

# Predictive Models for Assessing the Passive Solar and Daylight Potential of Neighborhood Designs: A Comparative Proof-of-Concept Study

Emilie Nault<sup>a,\*</sup>, Peter Moonen<sup>b,c</sup>, Emmanuel Rey<sup>d</sup>, Marilyne Andersen<sup>a</sup>

<sup>a</sup>*Interdisciplinary Laboratory of Performance-Integrated Design (LIPID), School of Architecture, Civil and Environmental Engineering (ENAC),*

*Ecole polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland*

<sup>b</sup>*Univ Pau & Pays Adour, LFCR-IPRA, 64000 Pau, France*

<sup>c</sup>*Univ Pau & Pays Adour, DMEX-IPRA, 64000 Pau, France*

<sup>d</sup>*Laboratory of Architecture and Sustainable Technologies (LAST), School of Architecture, Civil and Environmental Engineering (ENAC),*

*Ecole polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland*

---

## Abstract

Despite recent developments, neighborhood-scale performance assessment at the early-design phase is seldom carried out in practice, notably due to high computational complexity, time requirement, and perceived need for expert knowledge, ultimately limiting the integration of such a task into the design process. In this paper, we introduce a predictive modeling approach to rapidly obtain an estimate of the performance of early-design phase neighborhood projects, from simple geometry- and irradiation-based parameters. The performance criteria considered are the passive solar and daylight potential, respectively quantified by the energy need for space heating and cooling (given certain assumptions) and the spatial daylight autonomy at the ground-floor level. Two predictive models, or metamodels, are developed following distinct techniques: a multiple linear regression function and a Gaussian Processes regression model. These are developed from a reference dataset acquired through the parametric modeling and simulation of neighborhood design variants. When tested on designs provided by professionals, the metamodels with the highest accuracy within the compared types (MLR versus GPs) portray a prediction error below 10% in 87% (respectively 60%) of the cases for the passive solar (resp. daylight) potential. Results show this approach to be a promising alternative to running detailed simulations when comparing early-design variants.

---

\*Corresponding author. Tel.: +41 21 69 38041

Email address: [emilie.nault@epfl.ch](mailto:emilie.nault@epfl.ch) (Emilie Nault)

*Keywords:* predictive model, solar potential, neighborhood scale, early-design phase, simulation

---

## 1. Introduction

With the uprise of evermore demanding energy regulations particularly targeting the built environment sector, a shift towards sustainable urban planning and design is essential. To support this transition, adequate decision-support methods are required to ensure that energy considerations become an integral part of the practitioners' process.

The energy performance of the built environment is strongly conditioned by the morphological features of the buildings, such as their shape and layout, and their interaction with the climate and surrounding context, notably affecting the solar exposure of building surfaces. As such, the need and benefit of assessing building performance at the early-design phase and neighborhood level has well been acknowledged, as these respectively represent the moment and scale at which decisions are made on design parameters of influence (Zeiler et al., 2007; Lechner, 2009). Despite recent efforts to provide support to designers, practical use of design-support tools is still limited, particularly due to excessive computational complexity of urban-scale modeling and simulation, limited integration and guidance during the exploratory early-design phase, as well as insufficient interactivity with the designer (Beckers and Rodriguez, 2009; Hensen and Lamberts, 2011).

As an attempt to overcome these barriers, this paper proposes a neighborhood performance assessment approach based on predictive models (or metamodels) derived from a reference dataset. Our goal is to predict the value of a performance metric (quantifier), from the abstract and limited amount of information typically available at the conceptual neighborhood design stage. Two performance criteria are considered: the passive solar potential, quantified by the energy need for space heating and cooling (given certain assumptions e.g. on insulation levels), and the daylight potential, quantified by the spatial daylight autonomy at the ground-floor level (likely to be a conservative value for the whole design). Resorting to the use of metamodels allows temporarily bypassing the need for detailed simulations, which demand a higher level of both technical and design-related knowledge and of computational resources, but that should nevertheless occur later on in the design process.

