of the error $\mathbf{e}_i = \mathbf{P}_i^n - \widehat{\mathbf{P}}_i^n$, as a function of the sun position (azimuth and elevation). This map can then be used to evaluate how much a certain PV measurement can be trusted as a function of sun position. For the rest of the paper, we will refer to this map as the trust function.

The above considerations lead to the more general formula:

$$\mathbf{GHI}^{\star} = \arg\min_{\mathbf{GHI}} \left\| \frac{1}{n_{pv}} \sum_{i=1}^{n_{pv}} \mathcal{I}(\mathbf{E})\mathcal{F}_i(\boldsymbol{\gamma}_s, \theta_z)\mathbf{e}_i \right\|_2 \tag{4}$$

where $\mathcal{I} : \mathbb{R} \rightarrow \{0,1\}$ is an indicator function detecting the presence of outliers, $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ...\mathbf{e}_N]$ is the estimation error matrix, $\mathcal{F} : \mathbb{R} \rightarrow [0,1]$ is the trust function and $\boldsymbol{\gamma}_s$ and θ_z are, respectively, the azimuth and zenith of the sun. Here $\mathcal{F}$ can be interpreted as a dynamic weight function, since $\boldsymbol{\gamma}_s$ and θ_z are function of time. The role of the trust function is to attach less importance to the calculation of the ith signal if this is believed to be affected by shadows with a particular position of the sun. The construction of the $\mathcal{I}$ and $\mathcal{F}$ functions is described in Section 5.

3. Orientation assessment

PV plant orientation could theoretically be estimated from the PV plant power measurements, and from the GHI measurements, by means of the equations 7, 8, 9 and 10. The orientation estimation can be formulated as the following optimization problem:

$$\underset{\alpha_i,\beta_i}{\text{minimize}} \, \|\mathbf{P}_i - \widehat{\mathbf{P}}_i\|_2 \tag{5}$$

where $\widehat{\mathbf{P}}_i$ is the PV production estimated from the GHI signal. The following aspects must be taken into account:

1. We want to estimate PV plant orientation without knowing the actual GHI seen by the modules
2. PV plants can consist of groups of modules with different orientations, e.g. plants with an east-west configuration

3. The presence of shadows affects the relationships between the GHI projection on an oriented surface and the PV power output
4. Problem 5 is non-linear and non-convex

If the GHI seen by the modules is unknown, estimating their orientation would result in a blind identification problem [26, 27]. We exploit the fact that we can obtain a good approximation of GHI for clear-sky periods, using a model for the extra-terrestrial irradiation and for the air mass index. In this paper, we used time series obtained from the Soda-pro CAMS Mc-Clear service[1], which uses the McClear clear sky model [28].

We can thus identify the PV plant orientations if we can select clear-sky periods, using only the PV plant power output. Different methods can be used to exploit PV power signals in order to identify clear-sky radiation periods. In [29], a period is considered to be clear if the measured PV power is higher than the 80% percentile of the set of measurements taken at the same time of day, during the previous 15 days. In [23] this method is combined with the clear-sky detection routine described in [30], which uses GHI observation as input and a set of 5 extraction parameters. In this paper, we first developed a selection based on the smoothed power signals: the power output of each PV plant is filtered using a second order low-pass Butterworth filter [31]. We then considered a period to be clear if the root mean squared relative error between the original and filtered signal was lower than a threshold value.

The main drawback of this method is that its performance is influenced by three parameters, namely the threshold value, the low-cut frequency period and the length of the period, which have to be tuned. Moreover, in the event of PV curtailment, these curtailment periods can be identified as clear periods. In order to overcome these issues, we developed a different method. PV power signal distribution as a function of the sun position is typically bimodal, due to the presence of clouds during data acquisition. On the other hand, a unimodal distribution could indicate a systematic shadow for the corresponding sun position. In order to select clear data periods, we fit a gaussian mixture probability density function with two components $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ for each sun position, with a discretization of $5°$. Then, for each sun position we identify the observations lying in the one sigma interval of the gaussian distribution with the largest μ as clear observations. We chose to discard other values since they could have been potentially caused by cloud enhancement events (higher power) or by the presence of haze or high clouds (lower power). An example of a PV power distribution for a particular sun position is depicted in figure 1. Despite the second model being more robust in terms of selection of clear sky periods, the task of predicting the GHI seen by PV panels through a clear sky model presents some intrinsic errors. Clear sky models are not perfect and it is not possible to guarantee that the GHI seen by the PV panels is exactly the one predicted by the clear sky models. In order to overcome this problem and the others referred to above, instead of di-

---

[1]http://www.soda-pro.com/web-services/radiation/cams-mcclear

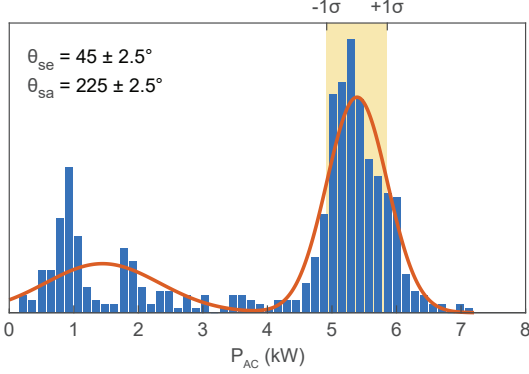rectly solving the optimization problem 5, we reformulated it as a robust linear regression:

Figure 1: Distribution of the power signal of one particular PV power plant of the Biel-Benken case study, for $\theta_z = 45 \pm 2.5°$ and $\gamma_s = 225 \pm 2.5°$. The distribution presents a bimodal shape. We fit two gaussian distributions and considered the leftmost peak to be caused by clouds, while we select the data between one standard deviation from the mean of the rightmost gaussian distribution as belonging to clear sky periods. Values outside the one standard deviation interval were discarded, since they could have been potentially generated during cloud enhancement events (higher power) or in the presence of haze or high clouds (lower power).

$$\underset{\omega_i \in \mathbb{R}^{n_p}_+}{\text{minimize}} \; \mathcal{L}\left(\mathbf{P}_i - \mathbf{Pr}\omega_i\right) \qquad (6)$$

where $\mathcal{L}$ is a robust loss function [32], $\mathbf{Pr} \in \mathbb{R}^{T \times n_p}$ is the proxy matrix, each proxy being the estimated power produced by a panel with a given orientation, $\omega_i \in \mathbb{R}^{n_p}_+$ is the coefficient vector for the ith PV installation, n_p and T being the number of proxies and total number of temporal observations. The additional requirement $\omega_i \in \mathbb{R}^{n_p}_+$ forces vector ω_i to have all positive values. Thus, we can interpret the components of vector ω_i as coefficients describing the significance of each proxy in explaining the power output of the ith PV plant, rescaled for its nominal power. Another interpretation is that non-zero entries of ω_i are the estimates of the nominal power of the ith PV plant oriented as its corresponding proxies. Furthermore, problem 6 forces ω_i to be sparse and it is robust with respect to the presence of partial shadows.

If we use a loss function of an M-estimator for $\mathcal{L}$, we can solve problem 6 using an efficient iterative reweighted least square algorithm [33]. As previously anticipated, we could use more than one PV signal to estimate GHI. In this case, we need to identify a set of coefficients for each of the n signals, and some methods to assign different weights to the estimation errors arising from the different PV signals, in order to calculate GHI more accurately.

4. Proxy Model for PV Performance

The proxies are an estimate of the electrical power generated by a panel with a given orientation and GHI. In order to

effectively solve problem 5, we need to select the most representative proxy orientations. The tilts and azimuths of the proxies, respectively $\alpha_{pr,i}$ and $\beta_{pr,i}$, are obtained by generating a triangular mesh of an icosahedron on a unit sphere. This is later refined through subdivisions, in order to increase the number of points. In this paper, the most north-facing orientations are discarded, as shown in figure 6.

The sun azimuth and elevation are calculated based on the current time and the altitude, longitude and latitude of the given location. For this task we have used the `pvl_ephemeris.m` matlab function from the freely available Sandia National Laboratories PV Collaborative Toolbox [34], which is based on the 1985 Grover Hughes' Engineering Astronomy course at Sandia National Laboratories. The direct normal irradiance (DNI) is then calculated by means of the empirical disc model [35]. The diffuse horizontal irradiance at time t DHI_t is then calculated as:

$$DHI_t = GHI_t - \cos\left(\theta_{z,t}\right)DNI_t \qquad (7)$$

where $\theta_{z,t}$ is the zenith angle of the sun at time t. DHI is then used to estimate the projection of the diffuse radiation on the given surface I_d, using the Hay and Davies' model [25]. The overall radiation on the given surface is then given by the sum of the diffuse, direct and ground-reflected radiation.

$$I_{i,t} = I_{b,i,t} + I_{d,i,t} + I_{g,i,t} \qquad (8)$$

where I_g is the ground reflected component, calculated as:

$$I_{g,i,t} = \rho GHI_t \frac{(1 - \cos(\alpha_i))}{2} \qquad (9)$$

where ρ is the albedo, which was fixed to a typical value of 0.2. The direct irradiation on the oriented surface I_b is obtained from the DNI:

$$I_{b,i,t} = DNI_t \cos(AOI_{i,t}) \qquad (10)$$

where AOI_i is the angle of incidence of the oriented surface i. To calculate DNI and $I_{d,i,t}$ we used the PV Performance Modeling Toolbox by Sandia National Laboratories [34]. Since reflection losses can significantly increase at large AOI [22], we applied an AOI correction, independent from the module technology [36]:

$$I_{i,t}^{AOI} = IAM_{i,t}I_{b,i,t} + 0.95(I_{d,i,t} + I_{g,i,t}) \qquad (11)$$

where IAM is the angle of incidence modifier. We use the following ASHRAE approximation [37]:

$$IAM_{i,t} = \max\left(1 - k_1\left(\cot(\min(AOI_{i,t}, \pi/2)) - 1\right), 0\right) \qquad (12)$$

and k_1 is 0.05.

Finally, in order to obtain a proxy for the electrical power produced by a field with the ith orientation, we apply a correction taking into account the ambient temperature and the inverter and module efficiencies. The cell temperature is first estimated from the ambient temperature, then a linear correction is applied [38]:

$$T_{cell,i,t} = T_{a,t} + \phi I_{i,t}^{AOI} \qquad (13)$$

$$I_{i,t}^{\text{AOIT}} = I_{i,t}^{\text{AOI}} \left[1 + \gamma (T_{cell,i,t} - T_{ref}) \right] \tag{14}$$

$T_{ref} = 25°C$ a reference temperature, ϕ and γ two coefficients. In this study, ϕ and γ are not estimated and are set respectively to the values of 3.14e-2 $[Km^2/W]$ and -4.3e-3 $[1/K]$, which represent crystalline silicon framed PV modules. Finally, the proxies are corrected for the module and inverter efficiencies, using the following equation:

$$\text{Pr}_{i,t} = \eta_t I_{i,t}^{\text{AOIT}} \tag{15}$$

where η_t is the combined module and inverter efficiency. In order to reduce the number of parameters, we modeled it as a function of the irradiance $I_{i,t}^{AOIT}$ using the following equation:

$$\eta_t = k_2 + k_3 \ln(I_{i,t}^{\text{AOIT}}/I_{STC}) + k_4(\ln(I_{i,t}^{\text{AOIT}}/I_{STC}))^2 \tag{16}$$

where $I_{STC} = 1000\text{W/m}^2$ is the reference irradiance and k_2, k_3, k_4 are free parameters. By fitting equation 16 to typical inverter and polycrystalline module data, we obtained the following values: $k_2 = 0.942$, $k_3 = -5.02\text{e-}2$, $k_4 = -3.77\text{e-}2$.

5. Numerical optimization

5.1. Algorithm description

As stated in Section 2, function f is not always invertible. For this reason, we solve 2, or 4 if more than one PV signal is available. These minimizations can be naturally decomposed in time: that is, the norm operator can be written as the sum of the objective functions related to a single observation in time. To keep the discussion general, we can restate the left-hand side of equations 2, 3 and 4 as:

$$\underset{\textbf{GHI}}{\arg\min}\, h(\textbf{GHI}) \tag{17}$$

where h is a placeholder for one of the objective functions defined in equations 2, 3 or 4. The overall objective function can be restated as

$$h(\textbf{GHI}) = \sum_{t=1}^{T} h_t(\text{GHI}_t) \tag{18}$$

where T is the total number of observations. Derivative-free optimization algorithms such as genetic algorithms, the Nelder Meads simplex method and particle swarm optimization algorithms are badly affected by increasing numbers of decision variables [39, 40]. General purpose nonlinear solvers usually rely on calculating the objective function derivatives for all the values of the decision function. This means that $\frac{\partial h(\text{GHI}_i)}{\text{GHI}_j}$ must be calculated at each step. Even if it is possible to specify a pattern for the Hessian matrix to the trust-region-reflective algorithm in Matlab, which would significantly speed up the optimization, this algorithm requires the analytical gradient for h, which we do not possess [41].

For this reason, we implemented a solver for our problem. A comparison between our solver and fmincon computational time, when fmincon solves 4 for each time step individually, is shown in table 1. The results are related to 1500 points. The relative

Table 1: Computational times comparison

	mean s/sample	std s/sample
fmincon	2.32	3.1e-2
our solver	1.8e-2	8.5e-3

difference in the solutions was below 1% with associated standard deviation of 2.02e-1 %. Our algorithm took approximately 15.7 minutes to process one year of data with a temporal resolution of 10 minutes, on an Intel Xenon CPU E5-2697 v2 @ 2.70 GHz with 32.0 GB of RAM.

Since our objective function h is in the form described in 17, our solver simply minimizes $h(x_i) \quad \forall i \in [1, T]$ with a steepest-descent solution strategy. Since $h(GHI_i)$ could present local minima as previously stated in 2, we initialized the solution with a grid search over the possible values of GHI, as shown in the pseudocode 1.

For each time step, we searched in a discrete space of possible values of GHI, uniformly sampled from 0 to

$$\text{GHI}_{max,t} = k\text{GHI}_{clear,t} \tag{19}$$

where $\text{GHI}_{clear,t}$ is the GHI calculated from the clear sky model, and k is a safety factor accounting for the fact that particular cloud configurations can increase the measured GHI to above the clear sky values [42]. We use a 30-step discretization for the grid search. Considering a standard irradiation of 1000 W/m2, this would result in an approximate accuracy level of 33.3 W/m2. We obtain a set of n_g guess vectors for GHI, and for each of them we calculate the proxies and assess the hypothetic power produced by each PV plant (line 3 and 4 of algorithm 1), and the PV estimation error matrix $\mathbf{E}_g \in \mathbb{R}^{T \times n_{pv}}$ (line 5). Then, for each time t we find the best guess GHI_t^*, which is the one that minimizes the average PV estimation error among the different PV plants (line 8-11 of algorithm 1). Note that the inner minimization of line 9 is inexpensive, since it consists of finding the position of the minimum element of the average of $E_{g,t,i}$ over g, at a given t, where i refers to the ith PV plant.

Once the initial guess for GHI has been obtained, the solution is refined as shown in the pseudocode 2. Starting from the first guess solution, we once again determine the PV estimation error at iteration ν, and then calculate the gradient of the objective function introduced in 4 with respect to GHI:

$$\nabla_{\textbf{GHI}}\mathbf{h} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{I}(\mathbf{E})\mathcal{F}_{i,t}\nabla_{\textbf{GHI}}\mathbf{e}_i \tag{20}$$

where

$$\nabla_{\textbf{GHI}}\mathbf{e}_i = -\left(\nabla_{\textbf{GHI}}\mathbf{Pr}\right)\omega_i \tag{21}$$

note again that, since the function is time-separable, the only non-zero elements of $\nabla_{\textbf{GHI}}\mathbf{Pr}$ are those related to observations occurring at the same time-step

$$\frac{\partial \text{Pr}_{i,j}}{\partial \text{GHI}_k} = 0 \qquad \forall \quad j \neq k \tag{22}$$

6

For this reason, the resulting tensor can be rewritten in matrix form such that $\nabla_{\mathbf{GHI}}\mathbf{Pr} \in \mathbb{R}^{T,n_p}$. Each element of $\nabla_{\mathbf{GHI}}\mathbf{Pr}$ is calculated as

$$\frac{\partial \Pr_{i,j}}{\partial \text{GHI}_i} = \frac{\Pr_{i,j}^+ - \Pr_{i,j}}{\delta_{ghi}} \tag{23}$$

where $Pr_{i,j}^+$ is the jth proxy at ith observation computed from $\text{GHI}^+ = \text{GHI} + \delta_{ghi}$.

Algorithm 1 Initialize GHI

1: **for** $g \in [1, n_g]$ **do**   ▷ grid search initialization
2:      $\mathbf{GHI}_g = g/n_g \mathbf{GHI}_{max}$ ▷ linear rescale
3:      $\mathbf{Pr}_g \leftarrow \mathbf{GHI}_g$
4:      $\widehat{\mathbf{P}}_g = \mathbf{Pr}_i \mathbf{\Omega}$
5:      $\mathbf{E}_g = \mathbf{P} - \widehat{\mathbf{P}}_g$
6: **end for**
7: Find GHI_g that minimize the mean PV error at each t
8: **for** $t \in [1, T]$ **do**
9:      $g^\star \leftarrow \arg \min_g \frac{1}{n_p} \sum_{p=1}^{n_p} E_{g,t,i}$
10:      $\text{GHI}_t^\star \leftarrow \text{GHI}_{g^\star, t}$
11: **end for**

Algorithm 2 Estimate GHI

1: **while** $err_\nu \leq err_{\nu-1}$ **and** $\nu < \nu_{max}$ **do**
2:      $\mathbf{Pr}_\nu \leftarrow \mathbf{GHI}_\nu$
3:      $\widehat{\mathbf{P}}_\nu \leftarrow \mathbf{Pr}_\nu \mathbf{\Omega}$
4:      $\mathbf{E}_\nu \leftarrow \mathbf{P} - \widehat{\mathbf{P}}_\nu$ ▷ PV estimation error
5:      $\nabla_{\mathbf{GHI}}\mathbf{h}_\nu \leftarrow \mathbf{GHI}_\nu + \delta_{ghi}, \mathcal{I}(\mathbf{E}), \mathcal{F}_{i,t}$
6:      $err_\nu = \|\mathbf{E}\|_{fr}$
7:      **for** $t \in [1, T]$ **do**
8:          $\epsilon_{t,\nu} = \frac{1}{N} \sum_{i=1}^N |(\mathbf{e}_{i,\nu} - \mathbf{e}_{i,\nu-1})\mathcal{I}(\mathbf{E})\mathcal{F}_{i,t}| \leq 0$
9:          **if** $\sim \epsilon_{t,\nu}$ **then**
10:             $\lambda_{t,\nu+1} = k\lambda_{t,\nu}$
11:             $\lambda_{t,\nu} = 0$
12:          **end if**
13:      **end for**
14:      $\mathbf{GHI}_{\nu+1} = \mathbf{GHI}_\nu - \lambda_\nu \nabla_{\mathbf{GHI}}\mathbf{h}_\nu$
15: **end while**

Lines 7 to 13 in algorithm 2 describe the λ_ν update strategy, where λ_ν is a vector of coefficients describing how much GHI must be shifted in the direction of the objective function gradient, at the ν iteration. Instead of using a backtracking strategy, which performs a line search on parameter $\lambda_{t,\nu}$, we applied an exponential decay on $\lambda_{t,\nu}$, in the attempt to reduce the total number of function evaluations. Since the objective function is not monotone in GHI, at each iteration ν we check if the mean estimation error has decreased. In this case, $\lambda_{t,\nu}$ is unchanged, otherwise $\lambda_{t,\nu}$ is set to zero (which results in not updating $\text{GHI}_{\nu,t}$) and the new $\lambda_{t,\nu+1}$ decreases by a factor $k < 1$.