## 2. State-of-the-Art

Most existing building performance assessment methods are based on the resolution of equations that simulate the thermal behavior or the (day)light conditions of a building. Such methods lead to high accuracy, so long as the required significant amount of details is available to the user, which is typically not the case at the early-design phase (Zhao and Magoulès, 2012). To overcome this limitation, machine learning methods can be used to bypass the need for full simulations. Machine learning is a multidisciplinary field that employs techniques from analogous domains (e.g. statistics and data mining) to build algorithms that can learn from data with minimal human intervention (Hall et al., 2014). We are here interested in the specific branch of supervised machine learning, which consists of algorithms obtained by training an equation from a set of data. Such equations, known as metamodels, surrogate models or emulators, can then be used for prediction<sup>1</sup>.

When using such methods, a basic requirement is the access to a database, e.g. of buildings, containing the necessary information, e.g. building characteristics and performance results. There are various ways of generating this data. Decisions must be made on the base case building(s), on the parameters to vary and their range. This defines the solution space which will have to be sampled in a certain chosen way, for instance using Monte Carlo. Parameters to be kept constant must also be fixed. The tool(s) used for simulating the different cases and the desired output(s) must be selected. As will be seen later, these decisions define the types of analysis that can be conducted on the results and delimit the validity boundaries of the generated predictive model.

The use of statistical methods and machine learning techniques in the field of building performance is fairly recent. Multiple studies can be found at the building scale; for an exhaustive review, we refer the reader to Fouquier et al. (2013); Zhao and Magoulès (2012). However, fewer studies address larger scales such as a neighborhood Martins et al. (2014) or city (Howard et al., 2012). Table 1 summarizes some of the relevant studies that can be found in the literature. The table contains the type of method used and analysis conducted on the specified inputs and outputs obtained by varying, according to a sampling method, a base case design.

---

<sup>1</sup>The term metamodel is used in the following, sometimes replaced by (predictive) function or simply model to alleviate the text, the latter not to be confused with a 3D model (the sentence's context should prevent any mix-up)

The main techniques used in the field of building energy forecasting are multiple linear regression, artificial neural networks (ANNs), genetic algorithms (GAs) and support vector machines (SVMs) (Foucquier et al., 2013). Linear regression and more recently ANNs are often used to predict energy consumption from various sets of parameters (inputs) related to the building material, orientation, glazing area and others (Asadi et al., 2014; Hygh et al., 2012; Tsanas and Xifara, 2012; Ekici and Aksoy, 2009; Ritter et al., 2015). More complex techniques such as support vector regression (SVR) can also be found being used for prediction (Eisenhower et al., 2012). The application of Bayesian calibration based on Gaussian Processes (GPs) in the field of building performance prediction was first seen in (Heo et al., 2012). In contrast to ANNs and SVMs, GPs provide full probabilistic prediction and an estimate of uncertainty in the prediction. For an in depth comparison, we refer the interested reader to (Rasmussen and Williams, 2006). While advanced methods such as ANNs, SVMs, SVR and GPs are more powerful than simpler techniques like multiple linear regression, they are also more complex to implement, use and interpret (Foucquier et al., 2013; Zhao and Magoulès, 2012).

This paper proposes a neighborhood performance assessment approach based on meta-models. According to the authors' knowledge, very little similar research has been done at this scale. A notable example is the work of Martins et al. (2014), who conducted a sensitivity analysis to examine the influence of various urban morphological parameters (e.g. floor area ratio, plot ratio, average building height) on the irradiation and illuminance levels on the envelopes of nine surrounded buildings. While they followed-up with an optimization of the built form using GAs, we propose to take a different approach, combining urban morphological parameters and irradiation levels as inputs to a metamodel allowing to predict the desired performance output. For both the energy need and spatial daylight autonomy metric, a multiple linear regression model is trained and compared to the more powerful Gaussian Processes model. This twofold metamodeling approach resembles the work of Tsanas and Xifara (2012), who compared linear regression with a non-linear/non-parametric method (random forests) for predicting the heating and cooling load of a residential building.

The next sections describe the methodology for development and calibration of the predictive models, followed by a presentation of the obtained results and a comparative assessment of the prediction accuracy of either technique.