5.2. Trust function and outlier detection

We now illustrate the method used to construct the trust function $\mathcal{F}(t)$ and the outlier detection function $\mathcal{I}(\mathbf{E})$. As previ-

ously stated in Section 2, $\mathcal{F}(t)$ weights the PV estimation differently based on the sun position. Greater importance is attached to signals with lower relative estimation errors during clear sky conditions for a given sun azimuth and elevation. Since we do not possess the real GHI, clear sky periods must be estimated. In order to use as much data as possible for shadow detection, a different method from the one introduced in 3 is used. We estimate the relative PV estimation error under clear sky conditions as:

$$\text{re}_{clear,i}(\boldsymbol{\gamma}_s, \boldsymbol{\theta}_z) = Q_{0.01,\boldsymbol{\gamma}_s,\boldsymbol{\theta}_z}\left(\frac{\widehat{\mathbf{P}}_i(\mathbf{GHI}_{clear}) - \mathbf{P}_i}{\mathbf{P}_i}\right) \tag{24}$$

where $Q_{0,01}$ is the 1% quantile, $\boldsymbol{\gamma}_s$ and $\boldsymbol{\theta}_z$ are the sun azimuth and zenith angles, discretized with a $2°$ step.

Since the $\text{re}_{clear}(\gamma_{z,t}, \theta_{z,t})$ map is assumed to be affected by noise, in order to obtain a more significant representation of the shadow pattern, we fit a gaussian process on top of it. We then apply a threshold to eliminate the lowest values in the map, which could be due to the lack of observed clear sky conditions in the corresponding sun position. An example of the resulting thresholded map for a given PV plant is shown in figure 2.
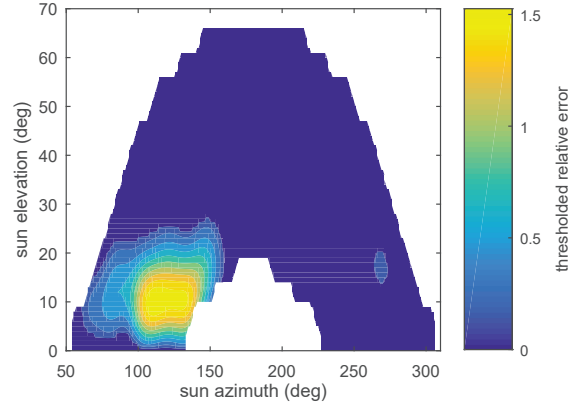
Figure 2: Thresholded smoothed estimated error as a function of sun position, for a given PV plant. This particular PV plant presents significant shadowing during morning hours

Once the PV estimation error has been established, we use it to map those signals that are more accurately calculated in a given combination of $\gamma_{s,t}$ and $\theta_{z,t}$, through an inverse relation:

$$d_{t,i} = re_{clear,i}(\gamma_{s,t}, \theta_{z,t})^{-1} \tag{25}$$

Finally, since we do not want the change in GHI to be unbounded , we normalize the obtained distances:

$$\mathcal{F}_i(\gamma_{s,t}, \theta_{z,t}) = \frac{d_{i,t}}{\sum_{i=1}^N d_{i,t}} \tag{26}$$

where $d_i(t)$ is the distance of signal i at time step t. The second strategy for improving the GHI estimation accuracy is to detect outliers in different signals at each time-step t. Intuitively, if we possess more than one power signal, we can study the distribution of the various estimation errors and exclude from

the objective function those signals that at time-step t are labeled as outliers, applying a standard outlier detection method. We used Tukey's outlier detection method [43], which is based on the interquartile range, and which can be applied to non-symmetric data distributions. We can now define the outliers detection function $\mathcal{I}$ as:

$$\mathcal{I}(e_{i,t}) = \begin{cases} 0, & \text{if } e_{i,t} \le Q_{0.25} - k_q\text{IQ} \ \vee \ e_{i,t} \ge Q_{0.75} + k_q\text{IQ} \\ 1, & \text{otherwise} \end{cases}$$

(27)

where IQ is the interquartile range: i.e. $Q_{0.75} - Q_{0.25}$ and k_q is a parameter dependent on the assumed distribution of e_t. We used $k_q = 1.5$, which corresponds to considering approximately 1% of the points as outliers, under normal data distribution conditions. When a point is identified as an outlier, it is not used to correct GHI. This is done by setting to zero the corresponding elements of the objective function gradient. Formally:

$$\nabla_{\mathbf{GHI}} h_{v,i_o,t_o} = 0$$

(28)

where i_o, t_o are the signal and timestep marked as outliers.

6. Evaluation on case studies

The proposed method has been tested on two case studies with multiple power signals. Both case studies are located in Switzerland and are particularly challenging for GHI estimation since they present:

- multiple PV orientations (even at single inverter level)

- different nominal powers for each PV installation

- significant shading, due to nearby objects and/or horizon profile

- partial PV curtailment

The first case study is located in Biel-Benken, on the northern Swiss plateau close to the German and French borders. It consists of 4 residential rooftop PV installations with nominal powers ranging from 6.6 to 10.7 kWp. The mean distance between the PV plants and the pyranometer is approximately 150 meters. The PV plants are affected by local shading, due to the presence of chimneys and nearby buildings. One PV plant has two different module orientations (mounted on two folds of the same roof). The effect of the horizon in this region is negligible. The second case study is located in Lugano, a hilly region in the alpine foothills. In this case the power signals are related to 5 different industrial PV plants, with nominal power ranging from 126 to 275 kWp. Several inverters are installed in each plant, making a total number of 50 inverters. The mean distance between the PV plants and the pyranometer is approximately 300 meters. In this case the horizon is non-negligible, due to the presence of significant topographical relief formations.
For each case study, both on-ground measurements and satellite data are used for performance assessment. At each location, an ISO 9060 secondary standard pyranometer (CMP10 and CMP21, Kipp&Zonen, Delft, The Netherlands) is used as

Table 2: Data splits for Biel-Benken Ω identification

split [days]	365	182	121	**91**	73
PV$_{RMSE}$ [-]	6.6e-2	5.7e-2	5.64e-2	**5.6e-2**	6.4e-2
GHI$_{RMSE}$ [W/m2]	41.7	34.8	34.4	**33.5**	35

ground-truth reference. In the first case-study, the pyranometer is mounted on one of the roofs hosting the PV installations, while in the second case the pyranometer is located in the SUPSI Trevano campus. The results were also compared against two different satellite-based irradiation models: MACC-RAD and SICCS. MACC-RAD uses the Heliosat-4 method [44], while SICCS is based on a Cloud Physical Properties model [45]. Both models are based on Meteosat satellite images. The MACC-RAD data are freely accessible, while SICCS data are sold by 3E.

The data for the Biel-Benken case study refer to the period from August 1st, 2015 to August 1st, 2016, with a 1-minute sampling time. Since this case presents a great annual variation in terms of shadow pattern, in order to increase the method accuracy, we repeatedly identified Ω for different time splits of the data. Then, we chose Ω as the one that minimizes the total RMSE on the mean PV estimation error. In particular, table 2 shows the attempted split and the achieved RMSE for the PV estimation error and for the GHI estimation error, reported here only for comparison. The chosen split uses 3 months data folds
.

The data for the Lugano case study refer to the period from January 1st, 2016 to June 1st, 2016, with a 10-minute sampling time. In both case studies, equation 2 is solved for each PV plant separately, obtaining 4 and 5 different GHI estimations, respectively. Equation 4 is then solved using all the signals from the different PV plants, in an attempt to improve the GHI estimation.
Figure 3 summarizes the main results of two case studies, for a sampling time of 10 minutes. The top graphs show the normalized error distributions between GHI$_{py}$ and the two satellite-based models and between GHI$_{py}$ and the solution of problem 4, where GHI$_{py}$ is the signal measured by the pyranometers. The normalization is obtained by dividing the error distribution by a constant value k_n, defined as the mean value of non-zero GHI observations:

$$k_n = \frac{1}{T^*} \sum_{t=1}^{T^*} \text{GHI}_{py,t} \qquad \forall \ \text{GHI}_{py,t} > 0$$

(29)

where T^* is the number of non-zero elements of vector $\mathbf{GHI}_{py}$. In both case studies, estimating GHI from on-ground measurements significantly narrows the error distribution. The lower part of figure 3 shows the estimated probability density function for the absolute relative errors. Each line represents the GHI calculated using one single PV plant, while the thick slashed line is related to the GHI calculated using all the PV signals. The red band represents the typical pyranometer level of ac-
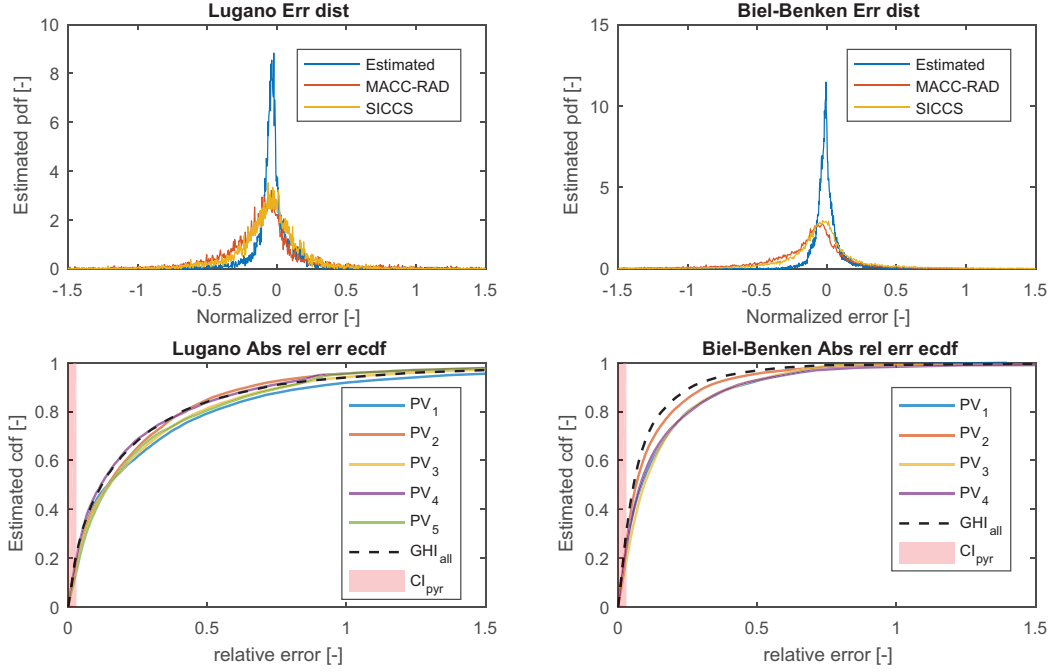
Figure 3: Error distribution for the two case studies with sampling period of 10 minutes. Upper part: empirical probability density functions of the normalized error distributions for the estimated GHI signal and for the GHI signal from MACC-RAD and SICCS models. Lower part: empirical cumulated density functions for the absolute relative errors of the estimated signals. Colored lines refers to the GHI signal estimated using single PV plants power signals, the black dashed line refers to the GHI signal estimated with power signals of all the PV plants. Pale red bands refers to the confidence interval of the pyranometer

curacy. The results suggest that using more than one signal increases the robustness and accuracy of the method. This is partly due to the trust functions and outlier detection function, as better explained in figure 4. As we can see, the normalized root mean squared error decreases when we remove erratic observations from the objective function, and when we use the trust function to weight the signals.

Figure 3 shows that the proposed method has low bias with respect to the satellite-based methods. In order to gain additional insight into the error distributions, we performed a bias-variance error decomposition:

$$RMSE_{\mathcal{D}}^2 = (\mathbb{E}_{\mathcal{D}}e)^2 + \left[\mathbb{E}_{\mathcal{D}}\left(e - \mathbb{E}_{\mathcal{D}}e\right)^2\right]^2 \tag{30}$$

$$= \text{bias}_{\mathcal{D}}^2 + \text{std}_{\mathcal{D}}^2 \tag{31}$$

where e is the estimation error, $\mathcal{D}$ is a given dataset and $\mathbb{E}_{\mathcal{D}}$ is the expectation over the dataset $\mathcal{D}$. We calculated $\text{bias}_{\mathcal{D}}$ and $\text{std}_{\mathcal{D}}$ using daily datasets. We performed the calculation using the maximum available time resolution, for our method with trust function and outlier detection and for the two satellite-based methods. This procedure generates bivariate distributions in terms of bias and standard deviation. For visualization reasons, instead of showing all the points generated by this daily decomposition, estimations of the regions containing 25, 50 and 75 % of the points, respectively, are plotted in figure 5. These

estimations were obtained using the kernel density smoother Matlab function `ksdensity`. Normalized bias and normalized standard deviation for all the observations in the datasets are shown (filled circles). We also show the daily expected values for the normalized bias and normalized standard deviation (diamonds). We can see from figure 5 that, in comparison with other methods, the proposed method has a narrower distribution in terms of bias and standard deviation.

For the Biel-Benken case study, the exact tilt, azimuth and nominal power of the installed PV plants are known. In order to check if these values were estimated correctly, we compared the ground truth with the identified $\mathbf{\Omega}$, plotted as a function of azimuth and elevation in figure 6. It can be seen that the non-zero coefficients of $\mathbf{\Omega}$ are close to those of the real PV plants. In fact, except for the second PV plant, whose PV panels present more than one orientation, the real orientations lie in the convex-hull formed by the non-zero coefficients.

7. Conclusions

In this paper, we present an unsupervised method for estimating global horizontal irradiance from the AC measurements of one or more PV plants, consisting of PV modules of unknown nominal power and orientation. The only inputs to the method are the AC power signals from the PV plants, with corresponding timestamp, and their approximate location in terms
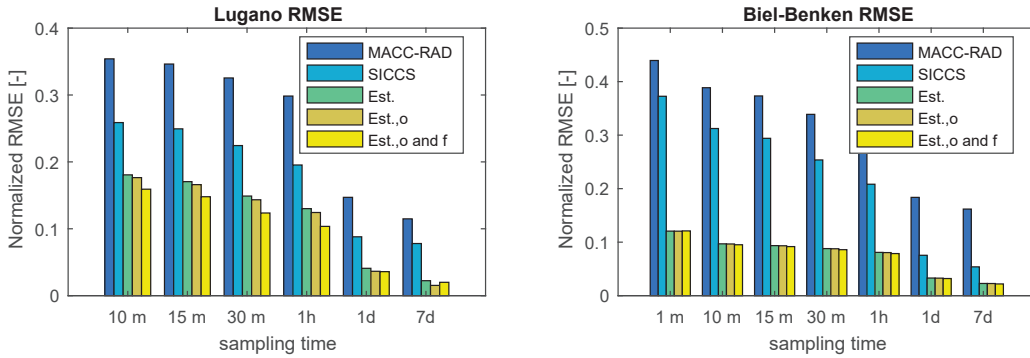
Figure 4: Normalized RMSE for the two test cases, for different aggregation times. The blue and light blue bars represent the RMSE of the MACC-RAD and SICCS models. The green, brown and yellow bars represent the RMSE of the estimated GHI using our method without corrections, with the outliers detection function and with both outliers detection and trust functions, respectively.
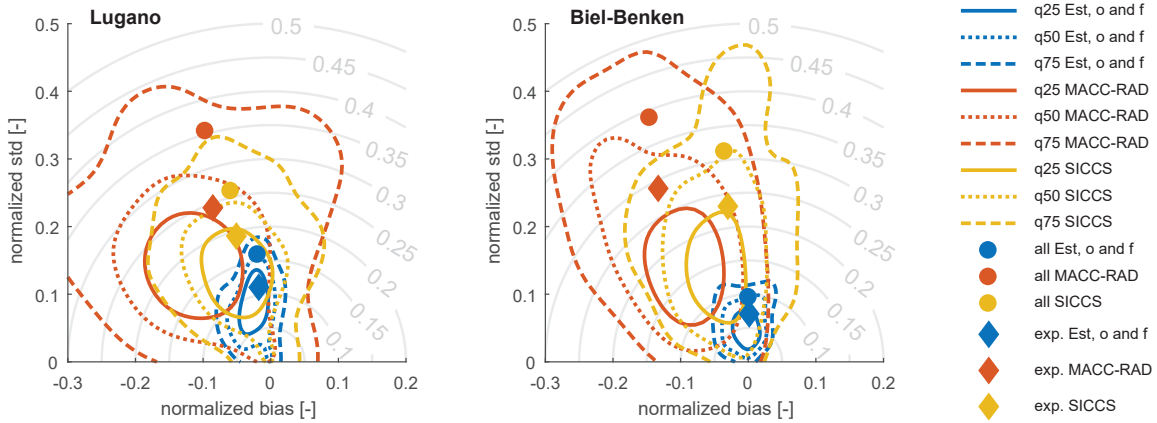
Figure 5: Normalized bias and standard deviation of the MACC-RAD and SICCS services and the proposed method (using outlier detection and trust function), for the two test sites. Colored lines are estimated regions containing 25, 50 and 75 % of the observations, when considering daily subsets. Filled circles represent normalized squared bias and normalized standard deviation for the whole datasets, while diamonds represent the expected daily normalized squared bias and normalized standard deviation. Grey lines show contour lines for the normalized RMSE.

of latitude, longitude and altitude.

An algorithm was developed to speed up the optimization of the underlying non-linear non-convex problem. In terms of computational time, this compares favorably with existing general-purpose solvers.

The method was tested in two different case-studies, both presenting shading and partial curtailment. With respect to other existing satellite-based methods, the results show a significant improvement in the GHI estimation, in terms of RMSE, as shown in figure 4. In both case studies, the relative calculation error is within the secondary standard pyranometer confidence interval for roughly 20-30% of the observations, as shown in 3. The method can correctly identify the orientation and nominal power of the PV modules, even when the PV plant presents PV fields with different orientations. See figure 6.

The method relies on constructing proxies for the electrical power output of the PV modules. This depends on a set of parameters,

namely β, γ, k_1, k_2, k_3, k_4, which in this study were kept fixed. Further work is required in order to determine the influence of these parameters on the performance of the algorithm in terms of estimation accuracy. In this work the Maxwell empirical disc model has been used to disaggregate the direct and diffuse irradiance. In recent years, other models have been developed, which have been shown to provide better results [46], as for example the ENGERER2 model [47]. In future work we will study the effect of different separation models on the algorithm performance.