Table 1: Overview of studies using machine learning techniques in the field of building performance prediction and sensitivity analysis. \*Sample size of complete database including training and testing sets. (?): not clearly stated; bldg: building; incl.: including; MLR: multiple linear regression; RSM: response surface methodology; NYC: New York City; GPs: Gaussian Processes; LHS: Latin Hypercube Sampling.

Reference	Model/analysis type	Base case	Inputs	Outputs	Sampling method; Sample size*
Asadi et al. (2014)	MLR (per building shape)	two-story office bldg	materials (incl. thickness), schedule	energy consumption (heating, cooling, lighting, ventilation)	Monte Carlo; 70 000
Hygh et al. (2012)	multivariate regression (per climate)	office bldg	27 parameters (incl. depth, orientation, U-value)	energy consumption (heating, cooling, total)	Monte Carlo; 20 000
Karteris et al. (2013)	linear regression (per roof class)	multi-story-family bldg	unavailable roof areas, orientation of roof near south side, stairwell and elevator shaft position	solar suitable flat roof-top area	209 representative real bldgs
Lam et al. (2010)	multiple regression (per climate)	40-story office bldg	12 bldg design parameters (incl. window-to-wall ratio, U-value)	annual bldg energy use	full factorial, pseudo-random number generator; 1021
Ritter et al. (2015)	MLR (2nd degree polynomial); least squares; RSM	office (?) bldg	n user-defined bldg parameters (e.g. glazing factor, window U-value)	user-defined energy performance metric (?) (e.g. heating energy use)	3 level full factorial ( $3^n$ ); according to user-inputs
Howard et al. (2012)	MLR (per city area and bldg function)	NYC urban areas incl. 8 bldg functions	bldg floor area and function	annual energy consumption (electricity and total fuel)	NYC database
Heo et al. (2012)	Bayesian calibration using GPs	office bldg	4 parameters defining ventilation, infiltration, and indoor temperature	gas consumption	Morris and LHS; 150

Table 1: (Continued) Overview of studies using machine learning techniques in the field of building performance for prediction and sensitivity analyzes. \*Sample size of complete database including training and testing sets. (?): not clearly stated; bldg: building; incl.: including; FAST: Fourier Amplitude Sensitivity Test; MLR: multiple linear regression; LHS: Latin Hypercube Sampling; ANN: Artificial Neural Networks; ANFIS: Adaptive Neuro-Fuzzy Inference System; SVR: Support Vector Regression; PMV: Predicted Mean Vote; DoE: Design of Experiment; GSA: Global Sensitivity Analysis; GA: Genetic Algorithm.

Reference	Model/analysis type	Base case	Inputs	Outputs	Sampling method; Sample size*
Capozzoli et al. (2009)	variance decomposition with FAST (per climate); MLR	middle floor of multi-story office bldg	6 parameters (incl. compactness, orientation)	energy needs for space heating and cooling	LHS and Monte Carlo; 600 (?)
Tsanas and Xifara (2012)	linear regression and random forests	residential bldg	8 parameters (incl. surface area, orientation, glazing area)	heating and cooling load	full factorial; 768
Ekici and Aksoy (2009)	ANN (per building aspect ratio)	bldg	transparency ratio, insulation thickness, orientation	heating energy	full factorial; 896
Ekici and Aksoy (2011)	ANFIS	bldg	aspect ratio, transparency ratio, insulation thickness, orientation	heating and cooling energy	full factorial; 225
Eisenhower et al. (2012)	SVR with Gaussian kernel	two-story office and gym bldg	1009 parameters initially (incl. material, equipment), then 20 most influential	thermal comfort (PMV) and annual energy consumption	quasi-Monte Carlo; 5000
Martins et al. (2014)	sensitivity analysis; significance test	real high-rise urban area (9 bldgs)	12 parameters (incl. height, street width, floor area ratio)	irradiation and illuminance on bldg envelope	DoE fractional factorial; 8192 (?)
Hemsath and Alagheband Bandhosseini (2015)	local sensitivity index and Morris GSA (per climate)	residential bldg	9 parameters (incl. geometric e.g. aspect ratio and material e.g. insulation)	heating and cooling loads	GA (?); not stated