The proposed method will be used in future studies, in order to disaggregate PV generation from electricity demand, in an attempt to increase the accuracy of aggregated net load short-term forecasts in a low voltage grid. The developed algorithm is freely available as open-source code at

`https://github.com/supsi-dacd-isaac/GHIEstimator`.
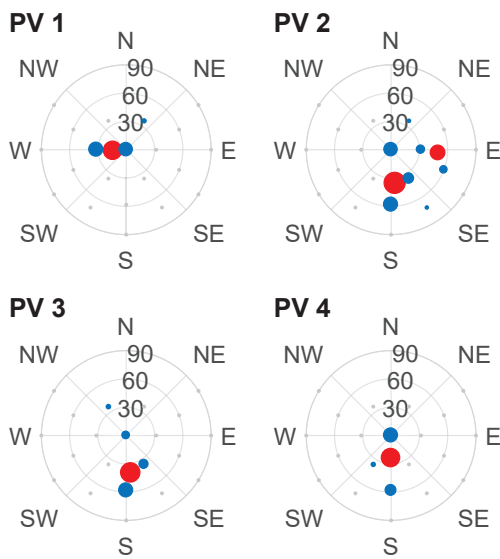
Figure 6: Identified Ω for the Biel-Benken case study. The grey dots indicate the orientations of the proxies, each dot referring to a different column in the **Pr** matrix. The blue dots represent the identified coefficients of Ω while red dots represent the ground truth. The size of the dots is proportional to the identified/nominal power for a given orientation.

Acknowledgments

References

[1] J. N. Mayer, P. Simon, N. S. H. Philipps, T. Schlegl, and C. Senkpiel, "Current and future cost of photovoltaics," *Long-term Scenarios for Market Development, System Prices and LCOE of Utility-Scale PV Systems (Fraunhofer ISE, Study on behalf of Agora Energiewende, Freiburg, 2015)*, 2015.

[2] S. Eftekharnejad, V. Vittal, G. T. Heydt, B. Keel, and J. Loehr, "Impact of increased penetration of photovoltaic generation on power systems," *IEEE transactions on power systems*, vol. 28, no. 2, pp. 893–901, 2013.

[3] IEA, "Global EV Outlook 2016: Beyond one million electric cars," 2016. [Online]. Available:

[4] P. Richardson, D. Flynn, and A. Keane, "Impact assessment of varying penetrations of electric vehicles on low voltage distribution systems," in *Power and Energy Society General Meeting, 2010 IEEE*. IEEE, 2010, pp. 1–6.

[5] M. Bahramipanah, R. Cherkaoui, and M. Paolone, "Decentralized voltage control of clustered active distribution network by means of energy storage systems," *Electric Power Systems Research*, vol. 136, pp. 370–382, 2016.

[6] L. Gelazanskas and K. A. Gamage, "Demand side management in smart grid: A review and proposals for future direction," *Sustainable Cities and Society*, vol. 11, pp. 22–30, 2014.

[7] B. G. Kim, S. Ren, M. Van Der Schaar, and J. W. Lee, "Bidirectional energy trading for residential load scheduling and electric vehicles," *Proceedings - IEEE INFOCOM*, vol. 31, no. 7, pp. 595–599, 2013.

[8] D. F. A. Monforte, M. C. Fordham, M. J. Blanco, and ..., "Improving Short- Term Load Forecasts by Incorporating Solar PV Generation," California Energy Commission, Tech. Rep., 2017.

[9] H. Shaker, H. Chitsaz, H. Zareipour, and D. Wood, "On comparison of two strategies in net demand forecasting using Wavelet Neural Network," *2014 North American Power Symposium, NAPS 2014*, 2014.

[10] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised Disaggregation of PhotoVoltaic Production from Aggregated Power Flow Measurements of Heterogeneous Prosumers,"*IEEE Transactions on Industrial Informatics*,2018.

[11] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 2, no. 1, pp. 2–10, 2009.

[12] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.pecs.2013.06.002

[13] J. R. Mecikalski, K. M. Bedka, W. M. Mackenzie, and S. J. Paech, "Convective Initiation and Lightning Prediction: The potential of the MTG-FC imagery mission," *EUMETSAT 5th MTG MTM*, no. EUMETSAT 5th MTG MTM, October 2007, pp. 1–39, 2007. [Online]. Available: http://doi.wiley.com/10.1029/2010GL042753

[14] C. Vernay, P. Blanc, and S. Pitaval, "Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3," *Renewable Energy*, vol. 57, pp. 339–347, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2013.01.049

[15] T. Mieslinger, F. Ament, K. Chhatbar, and R. Meyer, "A new method for fusion of measured and model-derived solar radiation time-series," *Energy Procedia*, vol. 48, pp. 1617–1626, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.egypro.2014.02.182

[16] A. T. Lorenzo, M. Morzfeld, W. F. Holmgren, and A. D. Cronin, "Optimal interpolation of satellite and ground data for irradiance nowcasting at city scales," *Solar Energy*, vol. 144, pp. 466–474, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2017.01.038

[17] A. Laudani, F. R. Fulginei, A. Salvini, M. Carrasco, and F. Mancilla-David, "A fast and effective procedure for sensing solar irradiance in photovoltaic arrays," *EEEIC 2016 - International Conference on Environment and Electrical Engineering*, no. 3, pp. 0–3, 2016.

[18] E. C. Kara, C. M. Roberts, M. Tabone, L. Alvarez, D. S. Callaway, and E. M. Stewart, "Towards Real-Time Estimation of Solar Generation From Micro-Synchrophasor Measurements," pp. 1–32, 2016. [Online]. Available: http://arxiv.org/abs/1607.02919

[19] P. Traxler, P. Gómez, and T. Grill, "A Robust Alternative to Correlation Networks for Identifying Faulty Systems," in *26th International Workshop on Principles of Diagnosis A*, 2015, pp. 11–18.

[20] N. A. Engerer and F. P. Mills, "ScienceDirect K PV : A clear-sky index for photovoltaics," *Solar Energy*, vol. 105, pp. 679–693, 2014.

[21] S. Killinger, F. Braam, B. Muller, B. Wille-Haussmann, and R. McKenna, "Projection of power generation between differently-oriented PV systems," *Solar Energy*, vol. 136, pp. 153–165, 2016.

[22] B. Marion and B. Smith, "Photovoltaic system derived data for determining the solar resource and for modeling the performance of other photovoltaic systems," *Solar Energy*, vol. 147, pp. 349–357, 2017.

[23] S. Killinger, N. Engerer, and B. Muller, "QCPV: A quality control algorithm for distributed photovoltaic array power output," *Solar Energy*, vol. 143, pp. 120–131, 2017.

[24] D. Yang, "Solar radiation on inclined surfaces: Corrections and benchmarks," *Solar Energy*, vol. 136, pp. 288–302, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2016.06.062

[25] P. G. Loutzenhiser, H. Manz, C. Felsmann, P. A. Strachan, T. Frank, and G. M. Maxwell, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254–267, 2007.

[26] A. Wills, T. B. Schön, L. Ljung, and B. Ninness, "Blind identification of wiener models," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 18, no. PART 1, 2011, pp. 5597–5602.

[27] H. Ohlsson, L. Ratliff, R. Dong, and S. S. Sastry, "Blind identification via lifting," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 19, no. 3. IFAC, 2014, pp. 10 367–10 372. [Online]. Available:

http://dx.doi.org/10.3182/20140824-6-ZA-1003.02567

[28] M. Lef, A. Oumbe, P. Blanc, B. Espinar, Z. Qu, L. Wald, M. S. Hom-scheidt, A. Arola, M. Lef, A. Oumbe, P. Blanc, and B. Espinar, "McClear : a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, pp. 2403—-2418, 2013.

[29] V. P. Lonij, A. E. Brooks, K. Koch, and A. D. Cronin, "Analysis of 80 rooftop PV systems in the Tucson, AZ area," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 549–553, 2012.

[30] M. J. Reno and C. W. Hansen, "Identification of periods of clear sky irradiance in time series of GHI measurements," *Renewable Energy*, vol. 90, pp. 520–531, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2015.12.031

[31] S. Butterworth, "On the theory of filter amplifiers," pp. 536–541, 1930.

[32] P. J. Huber and E. M. Ronchetti, *Robust Statistics 2e*, 2009, vol. 523, no. 3. [Online]. Available: http://doi.wiley.com/10.1002/0470010940

[33] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics - Theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03610927708827533

[34] J. S. Stein, "The photovoltaic Performance Modeling Collaborative (PVPMC)," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 3048–3052, 2012.

[35] E. L. Maxwell, "A quasi-physical model for converting hourly global to direct normal insolation," pp. 35–46, 1987. [Online]. Available: http://rredc.nrel.gov/solar/pubs/PDFs/TR-215-3087.pdf

[36] G. Valentin, "Design and Simulation of Photovoltaic Systems Manual," pp. 1–82, 2012.

[37] ASHRAE, "Methods of Testing to Determine the Thermal Performance of Solar Collectors, ASHRAE Standard 93-77," 1978.

[38] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy*, vol. 83, no. 5, pp. 614–624, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2008.10.008

[39] N. Pham, A. Malinowski, and T. Bartczak, "Comparative study of derivative free optimization algorithms," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 592–600, 2011.

[40] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: A review of algorithms and comparison of software implementations," *Journal of Global Optimization*, vol. 56, no. 3, pp. 1247–1293, 2013.

[41] Y.-x. Yuan, "A Review of Trust Region Algorithms for Optimization," *ICIAM 99: Proceedings of the Fourth International Congress on Industrial & Applied Mathematics, Edinburgh, 2000 Oxford University Press, USA*, vol. 99, pp. 271–282, 2000.

[42] R. H. Inman, Y. Chu, and C. F. M. Coimbra, "ScienceDirect Cloud enhancement of global horizontal irradiance in California and Hawaii," *SOLAR ENERGY*, vol. 130, pp. 128–138, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2016.02.011

[43] J. W. Tukey, "Exploratory Data Analysis," *Analysis*, vol. 2, no. 1999, p. 688, 1977. [Online]. Available: http://www.springerlink.com/index/10.1007/978-1-4419-7976-6

[44] C. Thomas, E. Wey, P. Blanc, L. Wald, and M. Lefèvre, "Validation of HelioClim-3 Version 4, HelioClim-3 Version 5 and MACC-RAD Using 14 BSRN Stations," *Energy Procedia*, vol. 91, pp. 1059–1069, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.egypro.2016.06.275

[45] W. Greuell, J. F. Meirink, and P. Wang, "Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations," *Journal of Geophysical Research Atmospheres*, vol. 118, no. 5, pp. 2340–2361, 2013.

[46] C. A. Gueymard and J. A. Ruiz-Arias, "Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance," *Solar Energy*, vol. 128, pp. 1–30, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2015.10.010

[47] N. A. Engerer, "Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia," *Solar Energy*, vol. 116, pp. 215–237, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2015.04.012

C Appendix C

Paper C: *L. Nespoli and V. Medici, "Constrained hierarchical networked optimization for energy markets," IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2018.*
List of authors:

- Nespoli, Lorenzo: Swiss Federal Institute of Technology of Lausanne (EPFL), University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

- Medici, Vasco: University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

Personal contributions:

- State of the art, theory, simulations.

Constrained hierarchical networked optimization for energy markets

Lorenzo Nespoli
Swiss Federal Institute of Technology
Lausanne, Switzerland
University of Applied Sciences and
Arts of Southern Switzerland
Lugano, Switzerland
Email: lorenzo.nespoli@epfl.ch

Vasco Medici
University of Applied Sciences and
Arts of Southern Switzerland
Lugano, Switzerland
Email: vasco.medici@supsi.ch

Abstract—In this paper, we propose a distributed control strategy for the design of an energy market. The method relies on a hierarchical structure of aggregators for the coordination of prosumers (agents which can produce and consume energy). The hierarchy reflects the voltage level separations of the electrical grid and allows aggregating prosumers in pools, while taking into account the grid operational constraints. To reach optimal coordination, the prosumers communicate their forecasted power profile to the upper level of the hierarchy. Each time the information crosses upwards a level of the hierarchy, it is first aggregated, both to strongly reduce the data flow and to preserve the privacy. In the first part of the paper, the decomposition algorithm, which is based on the alternating direction method of multipliers (ADMM), is presented. In the second part, we explore how the proposed algorithm scales with increasing number of prosumers and hierarchical levels, through extensive simulations based on randomly generated scenarios.

I. INTRODUCTION

With the penetration of renewable energy sources (RES), a decentralized market design with self-dispatch components is developing in the distribution grid. The demand side is becoming increasingly capable of providing flexibility services and contributing to a reliable power system and price stability on power markets. As flexible generation and consumption capacity will be highly fragmented and distributed, to better exploit it and maximize its economical profitability a high number of prosumers will be required to coordinate with each other, when responding to demand response (DR) signals. A market design that rewards flexibility needs to be set up. The highly stochastic nature of RES generation calls for a market that is able to operate in near real-time.

The presence of highly correlated distributed generation increases the risk of local congestions and voltage fluctuations. Indeed, the highly stochastic nature of RES generation calls for near real-time control of flexibility. Many authors have tried to achieve prosumers coordination in different ways, for instance by modelling the market as Cournot games with constraints [1], [2], [3], seeking Nash equilibria in non-cooperative games [4], [5], generalized Nash equilibria [6], and as a distributed control problem. An optimal coordination of the prosumers can be achieved in different ways, among which making use of aggregators is one of the most promising ones [7]. The question of how the aggregators will achieve coordination, however, is still matter for research with a number of promising solutions being investigated. An interesting way to exploit flexibility of a pool of prosumers is to explicitly formulate a common target for their aggregated power profile, and give them economic incentives to follow this target. Energy retailers and balance responsible parties, which bid for purchase of energy in the energy market, would benefit from a reduction of uncertainty in the prosumers consumption or production. In this paper, we consider prosumers as cooperative agents not able to modify the control algorithm that optimizes the operations of their flexible loads. As such, we are not obliged to choose prosumers' utility functions that generate a unique generalized Nash equilibrium. We will rather focus on a distributed control protocol allowing prosumers coordination through multiple voltage levels. In this context, a good coordination protocol must ensure prosumers privacy while being scalable. Prosumers privacy is inherently guaranteed if they do not need to share their private information (e.g. size of batteries, desired set-point temperatures in their homes), or their forecasted power profile. Scalability ensures that the computational time of coordination scales near-linearly with increasing number of agents, allowing for fast control.

Most studies on the subject are focused on maximizing the welfare of a group of prosumers, by means of maximizing their utility functions. In the mathematical optimization framework, this problem can be modelled as an allocation or exchange problem [8]. In [9] the welfare maximization problem is considered with additional coupling constraints, modelling line congestions. The problem is solved using a primal-dual interior point method, considering that each agent has access to the dual updates of his neighbours. In [10] the same problem is solved with different multi-steps gradient methods. In recent years, other authors proposed decomposition techniques based on the theory of monotone operator splitting [11]. These algorithms are known to have more convenient features in terms of convergence with respect of the gradient-based counterparts. An example of such approach, is represented by the proximal algorithms, which are well

suited for non-smooth, constrained, distributed optimization [12]. In [13] the decomposition of the welfare maximization problem under uncertainty is considered, combining proximal gradient method and weighted gradient method. [14] proposes a decentralized energy market that makes use of the alternate direction method of multipliers (ADMM) [8] to split the problem. In [15] the unit commitment problem is solved through ADMM. In [16] a robust implementation of demand response mechanism is introduced and solved with ADMM. A multi-objective optimization problem aiming at maximizing prosumer's welfare while minimizing a system-level objective can be modelled as a sharing problem, see for example [8], [17]. The general sharing problem can be written as:

$$\underset{x \in \mathcal{X}}{\operatorname{argmin}} \, e(x) + \sum_{i=1}^{N} f_i(x_i) \qquad (1)$$

where $x_i \in \mathbb{R}^t$ are the prosumers' vector of decision variables, $x = [x_i^T]^T = [x_1^T, ... x_N^T]^T \in \mathbb{R}^{NT}$ is the concatenation of all the decision variables, $\mathcal{X}$ is a convex and compact set of constraints, $e : \mathbb{R}^{NT} \to \mathbb{R}$ is a system-level objective function and $f_i : \mathbb{R}^T \to \mathbb{R}$ is a prosumer specific objective function. For example, in [18] an application to the energy market is considered, where a dynamic version of the sharing problem with a tracking profile system-level objective is decomposed using the Douglas-Rachford splitting. The same problem is solved in [19], where an adaptation of the ADMM algorithm for the sharing problem is used. This is the same solution approach proposed in [8], §7.3. In [20] a high-level hierarchical control flow between the DSO, independent aggregators and prosumers is proposed, but no link is given between the different control signals. It is worth noting that some authors, as [16], [19], [15], use the term hierarchical to refer to a single level hierarchy, in which a the problem can be solve with a master-slave solution scheme.

In this paper, we introduce a hierarchical market design that exploits the flexibility of prosumers located in different voltage levels of the distribution grid. By aggregating a higher number of prosumers, we can better exploit their flexibility for grid regulation. Each prosumer communicates with an aggregator, i.e. with his parent node. The algorithm, which can be monolithically described as a sharing problem, effectively preserves privacy between the different levels, since only aggregated information is available at the higher levels of the communication structure. In the first part of the paper, we present the algorithm used to solve the coordination problem. In the second part, we present results from the coordination of prosumers in different hierarchical structures. We systematically vary the number of levels and draw the number of prosumers per level from a uniform distribution. Results on convergence and computational time are presented.

II. Problem Formulation

In this work we jointly maximize prosumers' specific objective functions and a system-level objective, taking into account grid constraints. Prosumer's flexibility is modeled by means of electrical batteries, but the approach can be generalized to other kinds of flexibilities. Prosumers communicate only indirectly, with the help of aggregators, located in the branching nodes of the hierarchical structure.

This problem can be formulated monolithically using the very general formulation in 1. However, 1 does not explicitly show the tree-like dependences of the problem we would like to solve. In order to express the hierarchical nature of the problem, we briefly introduce the nomenclature of rooted tree structure, from graph theory. A rooted tree τ is a unidirected acyclic graph, with every node having exactly one parent, except for the root node. Each node is identified by a tuple $(d_1..d_l...d_d)$ where l_d is the level to which the node belongs, and each entry represent the enumeration of its l_{th} level ancestor. In this paper, we keep $d_1 = 1$, and indicate the root node as $\emptyset$. Next, we introduce the definition of the set we will use in the description of the algorithm.