### 3. Methodology

#### 3.1. Main goal and steps

The main purpose of this study is to investigate the feasibility of generating predictive models for rapidly and simply assessing the solar performance of neighborhood-scale designs in their early-phase. We define solar performance in terms of three criteria, of which the first two are addressed in this paper: (i) the passive solar potential, quantified by the energy need for space heating and cooling, (ii) the daylight potential, quantified by an adapted and conservative ground-floor level spatial daylight autonomy, and (iii) the active solar potential, quantified by the estimated energy production by photovoltaic and solar thermal systems. Whereas the evaluation of the first two metrics typically requires a detailed model of all buildings followed by a time-consuming thermal/daylighting simulation, the latter metric can be estimated from a faster irradiation simulation on a basic building massing model. For this reason, no metamodel was developed for the active solar potential criterion. However, it forms part of our solar potential definition and is considered in the larger research framework in which this study falls (Nault et al., 2016a; Nault, 2016).

This paper also aims at comparing two distinct metamodeling techniques - multiple linear regression (MLR) and Gaussian Processes (GPs) - which were chosen because they cover a big portion of the spectrum of possible methods, ranging from simple to complex in terms of implementation and interpretation.

The subsequent sections discuss how predictive metamodels can be constructed, starting from the generation of a reference dataset, over the validation of its consistency, up to metamodel training and testing. These three main phases are illustrated in Fig. 1.

#### 3.2. Data collection

Data collection consists of three steps. First, variants of neighborhood designs are parametrically generated. These are subsequently used to extract descriptors, characterizing the neighborhood. As a final step, the performance of each design is simulated. Details on each step are given below.

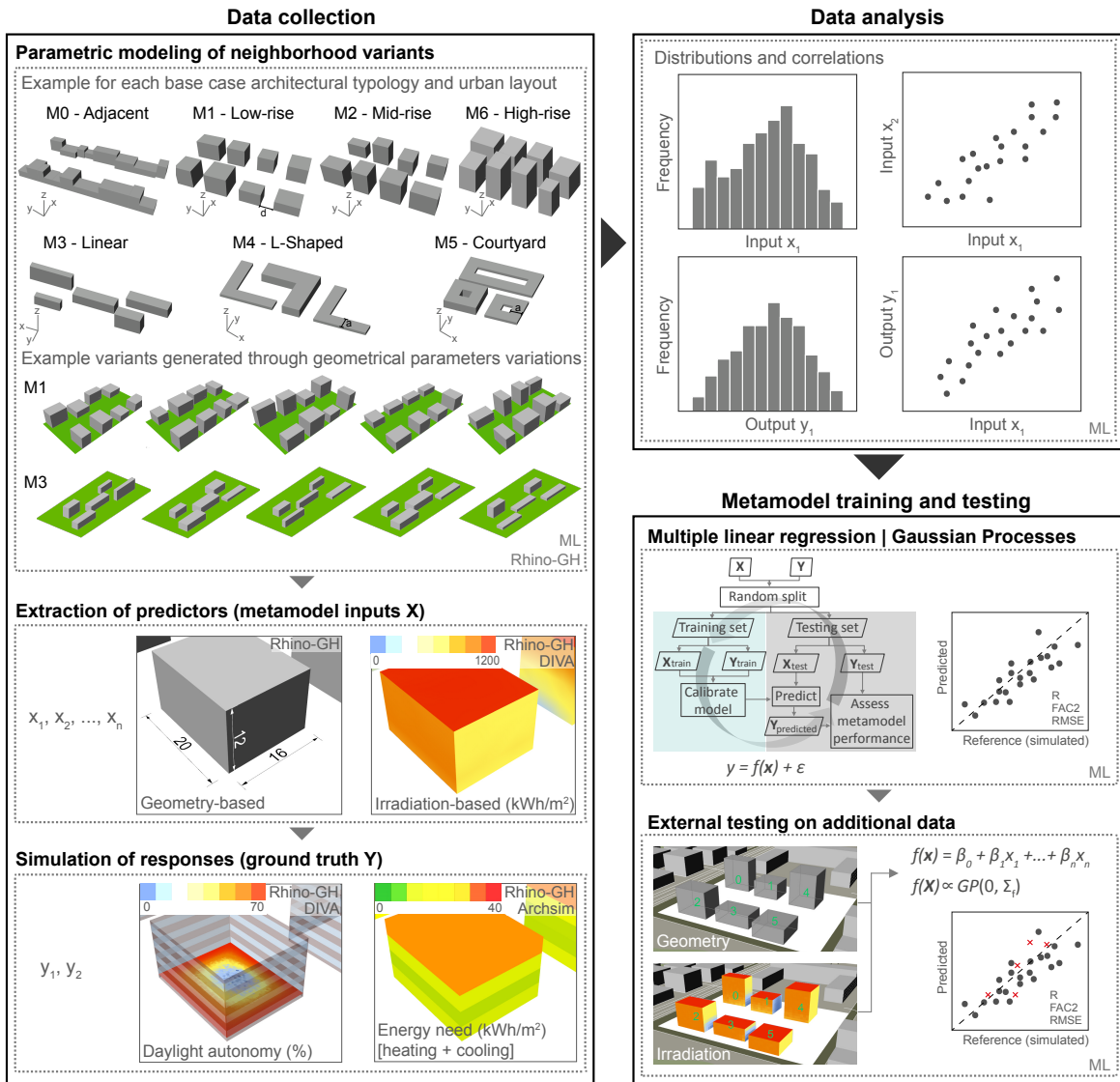


Figure 1: Overview of the steps in our approach. Tools used in each part are annotated (references can be found in the text): ML: Matlab; Rhino: Rhinoceros; GH: Grasshopper; DIVA: DIVA-for-Rhino/GH. Best viewed in color.

### 3.2.1. Parametric modeling of neighborhood variants

A parametric modeling workflow was set up in Grasshopper<sup>2</sup>, an algorithmic modeling platform for the 3D modeling interface Rhinoceros<sup>3</sup> (v5), to generate a series of variants starting from seven base case neighborhood designs. Each base case consists of a replicated building typology according to a certain urban layout. An example variant for each case is

<sup>2</sup>[www.grasshopper3d.com/](http://www.grasshopper3d.com/) (last accessed on March 20, 2016)

<sup>3</sup>[www.rhino3d.com/](http://www.rhino3d.com/) (last accessed on March 20, 2016)



illustrated in the upper left corner of Fig. 1. The base case designs come from two earlier studies: M0 to M2 stem from a collaboration with an architecture and urban design firm (Urbaplan), while M3 to M5 were generated as a preliminary dataset and inspired by student projects analyzed in the context of a collaborative study (Rey, 2013; Nault et al., 2013). M6 was later added to cover the tower typology, with the same layout as M0-M2. In what follows, we will refer to each base case as *case M $\alpha$*  ( $\alpha = [0, 6]$ ), as a term inclusive of both the building typology and the urban layout.

To generate a series of variants from the base cases, specific design variables had to be varied. To identify them, the following requirements were considered:

1. The design variables should represent parameters with which designers normally play at the early design phase of neighborhood-scale projects,
2. have a significant impact on the solar access / energy performance of buildings,
3. be reasonably easy to vary parametrically for a set of buildings.

Relevant design parameters which did not conform to these criteria were considered either as fixed or as constraints to respect.

Parameters related to point 1 were extracted from masterplan documents (PDL Gare-Lac, 2010; Gauthier, R., Atelier Poisson and SDOL, 2012), and through collaboration with the aforementioned firm. To identify the parameters most likely to affect building performance, we looked at multiple studies on the subject (Cheng et al., 2006; Sok Ling et al., 2007; Lobaccaro et al., 2012; van Esch et al., 2012; Takebayashi and Moriyama, 2012; Hachem et al., 2012; Pessenlehner and Mahdavi, 2003; Zhang et al., 2012; Martins et al., 2014; Peronato, 2014). The main parameters respecting points 1 and 2 were identified as the street width, the buildings shape, height and orientation, and the density.