Definition II.1 (Node sets). 1) Descendants of node $A = (d_1, ..d_L)$. $\mathcal{D}(A) = \{A_j = (j_1, ..j_{L_j}) : L_j > L, (j_1, .., j_L) = A\}$
2) Leaf node $\mathcal{L}(\tau) = \{A \in \tau : \mathcal{D}(A) = \emptyset\}$
3) Nodes in level l. $\mathcal{N}_l(\tau) = \{A \in \tau : L = l\}$
4) Branching nodes. $\mathcal{B}(\tau) = \backslash \mathcal{L}(\tau)$
5) Anchestors of node A. $\mathcal{A}(A) = \{A_j \in \tau : A \in \mathcal{D}(A_j)\}$
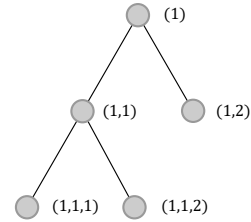
Fig. 1. Example of rooted tree hierarchical structure

Figure 1 shows a simple example of rooted tree hierarchical structure and the tuples associated with every node. Now, we can use the definition of branching node of level l to rewrite problem 1. For sake of simplicity and clarity of exposition, we assume to have only one constraint in each branching node. We will remove this assumption in the simulations.

$$\underset{x}{\operatorname{argmin}} \, e(S_\emptyset x) + \sum_{i=1}^{N} f_{c,i}(x_i) \qquad (2)$$
$$s.t. : S_B x \leq v_B \quad \forall \, B \in \mathcal{B}(\tau)$$

where x_i are the actions associated to the agent i in the terminal node A_i, B denotes the branching nodes of the tree τ and $S \in \mathbb{R}^{T \times N}$ are summation matrices defined as:

$$S_B = [M_{B,A_j}], \quad M_{B,A_j} = \begin{cases} a_{B,j} \mathbb{I}_T, & \text{if} \quad A_j \in \mathcal{D}(B) \\ \mathbf{0}_T, & \text{otherwise} \end{cases}$$

where $\mathbf{0}_T$ and $\mathbb{I}_T$ are the zero and identity matrix of size T respectively, $a_{B,j}$ is a weight associated to the j_{th} descendant

145

of branching node B and $f_{c,i}$ is the objective function of agent i, which takes into account his agent-specific constraints:

$$f_{c,i}(x_i) = \begin{cases} f(x_i), & \text{if} \quad x_i \in \mathcal{X}_i \\ \infty, & \text{otherwise} \end{cases}$$

We can now clarify how the notation used in 2 allows us to express grid constraints in a flexible way. Again, for sake of simplicity, in the following we only consider apparent power and no additional uncontrolled loads. Both these assumptions will be removed in the presented simulations. Since we are considering a radial grid, we can express the total power at a given branching node B as the sum of the power of its descendants, $\mathcal{D}(B)$. In this case, we can set $a_{B,j} = 1 \quad \forall j$ so that S_B becomes the time-summation matrix of powers of $\mathcal{D}(B)$. Imposing voltage constraints in terms of apparent power involves solving the power flow (PF) equations. Solving the exact PF equations would result in a non-convex optimization problem, which are in general difficult to solve. Different convex formulations of the PF exist, the most adopted being the DC PF model. Despite well suited for high and medium voltage grids, this model is typically inappropriate for distribution systems [21]. Furthermore, the DC PF still requires susceptances, voltage angles and the knowledge of the grid topology. Differently from the medium voltage grid, parameters and topology are hardly available for low voltage grids. A better linear approximation for low-voltage grids is represented by the first order truncation of the PF equations [22]:

$$|V| \approx V_0 + P^T \nabla_P |V| + Q^T \nabla_Q |V| \tag{3}$$

where $P \in \mathbb{R}^n$, $Q \in \mathbb{R}^n$ are the nodal active and reactive power in a grid of n nodes and V and V_0 are the voltage and reference voltage at a given point of the grid. $\nabla_P |V| \in \mathbb{R}^n$ and $\nabla_Q |V| \in \mathbb{R}^n$ are the gradients with respect to the nodal active and reactive power at each node, and are collectively called voltage sensitivity coefficients. It has been shown that they can be estimated using distributed sensor networks of phasor measurement units [23] or even smart meter data [24]. We can use this approximation, replacing $a_{B,j}$ with the voltage sensitivity coefficient of node j with respect to B. If we set $v_B = V_{max,B} - V_{0,B}$ we retrieve the formulation in 3.

III. PROBLEM DECOMPOSITION

A trivial way of decomposing the problem would consist in repeatedly applying existent decomposed formulation of the sharing problem for each level. This would result in an exponentially increasing computational time, with the number of considered levels, namely:

$$N_{tot} \sim tnN_i^L \prod_{l=1}^{L} N_b^l \sim N_i^L N_b^{\frac{L+L^2}{2}} \tag{4}$$

where t is the computational time for each agent for solving its local problem, n is the number of agents per level, N_i is the number of iterations before convergence for a single level, L is the number of levels, and N_b is the number of branches per level. Instead of following this strategy, we decompose the

monolithic formulation of the problem to obtain a near-linear increase of iterations with the number of levels. Problem 2 is not decomposable as it is, due to the first term $e(x)$ and the coupling constraints. To decompose it, we introduce additional variables, copy of the linear transformations of $\mathcal{DB}(\tau)$:

$$\underset{x}{\arg\min} \, e(y_\emptyset) + \sum_{i=1}^{N} f_{c,i}(x_i)$$
$$s.t. : y_{B,i} = a_{B,i} x_i \quad \forall \, B \in \mathcal{B}(\tau), \quad \forall i \in \mathcal{D}_\mathcal{L}(B) \tag{5}$$
$$\sum_{i \in \mathcal{D}_\mathcal{L}(B)} y_{B,i} \leq v_B \quad \forall \, B \in \mathcal{B}(\tau)$$

where $\mathcal{D}_\mathcal{L}(B)$ is the set containing the terminal nodes which descent from branch B, $\mathcal{D}_\mathcal{L}(B) = \mathcal{D}(B) \cap \mathcal{L}(\tau)$ We now formulate 5 as an unconstrained minimization problem. We do it using an augmented Lagrangian formulation for the equality constraints involving duplicated variables and explicitly consider inequality constraints of 5 by means of the indicator functions. See for example [12] §5.4 for a similar approach applied to the allocation problem.

$$\mathcal{L}_\rho = e(y_\emptyset) + \sum_{i=1}^{N} f_{c,i}(x_i) + I_{\mathcal{Y}_B}(\overline{y}_B)$$
$$+ \sum_{B \in \mathcal{B}(\tau)} \sum_{i \in \mathcal{D}_\mathcal{L}(B)} \frac{1}{2\rho} \| a_{B,i} x_i - y_{B,i} + \lambda_{B,i} \|_2^2 \tag{6}$$

where $\overline{y}_B$ is the sum of $y_{B,i}$, defined as $\overline{y}_B = \sum_{i \in \mathcal{D}_\mathcal{L}(B)} y_{B,i}$, $\lambda_{B,i}$ are the dual variables associated to the equality constraints of problem 5, $\rho \in \mathbb{R}$ is the augmented Lagrangian parameter, $\mathcal{Y}_B$ are the constraint sets of the inequalities of problem 5 and $I_{\mathcal{Y}_B}(y_B)$ are indicator functions, defined as:

$$I_\mathcal{Y}(y) = \begin{cases} 0, & \text{if} \quad y \in \mathcal{Y} \\ \infty, & \text{otherwise} \end{cases}$$

and $y_B = [y_{B,i}]$. We now follow the alternating direction method of multipliers (ADMM) strategy [8], and perform a joint minimization-maximization of $\mathcal{L}_\rho(x, y, \lambda)$. Note that the convergence results from the ADMM algorithm allow us to take into account extended-real-valued and non-differentiable functions, as the indicator function. The overall decomposed problem can be written as:

$$x_i^{k+1} = \operatorname*{argmin}_{x_i} f_{c,i}(x_i) +$$
$$\sum_{A \in \mathcal{A}(A_i)} \sum_{j \in \mathcal{D}_\mathcal{L}(A)} \frac{1}{2\rho} \|a_{A,j} x_j^k - y_{A,j}^k + \lambda_{A,j}^k\|_2^2 \quad (7)$$

$$y_\emptyset^{k+1} = \operatorname*{argmin}_{y_\emptyset} e(y_\emptyset) +$$
$$\sum_{i \in \mathcal{D}_\mathcal{L}(\emptyset)} \frac{1}{2\rho} \|a_{\emptyset,i} x_i^{k+1} - y_{\emptyset,i} + \lambda_{\emptyset,i}^k\|_2^2 + I_{\mathcal{Y}_\emptyset}(\overline{y_\emptyset}) \quad (8)$$

$$y_{B,i}^{k+1} = \operatorname*{argmin}_{y} \frac{1}{2\rho} \|a_{B,i} x_i^{k+1} - y + \lambda_{B,i}^k\|_2^2 + I_{\mathcal{Y}_B}(\overline{y}_B) \quad (9)$$

$$\lambda_{B,i}^{k+1} = \lambda_{B,i}^k + \rho(a_{B,i} x^{k+1} - y_{B,i}^{k+1}) \quad (10)$$

Note that due to the definition of the S_B matrices, the x_i update only involves constraints from the ancestors of node A_i. This is thanks to the fact that in our model, actions of node A_i do not influence agents in other subtrees, and is the ultimate reason that justifies a hierarchical communication structure. Note that for solving the first minimization problem, agent i must consider all the other agent actions x_{-i} as fixed. It has been shown that following a Gauss-Seidel like iteration, in which agents update their actions in sequence, considering all the available updated actions from the other agents, the above formulation converges [17]. Anyway, this requires to solve the subproblems in sequence, reducing the computational advantage of a distributed solution. We prefer to use a parallel formulation, in which agents solve their own problems simultaneously. This will obviously not decrease the overall computations, but rather the effective convergence time. We parallelize the problem fixing the average of the auxiliary variables $y_{B,i}$ during their update step. This will effectively reduce the overall number of variables and allows for a stable parallelization. Note that the resulting formulation can be interpreted as requiring each x_i to reduce the average constraint violations $S_B x^k - \overline{y}_B$ minus the scaled value of the previous iteration $a_{B,i} x^k$. See [19] and [8] §7.3 for a detailed description of the method. We reformulate the iterations noting that the $\overline{y}_B$ updates can be rewritten in terms of the proximal operator $\mathbf{prox}_{\rho f}$. Additionally, for all the $\overline{y}_B$ but the root node, the proximal operator of the indicator function reduces to the projection operator $\mathbf{\Pi}_\mathcal{X}$.

$$x_i^{k+1} = \operatorname*{argmin}_{x_i} f_{c,i}(x_i) + \sum_{B \in \mathcal{A}(A)} \frac{1}{2\rho} \|(S_B x^k - \overline{y}_B^k)/N_B$$
$$- a_{B,i} x_i^k + x_i + \overline{\lambda}_B^k\|_2^2 \quad (11)$$

$$\overline{y}_\emptyset^{k+1} = \mathbf{\Pi}_{\mathcal{Y}_\emptyset}(\mathbf{prox}_{\rho e}(S_\emptyset x^{k+1} + \overline{\lambda}_\emptyset^k)) \quad (12)$$

$$\overline{y}_B^{k+1} = \mathbf{\Pi}_{\mathcal{Y}_B}(S_B x^{k+1} + \overline{\lambda}_B^k) \quad (13)$$

$$\overline{\lambda}_B^{k+1} = \overline{\lambda}_B^k + \frac{\rho}{N_B}(S_B x^{k+1} - \overline{y}_B^{k+1}) \quad (14)$$

where N_B is the number of descendants of branch B. Since the root node update involves the minimization of system-level objective function e, equation 12 projects its proximal

minimization into the root node constraint set $\mathcal{Y}_\emptyset$, similarly to proximal gradient methods, as the forward-backward splitting [25].

The pseudocode of the update rule is summarized in Algorithm 1. The sum of norms in the agent update step can be reduced to a single norm:

$$x_i^{k+1} = \operatorname*{argmin}_{x_i} f_{c,i}(x_i) + \frac{1}{2\rho} \|r_i + R_i x_i - R_{a,i} x_i^k\|_2^2 \quad (15)$$

where $R_i = [\mathbb{I}_T] \in \mathbb{R}^{Tn_a \times T}$ is the concatenation of n_a identity matrices where n_a is the number of ancestors of agent i, $R_{a,i} = [a_{B,i} \mathbb{I}_T] \in \mathbb{R}^{Tn_a \times T}$, and $r_i = [r_B] \in \mathbb{R}^{Tn_a}$ is the concatenation of reference signals form its ancestors:

$$r_B = (S_B x - \overline{y}_B)/N_B + \lambda_B \quad (16)$$

Algorithm 1 Hierarchical optimization
1:	Initialize err $= tol * 2, \overline{y}_B = 0_T, \lambda_B = 0_T$
2:	**while** err $\leq tol$ **do**
3:	$\quad x_i^{k+1} \leftarrow x_i^k, r_i^k$ $\qquad\qquad\qquad\qquad$ ▷ agents
4:	$\quad \overline{y}_A^{k+1}, r_B^{k+1} \leftarrow \overline{\lambda}_B, x_{\mathcal{D}(B)}^{k+1}$ $\qquad\quad$ ▷ branch B
5:	$\quad \text{err}_B^{k+1} \leftarrow x_{\mathcal{D}(B)}^{k+1}, \overline{y}_B^{k+1}$ $\qquad$ ▷ primal err in branch B
6:	$\quad \overline{\lambda}_B^{k+1} \leftarrow \overline{\lambda}_B^k, \text{err}^{k+1}$ $\qquad$ ▷ dual variable in branch B
7:	$\quad err \leftarrow err_B$
8:	**end while**

We can see from the pseudocode in 1 that each agent requires only the reference signals from its ancestors to solve its optimization problem. Thanks to the hierarchical communication structure, these signal can be collected from the parent node of agent i. This allows the algorithm to be solved in a forward-backward passage. In the forward passage each branch B sends its reference signal r_B and the one received by its parent to his children, which propagate it downwards through the hierarchy. At the same time, prosumers in leaf nodes solve their optimization problem as soon as they receive their overall reference signal r_i. In the backward passage agents send their solutions to their parents, which collect them and send the aggregated solution upward. Note that r_B contains only aggregated information from branch B, which ensures privacy among prosumers.

IV. SIMULATION RESULTS

In this section we present the results of the numerical investigation of the proposed algorithm. In particular, we simulated 500 scenarios of different hierarchical structures in order to study the algorithm performances in terms of computational time. In each scenario the prosumers coordinate their actions for the day ahead. Each agent has a random generated power profile and an electrical battery with a random starting state of charge. The battery are considered to be dynamic linear systems, cyclic and calendar aging are not considered. For each scenario we built a random tree with at most 4 aggregator levels, which means that $l \in [2, 5]$. We only consider trees in which each branching node is the parent of at
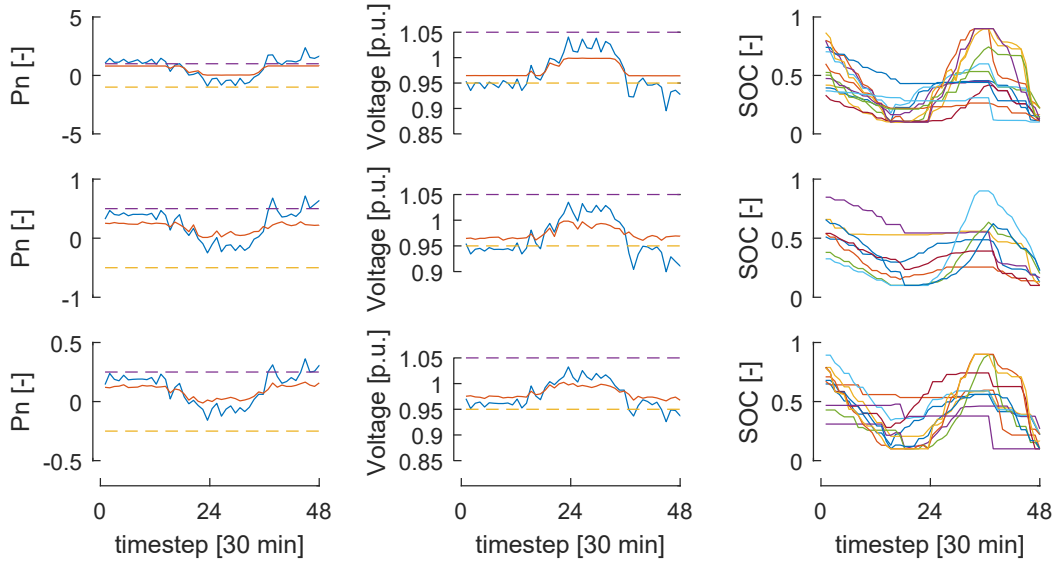
Fig. 2. Example case. The considered hierarchical structure has 4 levels, and a single branching node in the first 3 levels. For the first two columns, the first row refers to the root node, second and third rows to the second and third level. First column represents the aggregated power profiles. Blue line: no battery actions. Red line: optimized power profiles. Dashed lines: power constraints. The second column represent voltage profiles. Blue line: no battery actions. Red line optimized voltage profiles. voltages and the state of charge (SOC) of the batteries are shown. The third column represents state of charge of prosumers'batteries in the second, third and fourth levels.

most other 2 branching nodes, while the maximum number of leaf nodes per branch is 10. Only leaf nodes are considered to be flexible nodes, which means that all prosumers are located in leaf nodes, while branching nodes are aggregators. Voltage sensitivity coefficients for each level are randomly generated. With these rules, we obtain a tree with maximum number of
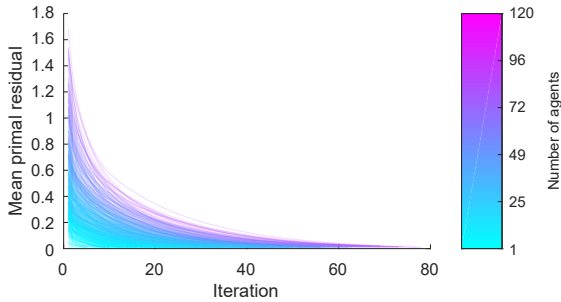
Fig. 3. Mean overall primal residual for all the simulations, plotted versus the number of iterations before convergence, colored against the number of agents.

15 branching nodes (including the root node). The objective function of each prosumers f_i is the sum of its electricity costs or revenues:

$$f_i(x_i) = \sum_{t=1}^{T} C_{i,t} \qquad (17)$$

where the cost at time t is defined as:

$$C_{i,t} = \begin{cases} (P_{u_i,t} + x_{i,t})p_b, & \text{if} \quad P_{u_i,t} + x_i \geq 0 \\ -(P_{u_i,t} + x_{i,t})p_s, & \text{otherwise} \end{cases}$$

where x_i is the overall power from the battery, P_{u_i} is the uncontrolled power of agent i, p_b and p_s are the buying and selling energy prices, respectively. Note that positive powers are considered as consumed quantities. The system-level objective is a tracking objective with a zero power profile, which results in a quadratic peak shaving:

$$e(x) = \|S_\emptyset(x + P_u)\|_2^2 \qquad (18)$$

where $P_u = [P_{u_i}]$. The simulations are carried out using an Intel Core i7-4790K CPU @ 4.00GHz with 32 GB of RAM.

In figure 2 an example of the coordination mechanism is shown. The considered hierarchical structure has 4 levels, and a single branching node in the first three levels. Aggregated power profiles, voltages and the state of charge (SOC) of the batteries are shown.

In figure 3 the mean overall primal residual for all the simulations is shown, where the primal residual in branch B is $err_B = S_B x - y_B$, is plotted versus the number of iterations before convergence, which is considered reached when $err \leq 1e-2$. The line color is related to the total number of prosumers in the related tree. As expected the number of iterations before convergence increases with the number of coordinated prosumers.
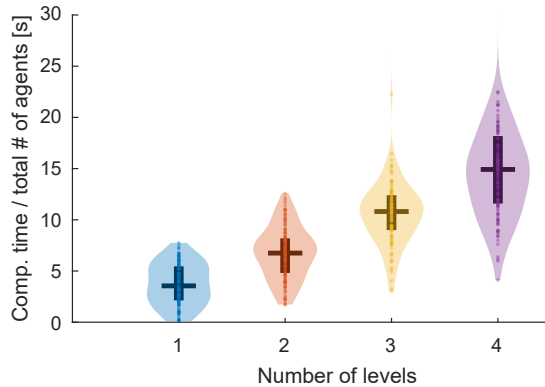
Fig. 4. Estimated probability densities of the computational time divided by the total number of agents, as a function of the number of branching levels. The vertical bar is the interquartile range, the horizontal line is the median.

Figure 4 shows the estimated probability densities of the agent-normalized computational time, as a function of the number of branching levels of the considered tree.

V. Conclusions

We presented a constrained networked optimization algorithm for the coordination of prosumers, which exploits a hierarchical structure, reflecting the hierarchy of the different voltage levels of the electrical grid. Prosumers are coordinated with the help of aggregators, located at the branching nodes. The monolithic optimization problem is decomposed and parallelized using the ADMM, resulting in a forward-backward communication flow in the hierarchy. The proposed mechanism ensures that prosumers' privacy is preserved, since communication between different levels involves only aggregated information. The numerical simulations show that the computational time normalized with the number of prosumers scales linearly with the number of levels. In future work the authors will investigate the algorithm performance using low and medium voltage test grids, by means of power flow simulations.

Acknowledgment

References

[1] J. Yao, S. S. Oren, and I. Adler, "Cournot equilibria in two-settlement electricity markets with system contingencies," *International Journal of Critical Infrastructures*, vol. 3, no. 1-2, pp. 142–160, 2007.

[2] B. Yang and M. Johansson, "Distributed optimization and games: A tutorial overview," in *Networked Control Systems*, 2010, vol. 406, pp. 109–148.

[3] A. R. Kian and J. B. Cruz, "Bidding strategies in dynamic electricity markets," vol. 40, pp. 543–551, 2005.

[4] B.-g. Kim, S. Member, S. Ren, M. V. D. Schaar, J.-w. Lee, and S. Member, "Bidirectional Energy Trading and Residential Load Scheduling with Electric Vehicles in the Smart Grid," *IEEE Journal on selected areas in communication*, vol. 31, no. 7, pp. 1219–1234, 2013.

[5] S. Park, S. Member, J. Lee, S. Bae, S. Member, and G. Hwang, "Contribution-Based Energy-Trading Mechanism in Microgrids for Future Smart Grid : A Game Theoretic Approach," vol. 63, no. 7, pp. 4255–4265, 2016.

[6] C. Li, X. Yu, W. Yu, and T. Huang, "Networked Optimization for Demand Side Management based on Non-cooperative Game."

[7] Y. Parag and B. K. Sovacool, "Electricity market design for the prosumer era," *Nature Energy*, vol. 1, no. 4, p. 16032, 2016.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[9] B. HomChaudhuri and M. Kumar, "A newton based distributed optimization method with local interactions for large-scale networked optimization problems," pp. 4336–4341, 2014.

[10] E. Ghadimi, I. Shames, and M. Johansson, "Multi-step gradient methods for networked optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5417–5429, 2013.

[11] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2011.

[12] N. Parikh and S. Boyd, "Proximal Algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[13] K. Margellos, A. Falsone, S. Garatti, and M. Prandini, "Distributed Constrained Optimization and Consensus in Uncertain Networks via Proximal Minimization," *IEEE Transactions on Automatic Control*, vol. 9286, no. c, pp. 1–16, 2017.

[14] F. Moret, P. Pinson, and S. Member, "Energy Collectives : a Collaborative Approach to Future Consumer-Centric Electricity Markets."

[15] J. Jian, C. Zhang, L. Yang, K. Meng, Y. Xu, and Z. Dong, "A Hierarchical Alternating Direction Method of Multipliers for Fully Distributed Unit Commitment," 2006.

[16] M. Diekerhof, F. Peterssen, and A. Monti, "Hierarchical Distributed Robust Optimization for Demand Response Services," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, pp. 1–1, 2017.

[17] M. Hong, Z. Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, pp. 337–364, 2016.

[18] R. Halvgaard, L. Vandenberghe, N. K. Poulsen, H. Madsen, and J. B. Jørgensen, "Distributed Model Predictive Control for Smart Energy Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 1–9, 2014.

[19] P. Braun, T. Faulwasser, L. Gr, C. M. Kellett, S. R. Weller, and K. Worthmann, "Hierarchical Distributed ADMM for Predictive Control with Applications in Power Networks," pp. 1–10.

[20] B. P. Bhattarai, B. Bak-Jensen, P. Mahat, J. R. Pillai, and M. Maier, "Hierarchical control architecture for demand response in smart grids," *Power and Energy Engineering Conference (APPEEC), 2013 IEEE PES Asia-Pacific*, pp. 1–6, 2013.

[21] D. K. Molzahn, F. Dorfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A Survey of Distributed Optimization and Control Algorithms for Electric Power Systems," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, pp. 1–1, 2017.

[22] H. Almasalma, J. Engels, and G. Deconinck, "Dual-decomposition-based peer-to-peer voltage control for distribution networks," *CIRED - Open Access Proceedings Journal*, no. 1, pp. 1718–1721, 2017.

[23] C. Mugnier, K. Christakou, J. Jaton, M. De Vivo, M. Carpita, and M. Paolone, "Model-less/measurement-based computation of voltage sensitivities in unbalanced electrical distribution networks," *19th Power Systems Computation Conference, PSCC 2016*, 2016.

[24] S. Weckx, S. Member, R. D. Hulst, J. Driesen, and S. Member, "Voltage Sensitivity Analysis of a Laboratory Distribution Grid With Incomplete Data," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1271–1280, 2015.

[25] L. Stella, A. Themelis, and P. Patrinos, "Forward backward quasi-Newton methods for nonsmooth optimization problems," *Computational Optimization and Applications*, vol. 67, no. 3, pp. 443–487, 2017.

D Appendix D

Paper D: *L. Nespoli, M. Salani, and V. Medici, "A rational decentralized generalized Nash equilibrium seeking for energy markets," in 2018 International Conference on Smart Energy Systems and Technologies, SEST 2018 - Proceedings, 2018.* List of authors:

- Nespoli, Lorenzo: Swiss Federal Institute of Technology of Lausanne (EPFL), University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

- Salani, Matteo: Dalle Molle Institute for Artificial Intelligence Research (IDSIA)

- Medici, Vasco: University of Applied Sciences and Arts of Italian Switzerland (SUPSI)

Personal contributions:

- State of the art, theory and theorems, simulations.

A rational decentralized generalized Nash equilibrium seeking for energy markets

Lorenzo Nespoli
EPFL
Lausanne, Switzerland
ISAAC, SUPSI
Campus Trevano, Canobbio, Switzerland
Email: lorenzo.nespoli@epfl.ch

Matteo Salani
IDSIA, USI/SUPSI
Galleria 2, Manno, Switzerland
Email: matteo.salani@idsia.ch

Vasco Medici
ISAAC, SUPSI
Campus Trevano, Canobbio, Switzerland
Email: vasco.medici@supsi.ch

Abstract—We propose a method to design a decentralized energy market which guarantees individual rationality (IR) in expectation, in the presence of system-level grid constraints. We formulate the market as a welfare maximization problem subject to IR constraints, and we make use of Lagrangian duality to model the problem as a n-person non-cooperative game with a unique generalized Nash equilibrium (GNE). We provide a distributed algorithm which converges to the GNE. The convergence and properties of the algorithm are investigated by means of numerical simulations.

I. INTRODUCTION

A. Motivations

The scientific community agrees that in the future the intelligent activation of demand response (DR) will contribute to a reliable power system and price stability on power markets. Actuation of DR requires to solve an optimization problem in order to maximize an economic objective, which typically results in a welfare maximization problem (WMP), in which the unweighted sum of the economic costs of a group of agents is minimized. A very similar, and perhaps more studied problem, is the optimal power flow (OPF) problem. The OPF is usually solved in a centralized way by an independent system operator (ISO), in order to minimize the generation cost of a group of distributed power plants, over the set of underlying grid constraints. When the number of generators increases, the problem could become computationally expensive. Furthermore, retrieving all the generator-specific parameters could become impractical for the ISOs. For these reasons, different decentralized formulations of the OPF exist [1], which can speed up the computation exploiting parallelization among the different units. Furthermore, solving the problem in a decentralized way allows the generators to keep most of their information and parameters private, increasing privacy and lowering cyber-security concerns. The main difference between the OPF and DR setting, is that the second one involves the participation of self-serving agents, which cannot

be a-priori trusted by the ISOs. This implies that if an agent find it profitable (in terms of its own economic utility), he will compute a different optimization problem from the one provided by the ISO. For this reason, some aspects of DR formulations are better described through a game theoretic framework.

B. Background and previous work

In this setting, we must consider that agents can adopt a strategy $s_i(\theta_i)$, which can be in general different from the one suggested by the ISO, based on their private information (or type), denoted as θ_i, and their belief about the strategy of the other prosumers. The well-known Vickrey-Clarke-Groves (VCG) mechanism [2]–[4] belongs to the strategy-proof class of mechanisms and presents other useful theoretical properties, among which being weakly budget-balanced. Anyway, to achieve this, it requires a value redistribution among agents under the form of monetary taxation, such that the tax which applies to agent i is directly or indirectly independent from its actions. This implies that N optimization problems must be solved, each of which is performed without considering a given agent. This makes the computational cost quadratic in N. Furthermore, VCGs are typically centralized and as such, they do not preserve the privacy of the agents. For example, in [5], a VCG mechanism for virtual inertia is considered, in which bidders send their bidding curves to a center, which solves N independent optimization problems. Since the VCG mechanism guarantees that the best bidding strategy is bidding truthfully, they send their true cost curves $c_i(x_i, \lambda_i)$ to the center. Note anyway that, if the agent's system presents some constraints, $c_i(x_i, \lambda_i)$ must represent them. This means that the center must know all the agent constraint sets $\mathcal{X}_i$ in order to solve the VCG. The unfavorable computational cost makes the VCG impractical for combinatorial auctions [6] and problems with a large number of users with a nontrivial objective function. Despite this and other aspects which make it impractical in some cases [7], VCGs have been extensively studied since they are the only general purpose incentive compatible mechanisms which maximize social welfare [8]. In order to preserve agent's privacy, it is possible to retrieve a distributed formulation of VCG using

primal-dual decomposition algorithms. Note that distributing the mechanism aggravates the scalability problem of VCG, since the overall computation must now take into account communication delays. A second effect of adopting a decentralized formulation is that we cannot guarantee strategyproofness anymore. This is known as the cost of decentralization [9], which leads to a weaker notion of incentive compatibility, namely ex-post Nash equilibrium (EPNE). Although weaker than a dominant-strategy equilibrium, ex-post Nash equilibrium does not require agents to model the strategies nor types of other agents through belief functions, as it's done using Bayes-Nash equilibrium [10]. Following this concept, in [11] guidelines for distributed implementations of VCG mechanisms are derived. In [9] a distributed VCG mechanism which reuses part of the computation done in each subproblem is presented. More recently [12] has proposed a distributed VCG implementation based on dual decomposition, and applied the concept of multistage mechanism, in which different mechanisms are applied at each primal dual update. Also in this case, the proposed algorithm scales quadratically with the number of agents. Another field of research, started with the seminal work of Rosen on n-person non-cooperative games [13], adopt non-VCG mechanisms to reach EPNE [14], [15]. This involves allowing a loss in terms of efficiency [16], with the benefit of better scalability with respect to the number of agents.

In this paper we propose a method to guarantee participation constraint, also known as individual rationality (IR): all the prosumers must have a positive return participating in the proposed energy market, with respect to the base case. We ensure IR allowing a coordinator to limit the Lagrangian multipliers associated to the coupling constraints. The rest of the paper is structured as follows: in II the specific problem we address is formulated and we show that its associate game mapping is monotone, which is a condition for the uniqueness of the VGNE; in III we propose a new algorithm for reaching the GNE, based on the alternating direction method of multipliers (ADMM); in IV we compare the convergence of the aforementioned algorithm with a recently proposed [17] preconditioned forward backward (pFB) algorithm for distributed Nash equilibrium seeking.

II. PROBLEM FORMULATION

In this work we are interested in a more general problem with respect of the OPF. In particular, we consider the case in which a group of agents which produce and/or consume energy (prosumers now on) can sell their aggregated flexibility to third parties, for example to DSO through demand response programs or to balance responsible parties. The mathematical formulation of this problem is known as the sharing problem:

$$\underset{x \in \mathcal{X}}{\text{argmin}} \; e(x) + \sum_{i}^{N} c_i(x_i) \qquad (1)$$
$$s.t. \quad Ax \le b$$

where $\mathcal{X} = \prod_{i=1}^{N} \mathcal{X}_i$ is the Cartesian product of the prosumers feasible sets, $e(x)$ is a system level objective, $c_i(x_i)$

are the costs of each prosumers and the linear constraints are affine coupling constraints between the prosumers and $x = [x_1^T, ..x_N^T] = [x_i]_{i=1}^N$ is the vector of the concatenated actions of all the prosumers. Here the affine coupling constraints encode grid constraints, limiting voltage and power in a subset of selected nodes of the grid in which the agents are located. This is possible taking into account the linearized formulation of the power flow equations [1], [18], whose coefficients can be estimated using phasor measurement units [19], even using smart meter data [20]. The advantage of considering coupling constraints instead of agent-level constraints on voltage and power is given by the fact that the first approach can reach better solutions in terms of total welfare.

As anticipated in the introduction, we are interested in decomposing problem (1) among the self interested prosumers, in such a way that the induced game presents only one variational GNE, and in the algorithms leading to such an equilibrium. Being the equilibrium unique, rational agents will converge to the EPGNE. This is equivalent to assume that the agents believe their own influence on the prices broadcasted by the sequence of mechanism proposed by the algorithm are negligible, i.e. they are price takers.

A reasonable way to turn the centralized problem (1) into a non-cooperative game, is to reward each prosumer with a part of the system level objective $e(x)$, based on the amount of energy he produces or consumes during a give period of time:

$$v(x_i, x_{-i}) = c_i(x_i) + \frac{|x_i|}{\sum_{i=1}^N |x_i|} e(x) \qquad (2)$$

Anyway, this would result in a non-linear and non-convex game. As a first approximation we can replace this repartition rule with fixed (during each horizon) coefficients, based on a moving average:

$$v(x_i, x_{-i}) = c_i(x_i) + \alpha_i e(x) \qquad (3)$$

where

$$\alpha_i = \frac{\sum_{k=t-\tau}^t |x_{i,k}|}{\sum_{k=t-\tau}^t \sum_{i=1}^N |x_{i,k}|} \qquad (4)$$

Note that the game $\mathcal{G}(s_i(x), v_i(x))$ induced by the value functions in (2) defines an aggregative game [21], in which the each prosumer influence other's prosumers value only by means of the aggregated actions. The induced game can be described as the set of optimization problems (5) in which each prosumer minimizes its own value function $v(x_i, x_{-i})$ and associated KKT conditions (6).

$$\begin{cases} \min_{x_i \in \mathcal{X}_i} \; v(x_i, x_{-i}) \\ s.t \quad Ax \le b \end{cases} \quad \forall i \in N \qquad (5)$$

$$KKT(i) = \begin{cases} 0 \in \partial_{x_i} v_i(x_i, \mathrm{x}_{-i}) + \mathrm{N}_{\mathcal{X}_i} + A_i^T \lambda_i \\ 0 \le \lambda_i \perp -(Ax - b) \ge 0 \end{cases} \qquad (6)$$

where $A^T = \left[A_i^T \right]_{i=1}^N$ and $\mathrm{N}_{\mathcal{X}_i}$ is the normal cone operator. Before introducing the algorithms that can be used to solve 5,

we discuss some properties of the proposed objective function. It is known that a sufficient condition for the existence and uniqueness of a NE for a n-person non cooperative game is that the system-level objective function $\sigma(x) = \sum_{i=1}^{N} v_i(x_i)$ is diagonally strictly convex [13]. In the case of affine coupling constraints, authors in [22] and [23] have shown that the game has a unique variational GNE if the pseudogradient of $\sigma(x)$, $\mathcal{F} : \mathbb{R}^{NT} \rightrightarrows \mathbb{R}^{NT} = [\partial_{x_i} v_i(x_i)]_i$ also known as game mapping, is strictly monotone. Furthermore, the equilibrium can be reached making the agents pay $A_i \lambda_i$, that is, the value function of each agent coincides with the integral of the first row of KKT in (6), $\tilde{v}_i = v_i(x_i, x_{-i}) + \lambda^T A_i x_i$. In this case, it has been shown that the agents reach a variational GNE with unique Lagrangian multiplier λ. Note that the game mapping differs from gradient of $\sigma(x)$ since its components are the partial derivatives of the values of the ith agent with respect to its own actions. We now show that the game map generated by the agents' values defined in (2) inherits monotonicity from the convexity of $e(x)$.

Theorem II.1. *Let* $e(x) : \mathbb{R}^{NT} \to \mathbb{R}$ *be a (strictly/strongly) convex function and let the costs of the agents* $c_i(x_i) : \mathbb{R}^{T} \to \bar{\mathbb{R}}$ *be convex functions. Then any repartition* $[\alpha_i]_{i=1}^{N}$ *of* $e(x)$ *among the agents such that:*

1) $v_i(x_i, x_{-i}) = \alpha_i e(x) + c_i(x_i)$
2) $\alpha_i \geq 0 \qquad \forall i \in \{N\}$

generates a (strictly/strongly) monotone game map $\mathcal{F}$: $\mathbb{R}^{NT} \rightrightarrows \mathbb{R}^{NT}$

Proof. $\mathcal{F} = [\partial_{x_i} v_i(x_i, x_{-i})]_{i=1}^{N}$ can be seen as a sum of two operators: $\mathcal{E} = [\partial_{x_i} \alpha_i e(x_i, x_{-i})]_{i=1}^{N}$ and $\mathcal{C} = [\partial_{x_i} c_i(x_i)]_{i=1}^{N}$. Due to the separability of $\mathcal{C}$, it coincides with the gradient of $\sigma(x) = \sum_{i=1}^{N} c_i(x_i)$. Due to the convexity of $\sigma(x)$, $\mathcal{C}$ is a monotone map, since the gradient of a convex function is monotone (theorem 1 in [24]). Using the same reasoning, $\nabla_x e(x)$ is a monotone map due to the convexity of $e(x)$. From the definition of monotonicity, $\langle x - y | \nabla_x e(x) - \nabla_y e(y) \rangle \geq 0 \quad \forall \quad (x, y)$. Additionally, since any convex function must be convex along any path, we can state it component-wise: $(x_i - y_i)(\partial_{x_i} e(x) - \partial_{y_i} e(y)) \geq 0 \quad \forall \quad (i \in \{N\}, x, y)$. Since we defined all α_i as positive, $(x_i - y_i)\alpha_i(\partial_{x_i} e(x) - \partial_{y_i} e(y)) \geq 0 \quad \forall \quad (i \in \{N\}, x, y)$. Thus $\langle x - y | \mathcal{E}(x) - \mathcal{E}(y) \rangle \geq 0 \quad \forall \quad (x, y)$, and $\mathcal{F}$ is monotone being the sum of two monotone operators. $\square$

III. ALGORITHMS FOR GNE SEEKING

As demonstrated in [23], asymmetric projection algorithms [25] can be used to reach a GNE of an aggregative game with quadratic utilities. Recently, the same algorithm has been rigorously derived modeling the GNE as a monotone inclusion [17], showing that it coincides with a preconditioned forward backward (pFB) method (algorithm (1)), which is a special case of the Banach-Picard iteration [26] of two operators whose sum is the set value mapping associated to the KKT conditions in (6).

Algorithm 1 pFB

$x^{k+1} = \mathbf{\Pi}_\mathcal{X} \left[x^k - \alpha(\mathcal{F}(x^k) + A^t \lambda^k)) \right]$
$\lambda^{k+1} = \mathbf{\Pi}_{\mathbb{R}+} \left[\lambda^k + \beta(2Ax^{k+1} - Ax^k - b) \right]$

We compare algorithm (1) with a trivial modification of the ADMM algorithm [27], which convergence rate and properties have been extensively studied in the literature. For clarity of exposition, we start considering the version of problem (1) without coupling constraints. This can be solved in a centralized way through ADMM, applying the procedure in [27] §7.3, which results in the following parallelized formulation:

Algorithm 2 ADMM

$$x_i^{k+1} = \underset{x_i \in X_i}{\operatorname{argmin}} \, c_i(x_i) + \frac{\alpha_i}{2\rho} \|(Sx^k - y^k)/N$$
$$- x_i^k + x_i + \lambda^k\|_2^2$$
$$+ \frac{1}{2\rho} \|(Ax^k - y^k)/N - A_i x_i^k + x_i + \lambda_a^k\|_2^2 \qquad (7)$$

$$y^{k+1} = \underset{y}{\operatorname{argmin}} \, e(y) + \frac{1}{2\rho} \|y - Sx^{k+1} - \lambda^k\| \qquad (8)$$

$$\lambda^{k+1} = \lambda^k + Sx^{k+1} - y^{k+1} \qquad (9)$$

$$y_a^{k+1} = \underset{y}{\operatorname{argmin}} \, \mathcal{I}_{\mathcal{X}_a} + \frac{1}{2\rho} \|y_a - Ax^{k+1} - \lambda^k\| \qquad (10)$$

$$\lambda_a^{k+1} = \lambda_a^k + Ax^{k+1} - y_a^{k+1} \qquad (11)$$

---

where the only difference form the centralized algorithm is the α_i coefficient in the x_i update. We can write the KKT conditions at convergence

$$\begin{cases} \partial_{x_i} c_i(x_i^*) + \alpha_i \frac{\lambda^*}{\rho} + A_i^T \frac{\lambda_a^*}{\rho} + \mathrm{N}_{\mathcal{X}_i} = 0 & \forall i \in N \,\, (12a) \\[2mm] \partial_y e(y^*) - \frac{\lambda^*}{\rho} = 0 & (12b) \\[2mm] y^* = Sx^* & (12c) \\[2mm] 0 \leq \lambda_a^* \perp -(Ax - b) \geq 0 & (12d) \end{cases}$$

We can find λ^* from 12b and substitute it in 12a:

$$\partial_{x_i} c_i(x_i^*) + \alpha_i \partial_y e(y^*) + A_i^T \frac{\lambda_a^*}{\rho} + \mathrm{N}_{\mathcal{X}_i} = 0 \qquad (13)$$

then we can use 12c, and recalling that S is the summation matrix, we obtain:

$$\partial_{x_i} c_i(x_i^*) + \alpha_i \partial_{x_i} e(x^*) + A_i^T \frac{\lambda_a^*}{\rho} + \mathrm{N}_{\mathcal{X}_i} = 0 \qquad (14)$$

which, together with 12d are equivalent to the KKT 6 of the game 5, when $v_i = c_i(x_i) + \alpha_i e(x)$.

A. Pricing and individual rationality

In this paper we only consider the case in which the function $e(x)$ is the surplus that the agent community has in paying the

energy at the point of common coupling with the electrical grid:

$$e(x) = c \left(\sum_{i=1}^{N} x_i \right) - \sum_{i=1}^{N} c(x_i) \qquad (15)$$

where $x_i \in \mathbb{R}^{\mathrm{T}}$ is the vector of total power of the ith agent, $c(\cdot)$ is the energy cost function defined as:

$$c(z_t) = \begin{cases} p_{b,t} z_t, & \text{if} \quad z_t \geq 0 \\ p_{s,t} z_t, & \text{otherwise} \end{cases} \qquad (16)$$

where $p_{b,t}$ and $p_{s,t}$ are the buying and selling tariffs, respectively, at time t. In order to induce agents to follow the proposed mechanism, we must ensure that the energy tariff they pay participating in the market is always lower than the one they pay in the base case. This is always true when we are not taking into account grid constraints, since $e(x)$ as defined in (15) is always non-negative, when $p_{b,t} \geq p_{s,t}$, as usual in energy tariffs. However, if the agents are located in a grid with big voltage oscillations, the Lagrangian dual variables (which we can interpret as punishment prices) could be such that the cost paid by the agents is higher than $\alpha_i e(x)$. To ensure IR, we encode it in the optimization scheme. At each iteration, for each time step in the horizon, we increment the Lagrangians only if the following condition holds:

$$\alpha_i e(x_t^k) + A_i^T \lambda_t^k \leq 0 \qquad \forall i \in N, \quad \forall t \in T \qquad (17)$$

where a negative value means that the prosumer is gaining a reward. This obviously results in the impossibility to satisfy the coupling constraints. We can give the following straightforward economic interpretation to this mechanism: each agent would opt-out from the game as soon as the energy tariffs become unfavorable with respect to the existing one. Condition (17) prevent this from happening. In the presence of bad power quality, the DSO could provide favorable energy tariffs to prosumers participating in the mechanism, ensuring that condition (17) is met with high probability.

B. Prosumers problem formulation

In this paper, each prosumer's flexibility is modeled using an electric battery. Although simple, the model we used is not simplistic and we briefly describe it in this subsection. Since the effect of charging or discharging on the state of charge is not symmetric due to the efficiencies, the problem is usually formulated as a mixed integer linear program (MILP), introducing binary decision variables and using bilinear constraints, to avoid the simultaneous charge and discharge of the battery. Furthermore, the objective function of the agents is non differentiable at $P_m = 0$ and is mathematically described by the maximum operator. In order to speed up the computations, we reformulated all the control problems as a quadratic optimization.

We start considering that both the ADMM and the pFB formulations can be described by the following optimization problem:

$$\underset{x_i \in X_i}{\text{argmin}} \quad f(x_i, x_{-i}) + \frac{1}{2\rho} \|Dx_i - r^k\|_2^2 \qquad (18)$$

where r is a reference signal and $D \in \mathbb{R}^{\mathrm{T} \times 2\mathrm{T}} = \mathrm{I_T} \otimes [1, -1]$, performs the sum of the charging and discharging operations with appropriate signs. Here, with abuse of notation, we redefined the vector $x_i \in \mathbb{R}^{2\mathrm{T}}$ as the vector containing both the charging and discharging operators, in such a way that $x_i = [P_{in,t}; P_{out,t}]_{t=1}^{T}$, where $P_{in,t}$ and $P_{out,t}$ are the charging and discharging powers of the battery. For the ADMM formulation, it is easy to see that problem (18) can be used to solve (7). We can still use (18) for solving the pFB formulation recalling that the projected gradient descent is equivalent to a quadratic optimization problem in the form:

$$\underset{x_i \in X_i}{\text{argmin}} \quad \left(\mathcal{F}_i(x_i^k) + A_i^t \lambda^k \right)^T x_i + \frac{1}{2\rho} \|x_i - x_i^k\|_2^2 \qquad (19)$$

The battery is modeled as a discrete linear system, with the state of charge denoted by s:

$$s_{i,t+1} = A_{s,i} s_{i,t} + B_{s,i} x_{i,t} \qquad (20)$$

We can eliminate the dependence on the state of the optimization problem, using the standard batch formulation:

$$s_i = \Lambda s_{i,0} + \Gamma x_i \qquad (21)$$

where $x_i \in \mathbb{R}^{2\mathrm{T} \times 1}$ is the control vector for the whole time horizon T and $\Lambda \in \mathbb{R}^{\mathrm{T} \times 1}, \Gamma \in \mathbb{R}^{\mathrm{T} \times 2\mathrm{T}}$ are the batch matrices.

We can now describe the set $\mathcal{X}_i$ through the linear constraints $A_{c_i} x_i \leq b_{c_i}$, defined as:

$$A_{c_i} = \begin{bmatrix} I \\ I \\ -\Gamma \\ \Gamma \end{bmatrix} b_{c_i} = \begin{bmatrix} x_{min} \\ x_{max} \\ -e_{min} + \Lambda e_0 \\ e_{max} - \Lambda e_0 \end{bmatrix} \qquad (22)$$

where $x_{min}, x_{max}, e_{min}, e_{max} \in \mathbb{R}^{2\mathrm{T}}$ are the power and energy box constraints, while I is the identity matrix of appropriate dimensions. Now we can reformulate the non differentiable cost function (16) with a linear function, such as we can reuse it in both the ADMM and pFB formulations. We start considering that the if condition of the cost function in (16) can be equivalently formulated using the max operator. In turn, the max operator can be replaced by the sum of an auxiliary variable y and appropriate inequality constraints. We augment our decision variable such as $\tilde{x} = [x^T, y^T]^T$. Now the minimization of (16) is equal to the following optimization problem:

$$\begin{aligned} \min_{\tilde{x}} \quad & l^T \tilde{x} \\ s.t. : \quad & \tilde{A} \tilde{x} \leq \tilde{b} \end{aligned} \qquad (23)$$

where $\tilde{A} = [A_{c,i}; A_y]$ and $\tilde{b} = [b_{c,i}, b_y]$, and

$$A_y = \begin{bmatrix} D \circ P_b - I_T \\ D \circ P_s - I_T \end{bmatrix} \quad b_y = \begin{bmatrix} -p_b P_m \\ -p_s P_m \end{bmatrix} \qquad (24)$$

where the $P_b, P_s \in \mathbb{R}^{\mathrm{T} \times 2\mathrm{T}}$ t_{th} rows entries are identical to the buying and selling prices at time t. The effect of the matrices in (24) is that the new auxiliary variable y is now an upper envelope for the cost function (16). Since we require the cost to be minimized, y will coincide with the cost function

$c(\cdot)$ at optimality. We can now use A_y and b_y in both the ADMM and pFB formulations. While in the first $l^T x_i$ replaces $f(x_i, x_{-i})$ in (18), in the latter the prosumers' energy costs are considered as part of the pseudogradient: $F_i = l + \partial_{x_i} e(x)$.

This problem formulation prevent us from introducing binary variables for the charging and discharging powers, since optimal solutions of (23) does not require to simultaneously charge and discharge the battery at the same time. This is not true for the ADMM formulation, in which the quadratic penalty on the sum of $P_{in,t}$ and $P_{out,t}$ with respect to a reference signal r^k is present. In the case in which the reference signal r^k is negative, the battery is not only incentivized to charge itself, but to consume as much energy as possible. This will result in a simultaneous charge and discharge, due to the round-trip efficiency. To avoid this behavior, $f(\cdot)$ can be augmented with a linear therm, punishing the battery discharging operation when r is negative: $\tilde{f}(\tilde{x}_i) = f(\tilde{x}_i) + l_p^T \tilde{x}_i$ where $l_p \in \mathbb{R}^{1 \times 3*T}$ has non zero entries, all identical to a punishment therms, only when $r < 0$.

IV. NUMERICAL ANALYSIS

We test both the ADMM and the pFB and compare the performance to a centralized solution. The only difference from the ADMM and the centralized formulation are the α_i coefficients in equation (7), which are not present in the centralized solution. Since the system-level objective function $e(x)$, as defined in (15), is not differentiable in 0 and is not strictly nor strongly convex, the convergence of pFB is not guaranteed. To have an equal comparison, we replaced the system-level cost function (16) with a continuously differentiable function. We define it by means of its derivative:

$$\nabla_z \tilde{c}(z) = (p_{b,t} - p_{s,t})\frac{\tanh(kSx_t) + 1}{2} + p_{s,t} \qquad (25)$$

where k regulates the steepness of the function in $z = 0$. In our simulations $k = 10$, which provides a reasonable steepness for all the possible values of the power aggregate, since we did all the computations in per units, and the aggregate power constraint is $Sx \in [-1.1, 1.1]$. We stress out that this approximation is only used for the system-level objective, and not for the prosumers objective functions, where the cost (16) is modeled as described in subsection III-B. In order to fairly compare the algorithms, we used an equal stepsize ρ, fixed to 0.1. The power profiles of each prosumer are randomly chosen from a yearly dataset of real residential electrical consumption. Each prosumer is equipped with a PV field, with a nominal power uniformly distributed between 2 and 10 times its daily energy consumption. Furthermore, each prosumer is provided with an electric battery with size equal to the expected daily energy exceeding its consumption. Figure 1 shows the optimized time series from a single case. In the upper panel, the batteries' state of charge (SOC) are shown. Since the SOC is the time integral of the optimization variables (P_{in}, P_{out}), it is clear that the ADMM and the pFB converged exactly to the same solution. The middle panels shows the forecasted aggregated power profile and the optimized one.
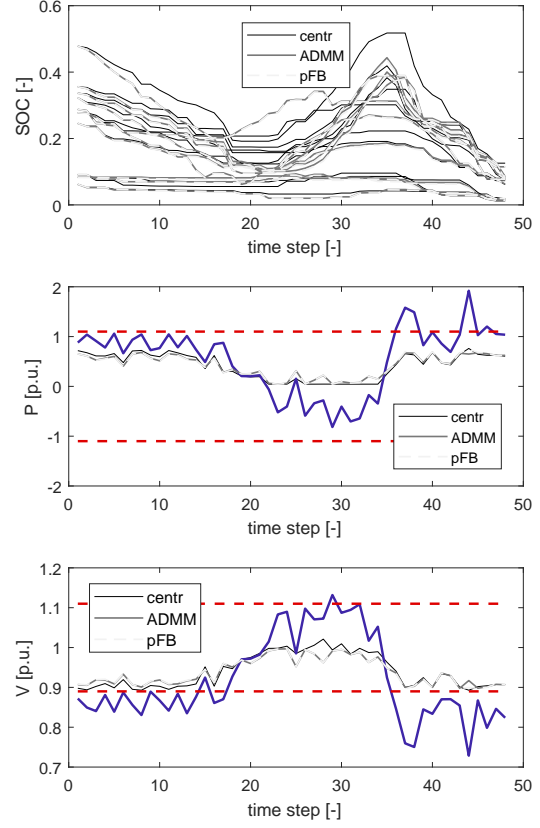
Fig. 1. Time series example, N = 10. Blue: forecasted profiles. Red: constraints. Grays: solutions of the centralized and decentralized approaches. Top: state of charge for each battery. Middle: power profiles. Bottom: voltage profiles.

Note that both the ADMM and pFB solutions are not far from the centralized solution, while differences are more evident in terms of single prosumers SOC. The last panel shows voltage profiles at the point of common coupling. In figure 2 the convergence of the two algorithms is shown, in terms of game objective function $\sigma(x)$. We ran a total of 50 simulations, each of which includes 10 prosumers with power profiles and battery sizes randomly chosen, as explained before. For each simulation s, we retrieve the best optimal value of $\sigma(x)$, p_{best}^s, defined as:

$$p_{best}^s = \text{minimum } \{p_{ADMM}^{*s}, p_{pFB}^{*s}\} \qquad (26)$$

where $p_{ADMM}^{*s}, p_{pFB}^{*s}$ are the solution of the two algorithms after 200 iterations (after which the relative change in $\sigma(x)$ for all the simulations was smaller than $1e - 5$). The thick

lines show the median, while the shadowed patches contain half of the simulations.
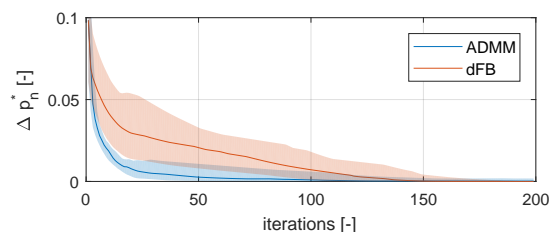
Fig. 2. Normalized optimal value p^*. The thick lines denote the median, while the shaded areas are the 25% and 75% quantiles.

V. Conclusions

We have proposed a method to enforce IR while reaching a EPGNE in a distributed way. The method and the related algorithm have been tested, and compared with pFB, a state of the art algorithm for GNE seeking. The simulations shows that the proposed algorithm reaches the same solutions of pFB, while showing faster convergence in most of the cases. In future research, we will extensively investigate the advantages of the proposed methodology.

References

[1] D. K. Molzahn, F. Dorfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A Survey of Distributed Optimization and Control Algorithms for Electric Power Systems," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, pp. 1–1, 2017.

[2] L. Makowski and J. M. Ostroy, "Vickrey-Clarke-Groves mechanisms and perfect competition," *Journal of Economic Theory*, vol. 42, no. 2, pp. 244–261, 1987.

[3] E. H. Clarke, "Multipart pricing of public goods," *Public Choice*, vol. 11, no. 1, pp. 17–33, 1971.

[4] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders.," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

[5] B. K. Poolla, S. Bolognani, L. Na, and F. Dorfler, "A Market Mechanism for Virtual Inertia," *arXiv*, 2017.

[6] V. Conitzer and T. Sandholm, "Failures of the VCG mechanism in combinatorial auctions and exchanges," *The 5th International Joint Conference on Autonomous and Multiagent Systems*, pp. 521–528, 2006.

[7] M. H. Rothkopf, "Thirteen Reasons Why the Vickrey-Clarke-Groves Process Is Not Practical," *Operations Research*, vol. 55, no. 2, pp. 191–197, 2007.

[8] N. Noam, T. Roughgarden, E. Tardos, and T. Wexler, *Algorithmic Game Theory*. Cambridge University Press, 2007.

[9] A. Petcu, B. Faltings, and D. C. Parkes, "M-DPOP: Faithful distributed implementation of efficient social choice problems," *Journal of Artificial Intelligence Research*, vol. 32, pp. 705–755, 2008.

[10] Y. Narahari, *Game Theory and Mechanism Design*, 2014, vol. 4.

[11] D. C. Parkes and J. Shneidman, "Distributed Implementations of Vickrey-Clarke-Groves Mechanisms," *Third Joint Conference on Autonomous and Multiagent Systems*, pp. 261–268, 2004.

[12] T. Tanaka, F. Farokhi, and C. Langbort, "Faithful implementations of distributed algorithms and control laws," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 2, pp. 191–201, 2017.

[13] J. . B. . Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[14] B.-g. Kim, S. Member, S. Ren, M. V. D. Schaar, J.-w. Lee, and S. Member, "Bidirectional Energy Trading and Residential Load Scheduling with Electric Vehicles in the Smart Grid," *IEEE Journal on selected areas in communication*, vol. 31, no. 7, pp. 1219–1234, 2013.

[15] B. Gharesifard, T. Baar, and A. D. Domínguez-García, "Price-based coordinated aggregation of networked distributed energy resources," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2936–2946, 2016.

[16] B. Yang and M. Johansson, "Distributed optimization and games: A tutorial overview," in *Networked Control Systems*, 2010, vol. 406, pp. 109–148.

[17] G. Belgioioso and S. Grammatico, "Projected-gradient algorithms for generalized equilibrium seeking in Aggregative Games are preconditioned Forward-Backward methods," 2018. [Online]. Available: http://arxiv.org/abs/1803.10441

[18] H. Almasalma, J. Engels, and G. Deconinck, "Dual-decomposition-based peer-to-peer voltage control for distribution networks," *CIRED - Open Access Proceedings Journal*, no. 1, pp. 1718–1721, 2017.

[19] C. Mugnier, K. Christakou, J. Jaton, M. De Vivo, M. Carpita, and M. Paolone, "Model-less/measurement-based computation of voltage sensitivities in unbalanced electrical distribution networks," *19th Power Systems Computation Conference, PSCC 2016*, 2016.

[20] S. Weckx, S. Member, R. D. Hulst, J. Driesen, and S. Member, "Voltage Sensitivity Analysis of a Laboratory Distribution Grid With Incomplete Data," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1271–1280, 2015.

[21] M. K. Jensen, "Aggregative games and best-reply potentials," *Economic Theory*, vol. 43, no. 1, pp. 45–66, 2010.

[22] G. Belgioioso and S. Grammatico, "Semi-Decentralized Nash Equilibrium Seeking in Aggregative Games With Separable Coupling Constraints and Non-Differentiable Cost Functions," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 400–405, 2017.

[23] D. Paccagnan, M. Kamgarpour, B. Gentile, F. Parise, J. Lygeros, and D. Paccagnan, "Distributed computation of Nash Equilibria in aggregative games with coupling constraints," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Las Vegas, USA, 2016, pp. 6123–6128.

[24] G. J. Minty, "On the Monotonicity of the Gradient of a Convex Function," *Pacific J. Math.*, vol. 14, no. 1, pp. 243–248, 1963.

[25] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, 2015, vol. 1542, no. 9.

[26] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2011.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

Bibliography

[1] R. Torquato, D. Salles, C. O. Pereira, P. C. M. Meira, and W. Freitas, "A Comprehensive Assessment of PV Hosting Capacity on Low-Voltage Distribution Systems," *IEEE Transactions on Power Delivery*, 2018.

[2] M. A. Eltawil and Z. Zhao, "Grid-connected photovoltaic power systems: Technical and potential problems-A review," 2010.

[3] K. Reddy, M. Kumar, T. Mallick, H. Sharon, and S. Lokeswaran, "A review of Integration, Control, Communication and Metering (ICCM) of renewable energy based smart grid," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 180–192, oct 2014.

[4] T. Aziz and N. Ketjoy, "PV Penetration Limits in Low Voltage Networks and Voltage Variations," 2017.

[5] C. W. Gellings, "The Concept of Demand-Side Management for Electric Utilities," *Proceedings of the IEEE*, 1985.

[6] H. A. Gabbar, "Perspectives of Demand-Side Management Under Smart Grid Concept," in *Energy Conservation in Residential, Commercial, and Industrial Facilities*, IEEE, 2018.

[7] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, 2008.

[8] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, 2011.

[9] A. F. Meyabadi and M. H. Deihimi, "A review of demand-side management: Reconsidering theoretical framework," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 367–379, 2017.

[10] A. Bari, J. Jiang, W. Saad, and A. Jaekel, "Challenges in the Smart Grid Applications: An Overview," *International Journal of Distributed Sensor Networks*, vol. 2014, pp. 1–11, 2014.

[11] A. H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, 2010.

[12] Z. Ming, G. Yajing, and L. Gengyin, "Study on improvement of available transfer capability by demand side management," in *3rd International Conference on Deregulation and Restructuring and Power Technologies, DRPT 2008*, 2008.

[13] P. Scott and S. Thiebaux, "Identification of Manipulation in Receding Horizon Electricity Markets," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, pp. 1–12, 2017.

[14] P. Velarde, J. M. Maestre, H. Ishii, and R. R. Negenborn, "Vulnerabilities in Lagrange-Based DMPC in the Context of Cyber-Security," *2017 IEEE International Conference on Autonomic Computing (ICAC)*, pp. 215–220, 2017.

[15] S. Weerakkody, X. Liu, S. H. Son, and B. Sinopoli, "A Graph-Theoretic Characterization of Perfect Attackability for Secure Design of Distributed Control Systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 60–70, 2017.

[16] N. P. Padhy, "Unit commitment - A bibliographical survey," *IEEE Transactions on Power Systems*, 2004.

[17] A. Castillo, C. Laird, C. A. Silva-Monroy, J. P. Watson, and R. P. O'Neill, "The Unit Commitment Problem with AC Optimal Power Flow Constraints," *IEEE Transactions on Power Systems*, 2016.

[18] Y. Fu, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with AC constraints," *IEEE Transactions on Power Systems*, 2005.

[19] Y. Bai, H. Zhong, Q. Xia, C. Kang, and L. Xie, "A decomposition method for network-constrained unit commitment with AC power flow constraints," *Energy*, vol. 88, pp. 595–603, 2015.

[20] F. Capitanescua, J. M. Ramos, P. Panciaticic, D. Kirschend, A. M. Marcolinib, L. Platbroode, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flo," *Electric Power Systems Research 81 (2011) 1731–1741 Contents*, vol. 81, pp. 1731–1741, 2011.

[21] H. Abdi, S. D. Beigvand, and M. L. Scala, "A review of optimal power flow studies applied to smart grids and microgrids," *Renewable and Sustainable Energy Reviews*, vol. 71, no. May 2015, pp. 742–766, 2017.

[22] M. Aien, A. Hajebrahimi, and M. F.-f. Fellow, "A comprehensive review on uncertainty modeling techniques in power system studies," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 1077–1089, 2016.

[23] A. R. Jordehi, "How to deal with uncertainties in electric power systems ? A review," *Renewable and Sustainable Energy Reviews*, vol. 96, no. August, pp. 145–155, 2018.

[24] Q. P. Zheng, J. Wang, S. Member, and A. L. Liu, "Stochastic Optimization for Unit Commitment — A Review," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913–1924, 2015.

[25] D. K. Molzahn, F. Dorfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A Survey of Distributed Optimization and Control Algorithms for Electric Power Systems," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, p. 1, 2017.

[26] S. Frank and S. Rebennack, "An introduction to optimal power flow: Theory, formulation, and examples," *IIE Transactions (Institute of Industrial Engineers)*, vol. 48, no. 12, pp. 1172–1197, 2016.

[27] M. Nick, R. Cherkaoui, J. Y. L. Boudec, and M. Paolone, "An Exact Convex Formulation of the Optimal Power Flow in Radial Distribution Networks Including Transverse Components," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 682–697, 2018.

[28] K. Christakou, D.-C. Tomozei, J.-Y. Le Boudec, and M. Paolone, "AC OPF in radial distribution networks – Part II: An augmented Lagrangian-based OPF algorithm, distributable via primal decomposition," *Electric Power Systems Research*, vol. 150, pp. 24–35, 2017.

[29] Q. P. Zheng, J. Wang, S. Member, and A. L. Liu, "Stochastic Optimization for Unit Commitment — A Review," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913–1924, 2015.

[30] D. Bertsimas and M. Sim, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, 2006.

[31] G. C. Calafiore, F. Dabbene, and R. Tempo, "Research on probabilistic methods for control system design," *Automatica*, 2011.

[32] L. Magni, D. Pala, and R. Scattolini, "Stochastic Model Predictive Control of Constrained Linear Systems with Additive Uncertainty," *Proceedings of the European Control Conference 2009*, 2009.

[33] M. Farina, L. Giulioni, L. Magni, and R. Scattolini, "An approach to output-feedback MPC of stochastic linear discrete-time systems," in *Automatica*, 2015.

[34] A. Nemirovski and A. Shapiro, "Convex Approximations of Chance Constrained Programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.

[35] J. M. Grosso, J. M. Maestre, C. Ocampo-Martinez, and V. Puig, *On the assessment of tree-based and chance-constrained predictive control approaches applied to drinking water networks*, vol. 19. IFAC, 2014.

[36] F. Oldewurtel, C. N. Jones, and M. Morari, "A tractable approximation of chance constrained stochastic MPC based on affine disturbance feedback," *2008 47th IEEE Conference on Decision and Control*, pp. 4731–4736, 2008.

[37] G. Schildbach, L. Fagiano, C. Frei, and M. Morari, "The scenario approach for Stochastic Model Predictive Control with bounds on closed-loop constraint violations," *Automatica*, vol. 50, no. 12, pp. 3009–3018, 2014.

[38] J. Matusko and F. Borrelli, "Scenario-based approach to stochastic linear predictive control," in *Proceedings of the IEEE Conference on Decision and Control*, 2012.

[39] M. Farina, L. Giulioni, and R. Scattolini, "Stochastic linear Model Predictive Control with chance constraints - A review," *Journal of Process Control*, vol. 44, pp. 53–67, 2016.

[40] J. Dupacova, G. Consigli, and S. W. Wallace, "Scenarios for multistage stochastic programs," *Annals of Operations Research*, 2000.

[41] N. Gröwe-Kuska and W. Römisch, *Stochastic unit commitment in hydrothermal power production planning.* 2005.

[42] D. Bernardini and A. Bemporad, "Scenario-based Model Predictive Control of Stochastic Constrained Linear Systems," in *Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, no. 2, pp. 6333–6338, 2009.

[43] P. Patrinos, S. Trimboli, and A. Bemporad, "Stochastic MPC for real-time market-based optimal power dispatch," in *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pp. 7111–7116, 2011.

[44] A. Bemporad, L. Bellucci, and T. Gabbriellini, "Dynamic option hedging via stochastic model predictive control based on scenario simulation," *Quantitative Finance*, 2014.

[45] J. M. Maestre, L. Raso, P. J. van Overloop, and B. De Schutter, "Distributed tree-based model predictive control on a drainage water system," *Journal of Hydroinformatics*, vol. 15, no. 2, p. 335, 2013.

[46] J. M. Maestre, L. Raso, P. van Overloop, and B. de Schutter, "Distributed tree-based model predictive control on an open water system," *American Control Conference (ACC)*, pp. 1985–1990, 2012.

[47] P. Velarde, L. Valverde, J. M. Maestre, C. Ocampo-Martinez, and C. Bordons, "On the comparison of stochastic model predictive control strategies applied to a hydrogen-based microgrid," *Journal of Power Sources*, 2017.

[48] N. Gröwe-Kuska, H. Heitsch, and W. Römisch, "Scenario reduction and scenario tree construction for power management problems," in *2003 IEEE Bologna PowerTech - Conference Proceedings*, 2003.

[49] J. Dupačová, N. Gröwe-Kuska, and W. Römisch, "Scenario reduction in stochastic programming An approach using probability metrics," *Mathematical Programming, Series B*, 2003.

[50] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap. Monographs on Statistics and Applied Probability, No. 57. Chapman and Hall, London, 436 p.* 1993.

[51] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," *Elements*, vol. 1, pp. 337–387, 2009.

[52] R. J. Hyndman, *Forecasting: Principles and Practice*. 2018.

[53] A. K. Sampathirao, J. M. Grosso, P. Sopasakis, C. Ocampo-Martinez, A. Bemporad, and V. Puig, "Water demand forecasting for the optimal operation of large-scale Drinking Water Networks: The barcelona case study," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2014.

[54] S. Takriti, J. R. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Transactions on Power Systems*, 1996.

[55] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, 2009.

[56] A. J. Patton and Others, "Copula methods for forecasting multivariate time series," *Handbook of economic forecasting*, vol. 2, no. April, pp. 899–960, 2012.

[57] R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting, "Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling," *Statistical Science*, 2013.

[58] Z. B. Bouallegue, T. Heppelmann, S. E. Theis, and P. Pinson, "Generation of scenarios from calibrated ensemble forecasts with a dual ensemble copula coupling approach," pp. 4737–4750, 2016.

[59] P. Arbenz, C. Hummel, and G. Mainik, "Copula based hierarchical risk aggregation through sample reordering," *Insurance: Mathematics and Economics*, vol. 51, no. 1, pp. 122–133, 2012.

[60] K. Bruninx, E. Delarue, and W. D'haeseleer, "A Practical Approach on Scenario Generation & Reduction Algorithms for Wind Power Forecast Error Scenarios." 2014.

[61] N. Parikh and S. Boyd, "Proximal Algorithms," *Foundations and Trends®in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[62] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2011.

[63] G. Belgioioso and S. Grammatico, "Semi-Decentralized Nash Equilibrium Seeking in Aggregative Games With Separable Coupling Constraints and Non-Differentiable Cost Functions," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 400–405, 2017.

[64] D. Paccagnan, M. Kamgarpour, B. Gentile, F. Parise, J. Lygeros, and D. Paccagnan, "Distributed computation of Nash Equilibria in aggregative games with coupling constraints," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, (Las Vegas, USA), pp. 6123–6128, 2016.

Bibliography

[65] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes," *SustKDD*, no. August, p. 6, 2012.

[66] T. V. Jensen and P. Pinson, "Data Descriptor: RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system," *Scientific Data*, vol. 4, pp. 1–18, 2017.

[67] S. Hosseini, S. Kelouwani, K. Agbossou, A. Cardenas, and N. Henao, "A semi-synthetic dataset development tool for household energy consumption analysis," *Proceedings of the IEEE International Conference on Industrial Technology*, pp. 564–569, 2017.

[68] T. Afjei, U. Schonhardt, C. Wemhöner, M. Erb, H. R. Gabathuler, H. Mayer, G. Zweifel, M. Achermann, R. von Euw, and U. Stöckli, "Standardschaltungen für Kleinwärmepumpenanlagen Teil 2: Grundlagen und Computersimulationen. Schlussbericht," tech. rep., 2002.

[69] L. Girardin, "A GIS-based Methodology for the Evaluation of Integrated Energy Systems in Urban Area," p. 218, 2012.

[70] K. Arendt, M. Jradi, H. R. Shaker, C. T. Veje, and S. Denmark, "Comparative Analysis of white- , gray- and black-BOX models for thermal simulation of indoor environment: teaching building case study," in *Building Performance Modeling Conference*, pp. 173–180, 2018.

[71] G. Reynders, J. Diriken, and D. Saelens, "Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals," *Energy and Buildings*, vol. 82, pp. 263–274, 2014.

[72] P. Bacher and H. Madsen, "Identifying suitable models for the heat dynamics of buildings," *Energy and Buildings*, vol. 43, pp. 1511–1522, jul 2011.

[73] F. S. Office, "Bâtiments: aperçu général selon les cantons 2016," 2016.

[74] L. Pampuri, N. Cereghetti, P. G. Bianchi, and P. Caputo, "Evaluation of the space heating need in residential buildings at territorial scale: The case of Canton Ticino (CH)," *Energy and Buildings*, vol. 148, pp. 218–227, 2017.

[75] European Commission, "European Building Database."

[76] F. P. Incropera, D. P. DeWitt, T. L. Bergman, and A. S. Lavine, *Fundamentals of Heat and Mass Transfer*. 2007.

[77] T. Cholewa, M. Rosiński, Z. Spik, M. R. Dudzińska, and A. Siuta-Olcha, "On the heat transfer coefficients between heated/cooled radiant floor and room," *Energy and Buildings*, vol. 66, pp. 599–606, 2013.

[78] P. Wallenten, "Heat Transfer Coefficients in a Full Scale Room With and Without Furniture," *Lund Institute of Thechnology*, pp. 1–8, 1999.

[79] T. Schütz, H. Harb, R. Streblow, and D. Müller, "Comparison of models for thermal energy storage units and heat pumps in mixed integer linear programming," *PROCEEDINGS OF ECOS 2015 - THE 28TH INTERNATIONAL CONFERENCE ON EFFICIENCY, COST, OPTIMIZATION, SIMULATION AND ENVIRONMENTAL IMPACT OF ENERGY SYSTEMS JUNE 30-JULY 3, 2015, PAU, FRANCE*, 2015.

[80] R. D. Coninck, R. Baetens, D. Saelens, A. Woyte, L. Helsen, A. Mechanics, and B. P. Section, "Rule-based demand side management of domestic hot water production with heat pumps in zero energy neighbourhoods," *Journal of Building Performance Simulation*, 2013.

[81] L. Nespoli, A. Giusti, N. Vermes, M. Derboni, E. Rizzoli, and L. M. Gambardella, "Distributed demand side management using electric boilers," *Computer Science - Research and Development*, 2016.

[82] J. S. Stein, "The photovoltaic Performance Modeling Collaborative (PVPMC)," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 3048–3052, 2012.

[83] E. L. Maxwell, "A quasi-physical model for converting hourly global to direct normal insolation," 1987.

[84] P. G. Loutzenhiser, H. Manz, C. Felsmann, P. A. Strachan, T. Frank, and G. M. Maxwell, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254–267, 2007.

[85] B. Marion and B. Smith, "Photovoltaic system derived data for determining the solar resource and for modeling the performance of other photovoltaic systems," *Solar Energy*, vol. 147, pp. 349–357, 2017.

[86] G. Valentin, "Design and Simulation of Photovoltaic Systems Manual," pp. 1–82, 2012.

[87] ASHRAE, "Methods of Testing to Determine the Thermal Performance of Solar Collectors, ASHRAE Standard 93-77," 1978.

[88] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy*, vol. 83, no. 5, pp. 614–624, 2009.

[89] R. S. Sutton, A. G. Barto, and F. Bach, *Reinforcement Learning: An Introduction, second edition.* A Bradford Book, 2018.

[90] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, "Optimal combination forecasts for hierarchical time series," *Computational Statistics and Data Analysis*, 2011.

Bibliography

[91] D. Yang, H. Quan, V. R. Disfani, and L. Liu, "Reconciling solar forecasts: Geographical hierarchy," *Solar Energy*, 2017.

[92] D. Yang, H. Quan, V. R. Disfani, and C. D. Rodríguez-Gallegos, "Reconciling solar forecasts: Temporal hierarchy," *Solar Energy*, 2017.

[93] P. Tatjewski, "Advanced control and on-line process optimization in multilayer structures," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 11, no. PART 1, pp. 13–26, 2007.

[94] M. Grochowski, M. A. Brdys, and T. Gminski, "Intelligent Control Structure for Control of Integrated Wastewater Systems," *IFAC Proceedings Volumes*, vol. 37, no. 11, pp. 245–250, 2004.

[95] C. C. Sun, V. Puig, and G. Cembrano, "Temporal multi-level coordination techniques oriented to regional water networks: application to the Catalunya case study," *Journal of Hydroinformatics*, vol. 16, no. 4, p. 952, 2014.

[96] V. Puig, C. Ocampo-Martinez, and S. Montes De Oca, "Hierarchical temporal multi-layer decentralised MPC strategy for drinking water networks: Application to the Barcelona case study," in *2012 20th Mediterranean Conference on Control and Automation, MED 2012 - Conference Proceedings*, 2012.

[97] F. Sossan, E. Namor, R. Cherkaoui, and M. Paolone, "Achieving the Dispatchability of Distribution Feeders Through Prosumers Data Driven Forecasting and Model Predictive Control of Electrochemical Storage," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, pp. 1762–1777, 2016.

[98] R. Cagienard, P. Grieder, E. C. Kerrigan, and M. Morari, "Move blocking strategies in receding horizon control," *Journal of Process Control*, vol. 17, no. 6, pp. 563–570, 2007.

[99] R. C. Shekhar and C. Manzie, "Optimal move blocking strategies for model predictive control," *Automatica*, vol. 61, pp. 27–34, 2015.

[100] A. Mathematics, "The Proximal Operator," 2017.

[101] T. Pan, *Exploring Model Predictive Control Applications for the Gridbox Pilot.* PhD thesis, ETH - Swiss Federal Institute of Technology Zurich, 2016.

[102] L. Ljung, *System Identification Theory for User.* 1987.

[103] J. D. Wichard, "Forecasting the NN5 time series with hybrid models," *International Journal of Forecasting*, 2011.

[104] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7067–7083, 2012.

[105] C. M. Bishop, *Patterns Recognition and Machine Learning*. 2006.

[106] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to Statistical Learning*, vol. 7. 2000.

[107] N. Meinshausen, "Quantile Regression Forests Nicolai," *Journal of Machine Learning Research*, vol. 131, pp. 65–78, 2014.

[108] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review*, 2015.

[109] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *International Journal of Forecasting*, 2004.

[110] E. S. Gardner, "Exponential smoothing: The state of the art," *Journal of Forecasting*, 1985.

[111] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation; With Application to Solar Energy," *Power Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–14, 2016.

[112] D. Yang, "Solar radiation on inclined surfaces: Corrections and benchmarks," *Solar Energy*, vol. 136, pp. 288–302, 2016.

[113] P. G. Loutzenhiser, H. Manz, C. Felsmann, P. A. Strachan, T. Frank, and G. M. Maxwell, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254–267, 2007.

[114] M. Lef, A. Oumbe, P. Blanc, B. Espinar, Z. Qu, L. Wald, M. S. Homscheidt, A. Arola, M. Lef, A. Oumbe, P. Blanc, and B. Espinar, "McClear : a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, pp. 2403—-2418, 2013.

[115] H. Shaker, H. Chitsaz, H. Zareipour, and D. Wood, "On comparison of two strategies in net demand forecasting using Wavelet Neural Network," *2014 North American Power Symposium, NAPS 2014*, 2014.

[116] Y. Wang, N. Zhang, Q. Chen, D. S. Kirschen, P. Li, and Q. Xia, "Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3255–3264, 2018.

[117] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "Trend Filtering.," *SIAM Review*, vol. 51, no. 2, pp. 339–360, 2009.

[118] R. J. Tibshirani, "Adaptive Piecewise Polynomial Estimation via Trend Filtering," *arXiv*, 2013.

[119] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 67, no. 1, pp. 91–108, 2005.

[120] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[121] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013.

[122] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[123] L. Nespoli and V. Medici, "An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements," *Solar Energy*, 2017.

[124] "MeteoBlue https://www.meteoblue.com/."

[125] S. L. Wickramasuriya and G. Athanasopoulos, "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization," *Working Paper*, no. June, 2017.

[126] S. B. Taieb, R. Rajagopal, S. Ben Taieb, J. Yu, M. Neves Barreto, and R. Rajagopal, "Regularization in Hierarchical Time Series Forecasting With Application to Electricity Smart Meter Data," *Aaai*, no. 2011, pp. 4474–4480, 2017.

[127] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends®in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[128] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Applications in Genetics and Molecular Biology*, 2005.

[129] A. Beck, *First-Order Methods in Optimization.* Society for Industrial and Applied Mathematics, 2017.

[130] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Coherent Probabilistic Forecasts for Hierarchical Time Series," *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, no. April, pp. 3348–3357, 2017.

[131] K. Pavel and S. David, "Algorithms for Efficient Computation of Convolution," in *Design and Architectures for Digital Signal Processing* (G. Ruiz and J. A. Michell, eds.), ch. 8, Rijeka: IntechOpen, 2013.

[132] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1460–1495, 2014.

[133] G. Belgioioso and S. Grammatico, "Projected-gradient algorithms for Generalized Equilibrium seeking in Aggregative Games are preconditioned Forward-Backward methods," *arXiv*, 2018.

[134] B. Yang and M. Johansson, "Distributed Optimization and Games: A Tutorial Overview," in *Networked Control Systems* (A. Bemporad, M. Heemels, and M. Johansson, eds.), pp. 109–148, London: Springer London, 2010.

[135] B. HomChaudhuri and M. Kumar, "A newton based distributed optimization method with local interactions for large-scale networked optimization problems," pp. 4336–4341, 2014.

[136] E. Ghadimi, I. Shames, and M. Johansson, "Multi-step gradient methods for networked optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5417–5429, 2013.

[137] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distribued optimization in netowrked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[138] R. Scattolini, "Architectures for distributed and hierarchical Model Predictive Control - A review," *Journal of Process Control*, vol. 19, no. 5, pp. 723–731, 2009.

[139] R. Negenborn and J. Maestre, "Distributed Model Predictive Control: An overview of features and research opportunities," *Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control*, no. April, pp. 530–535, 2014.

[140] R. Halvgaard, L. Vandenberghe, N. K. Poulsen, H. Madsen, and J. B. Jørgensen, "Distributed Model Predictive Control for Smart Energy Systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 1–9, 2014.

[141] P. Braun, T. Faulwasser, L. Gr, C. M. Kellett, S. R. Weller, and K. Worthmann, "Hierarchical Distributed ADMM for Predictive Control with Applications in Power Networks," pp. 1–10.

[142] M. Diekerhof, F. Peterssen, and A. Monti, "Hierarchical Distributed Robust Optimization for Demand Response Services," *IEEE Transactions on Smart Grid*, vol. 3053, no. c, p. 1, 2017.

[143] M. Kraning, E. Chu, and S. Boyd, "Dynamic Network Energy Management via Proximal Message Passing," *Foundations and Trends in Optimization*, vol. 1, no. 2, pp. 70–122, 2013.

[144] T. G. Hovgaard, K. Edlund, and J. B. Jørgensen, "The Potential of Economic MPC for Power Management," pp. 7533–7538, 2010.

[145] P. McNamara and S. McLoone, "Hierarchical Demand Response for Peak Decomposition," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2807–2815, 2015.

[146] K. Margellos, A. Falsone, S. Garatti, and M. Prandini, "Distributed Constrained Optimization and Consensus in Uncertain Networks via Proximal Minimization," *IEEE Transactions on Automatic Control*, vol. 9286, no. c, pp. 1–16, 2017.

[147] P. Jacquot, O. Beaude, S. Gaubert, and N. Oudjane, "Demand side management in the smart grid: An efficiency and fairness tradeoff," *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2017 - Proceedings*, vol. 2018-Janua, no. November, pp. 1–6, 2018.

[148] J. Jian, C. Zhang, L. Yang, K. Meng, Y. Xu, and Z. Dong, "A Hierarchical Alternating Direction Method of Multipliers for Fully Distributed Unit Commitment," 2006.

[149] R. Gupta, F. Sossan, E. Scolari, E. Namor, L. Fabietti, C. Jones, and M. Paolone, "An ADMM-based Coordination and Control Strategy for PV and Storage to Dispatch Stochastic Prosumers : Theory and Experimental Validation," in *Proceedings of the 20th Power Systems Computation Conference, -, -, 1-7*, pp. 1–7, 2018.

[150] D. Bao Nguyen, J. M. A. Scherpen, and F. Bliek, "Distributed Optimal Control of Smart Electricity Grids with Congestion Management," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 494–504, 2017.

[151] H. Almasalma, J. Engels, and G. Deconinck, "Dual-decomposition-based peer-to-peer voltage control for distribution networks," *CIRED - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1718–1721, 2017.

[152] E. Sorin, L. Bobo, and P. Pinson, "Consensus-based Approach to Peer-to-Peer Electricity Markets with Product Differentiation," *IEEE Transactions on Power Systems*, vol. X, no. X, pp. 1–8, 2018.

[153] F. Moret, P. Pinson, and S. Member, "Energy Collectives : a Collaborative Approach to Future Consumer-Centric Electricity Markets," *IEEE Transactions on Power Systems*, pp. 1–2, 2018.

[154] T. Erseghe, "Distributed optimal power flow using ADMM," *IEEE Transactions on Power Systems*, 2014.

[155] M. Ye, S. Member, and G. Hu, "Distributed Extremum Seeking for Constrained Networked Optimization and Its Application to Energy Consumption Control in Smart Grid," vol. 24, no. 6, pp. 1–11, 2016.

[156] L. Dai, Y. Xia, and Y. Gao, "Distributed model predictive control of linear systems with stochastic parametric uncertainties and coupled probabilistic constraints," *SIAM Journal on Control and Optimization*, vol. 53, no. 6, pp. 3411–3431, 2017.

[157] C. Li, X. Yu, W. Yu, and T. Huang, "Networked Optimization for Demand Side Management based on Non-cooperative Game," *IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015.

[158] S. Mhanna, G. Verbiˇ, and A. C. Chapman, "Response Aggregation," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1–11, 2016.

[159] M. Ye and G. Hu, "Distributed nash equilibrium seeking by a consensus based approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4811–4818, 2017.

[160] B.-g. Kim, S. Member, S. Ren, M. V. D. Schaar, J.-w. Lee, and S. Member, "Bidirectional Energy Trading and Residential Load Scheduling with Electric Vehicles in the Smart Grid," *IEEE Journal on selected areas in communication*, vol. 31, no. 7, pp. 1219–1234, 2013.

[161] P. Yi and L. Pavel, "Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms," *IEEE Transactions on Control of Network Systems*, 2018.

[162] S. Grammatico, "Dynamic Control of Agents Playing Aggregative Games with Coupling Constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4537–4548, 2017.

[163] S. Kar, G. Hug, J. Mohammadi, S. Member, and J. M. F. Moura, "Distributed State Estimation and Energy Management in Smart Grids : A Consensus Innovations Approach," *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 8, NO. 6*, vol. 8, no. 6, pp. 1022–1038, 2014.

[164] L. Makowski and J. M. Ostroy, "Vickrey-Clarke-Groves mechanisms and perfect competition," *Journal of Economic Theory*, vol. 42, no. 2, pp. 244–261, 1987.

[165] E. H. Clarke, "Multipart pricing of public goods," *Public Choice*, vol. 11, no. 1, pp. 17–33, 1971.

[166] W. Vickrey, "COUNTERSPECULATION, AUCTIONS, AND COMPETITIVE SEALED TENDERS," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

[167] N. Noam, T. Roughgarden, E. Tardos, and T. Wexler, *Algorithmic Game Theory*. Cambridge University Press, 2007.

[168] D. C. Parkes and J. Shneidman, "Distributed Implementations of Vickrey-Clarke-Groves Mechanisms," *Third Joint Conference on Autonomous and Multiagent Systems*, pp. 261–268, 2004.

Bibliography

[169] A. Petcu, B. Faltings, and D. C. Parkes, "M-DPOP: Faithful distributed implementation of efficient social choice problems," *Journal of Artificial Intelligence Research*, vol. 32, pp. 705–755, 2008.

[170] A. R. Kian and J. B. Cruz, "Bidding strategies in dynamic electricity markets," vol. 40, pp. 543–551, 2005.

[171] B. K. Poolla, S. Bolognani, L. Na, and F. Dorfler, "A Market Mechanism for Virtual Inertia," *arXiv*, 2017.

[172] S. Park, J. Lee, S. Bae, G. Hwang, and J. K. Choi, "Contribution-Based Energy-Trading Mechanism in Microgrids for Future Smart Grid: A Game Theoretic Approach," *IEEE Transactions on Industrial Electronics*, 2016.

[173] B. G. Kim, S. Ren, M. Van Der Schaar, and J. W. Lee, "Bidirectional energy trading for residential load scheduling and electric vehicles," in *Proceedings - IEEE INFOCOM*, 2013.

[174] P. Yi and L. Pavel, "A distributed primal-dual algorithm for computation of generalized Nash equilibria with shared affine coupling constraints via operator splitting methods," 2017.

[175] J. B. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[176] B.-g. Kim, S. Ren, M. V. D. Schaar, and J.-w. Lee, "Bidirectional Energy Trading for Residential Load Scheduling and Electric Vehicles," pp. 595–599, 2013.

[177] A. A. Kulkarni and U. V. Shanbhag, "On the variational equilibrium as a refinement of the generalized Nash equilibrium," *Automatica*, 2012.

[178] S. A. Arafeh, "Hierarchical Control of Power Distribution Systems," *IEEE Transactions on Automatic Control*, vol. AC-23, no. 2, pp. 333–343, 1978.

[179] B. P. Bhattarai, B. Bak-Jensen, P. Mahat, J. R. Pillai, and M. Maier, "Hierarchical control architecture for demand response in smart grids," *Power and Energy Engineering Conference (APPEEC), 2013 IEEE PES Asia-Pacific*, pp. 1–6, 2013.

[180] K. Baker, J. Guo, and S. Member, "Distributed MPC for Efficient Coordination of Storage and Renewable Energy Sources Across Control Areas," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 992–1001, 2016.

[181] X. Su, M. A. S. Masoum, and P. J. Wolfs, "Multi-Objective Hierarchical Control of Unbalanced Distribution Networks to Accommodate More Renewable Connections in the Smart Grid Era," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3924–3936, 2016.

[182] M. Hong, Z. Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, pp. 337–364, 2016.

172

[183] K. Christakou, J. Y. Leboudec, M. Paolone, and D. C. Tomozei, "Efficient computation of sensitivity coefficients of node voltages and line currents in unbalanced radial electrical distribution networks," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 741–750, 2013.

[184] C. Mugnier, K. Christakou, J. Jaton, M. De Vivo, M. Carpita, and M. Paolone, "Model-less/measurement-based computation of voltage sensitivities in unbalanced electrical distribution networks," *19th Power Systems Computation Conference, PSCC 2016*, 2016.

[185] S. Weckx, S. Member, R. D. Hulst, J. Driesen, and S. Member, "Voltage Sensitivity Analysis of a Laboratory Distribution Grid With Incomplete Data," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1271–1280, 2015.

[186] D. Deka, S. Backhaus, and M. Chertkov, "Learning topology of the power distribution grid with and without missing data," in *2016 European Control Conference, ECC 2016*, 2016.

[187] D. Deka, S. Backhaus, and M. Chertkov, "Structure Learning in Power Distribution Networks," *IEEE Transactions on Control of Network Systems*, vol. PP, 2015.

[188] IEEE, "IEEE test feeder http://sites.ieee.org/pes-testfeeders/resources/," 2015.

[189] J. Feigenbaum, S. Shenker, D. Algorithmic, and M. Design, "Distributed Algorithmic Mechanism Design : Recent Results and Future Directions," 2002.

[190] "Ordinanza sull'energia 730.01, https://www.admin.ch/opc/it/classified-compilation/20162945/index.html."

[191] Thoma-Emanuel, *Demand Side Management: Potential and impact on the Swiss transmission grid.* PhD thesis, ETH Zurich, 2015.

[192] L. Labib, M. Billah, G. Sultan Mahmud Rana, M. N. Sadat, M. G. Kibria, and M. R. Islam, "Design and implementation of low-cost universal smart energy meter with demand side load management," *IET Generation, Transmission & Distribution*, 2017.

[193] M. Walfish and A. J. Blumberg, "Verifying computations without reexecuting them," *Communications of the ACM*, vol. 58, no. 2, pp. 74–84, 2015.

[194] E. Ben-Sasson, I. Bentov, Y. H. M. S. . . . , and U. 2017, "Scalable, transparent, and post-quantum secure computational integrity," *Eprint.Iacr.Org*, 2018.

Lorenzo Nespoli

https://www.linkedin.com/in/lorenzonespoli/

Email : lorenzo.nespoli@supsi.ch

Mobile : +41 78 828 33 64

EDUCATION

Ecole polytechnique fédérale de Lausanne   Lausanne, CH
PhD Candidate   *2015 – 2019*

Thesis title: *Model based forecasting for demand side management strategies*

Politecnico di Milano   Milano, IT
Master in Energy Engineering (108/110)   *2012 – 2013*

Thesis title: *Analysis of heat and moisture transfer in building components by conjugate modeling*

Politecnico di Milano   Milano, IT
Bachelor in Energy Engineering (105/110)   *2008 – 2011*

EXPERIENCE

Hive Power   Lugano, CH
Co-founder   *Aug 2017 - Present*

- **Market design**: Design of energy market rules for local energy sharing communities, which are electric grid-aware, i.e. induce virtuous trades between users, which mitigate grid congestions.
- **Algorithmics**: Development of control algorithms for the coordination of distributed controllable loads, in a decentralized way.

University of Applied Sciences and Arts of Southern Switzerland (SUPSI)   Lugano, CH
Researcher   *Oct 2018 - Present*

- **Residential load and PV power forecasting**: Development and evaluation of parametric and non-parametric forecasters for the multi step-ahead prediction of power consumption/production
- **Distributed control algorithms**: Development of distributed algorithms for the control of residential loads under uncertainty.

University of Applied Sciences and Arts of Southern Switzerland (SUPSI)   Lugano, CH
Scientific assistant   *Jan 2014 - Oct 2018*

- **Multiphysics systems modeling and simulation**: Modeling and simulation of electro-thermal systems in Modelica and Python. The modeling activities included system identification of buildings thermal dynamics, modeling of heat pumps, boilers and heating system logic.
- **PV and battery sizing**: Optimal sizing of PV and electrical storage systems, sensitivity analysis on economic KPIs of relevant uncertain parameters. Development of algorithms for fast evaluation of investment economic feasibility.

Leopold-Franzens Universitt Innsbruck   Innsbruck, AT
Heat and mass transfer analyst   *Mar 2013 and Aug 2013*

- **CFD analysis**: Computational fluid dynamics analysis of moisture transfer in porous materials for the identification of diffusivity coefficients of insulation materials.

SELECTED PUBLICATIONS

[1] L. Nespoli, A. Giusti, N. Vermes, M. Derboni, A. E. Rizzoli, L. M. Gambardella, and V. Medici, "Distributed demand side management using electric boilers," *Computer Science-Research and Development*, vol. 32, no. 1-2, pp. 35–47, 2017.

[2] L. Nespoli and V. Medici, "An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements," *Solar Energy*, vol. 158, pp. 701–710, 2017.

[3] L. Nespoli and V. Medici, "Constrained hierarchical networked optimization for energy markets," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2018.

[4] L. Nespoli, M. Salani, and V. Medici, "A rational decentralized generalized nash equilibrium seeking for energy markets," in *2018 International Conference on Smart Energy Systems and Technologies (SEST)*, 2018.

[5] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised disaggregation of photovoltaic production from composite power flow measurements of heterogeneous prosumers," *IEEE Transactions on Industrial Informatics*, 2018.