When faced with the third requirement, a decision was made to generate variants by altering variables at the building scale (such as building height), as opposed to variables at the neighborhood scale (such as density). Neighborhood-scale descriptors, used in the metamodel development, can a posteriori easily be derived from the generated variants.

As such, the height, depth and width of individual buildings were taken as variables for the 3D modeling, as well as the grid orientation. A series of design variants were parametrically generated from each base case by modifying the building dimensions within the ranges listed in Table 2. All variants were then duplicated with a 90° grid rotation with respect to their initial alignment. Example variants for M1 and M3 are shown in Fig. 1.

Table 2: Range for each variable (z, y, x, a) and constraints (d, PR, FP) applied when generating the design variants, with respect to the initial grid orientation coded as 0° (see Fig. 1, also for definition of x, y, z, d and a). Dist.: distance; btw: between; bldg: building; PR: plot ratio; FP: footprint.

Case	Height (z) [m]	Length (y) [m]	Width (x) [m]	Depth (a) [m]	Dist. btw bldgs (d) [m]	Min PR [-]	Min bldg FP [m]
M0	6-12	6-24	8-15	-	0	0.9	50
M1	9-18	12-24	10-20	-	6-20	0.9	200
M2	12-24	12-24	10-20	-	6-20	0.9	200
M3	3-24	6-22	fixed	-	-	-	-
M4	3-24	-	-	6-22	-	-	-
M5	3-24	-	-	6-22	-	-	-
M6	30-60	15-29	15-29	-	6-20	-	-

For typologies M0-M2 and M6, the dimension values were selected by a random sampling algorithm scripted in Matlab<sup>4</sup>, which included a constraints verification to ensure all generated variants respected the specified minimum plot ratio and building footprint. For typologies M3 to M5, we followed the Design of Experiment (DoE) approach and used a 3-level Box-Behken design to generate the variants<sup>5</sup>. To increase the amount of data, a second Box-Behken design with a shifted range was used. It must be noted that, although both the random sampling and DoE-based approaches are valid experimental methods (Simpson et al., 2001), the added-value of the latter - i.e. that it provides the shape of the response surface and allows estimating interaction and quadratic effects<sup>6</sup> - is not exploited here, since the factors that are varied are not the ones that are later derived and used as metamodel inputs. More details on the sampling of the variables can be found in Nault et al. (2016a); Nault (2016).

### 3.2.2. Extraction of potential predictors (metamodel inputs X)

The design variations prompted by the parametric modeling procedure affect multiple other parameters at the neighborhood scale such as the street width, density, and solar exposure levels of the buildings. To capture these changes, an extensive amount of geometry- and irradiation-based parameters were computed at the neighborhood scale and used to populate the inputs dataset for the metamodel. Geometry-based parameters include density

<sup>4</sup>[mathworks.com/products/matlab/](http://mathworks.com/products/matlab/) (last accessed on March 20, 2016)

<sup>5</sup><http://www.itl.nist.gov/div898/handbook/pri/section3/pri3362.htm> (last accessed on August 15, 2016)

<sup>6</sup><http://www.itl.nist.gov/div898/handbook/pri/section3/pri33.htm> (last accessed on August 15, 2016)

measures (plot ratio: total building floor area divided by land area; site coverage: building footprint area divided by land area), facade ratios per orientation, and other values which were obtained through a Grasshopper script. Irradiation-based parameters include averages (per envelope area and per floor area) for roofs, facades, whole envelope and facades per orientation. The raw irradiation data was obtained through annual simulation using DIVA-for-Grasshopper (Jakubiec and Reinhart, 2011), with a 1-meter sensor grid resolution, 2 ambient bounces, and using the EnergyPlus weather data file for Geneva, Switzerland<sup>7</sup>. The full list of 25 parameters is graphically shown in the results section (§4.1). Our goal is to investigate the potential relationship between these design-specific values and the passive solar and daylight performance of buildings, ultimately exploiting these links to obtain a prediction of the performance from the easily computable parameters (or predictors).

### 3.2.3. Simulation of responses (ground truth $Y$ )

The performance of each neighborhood variant was then assessed. In this study we focus on the passive solar potential, quantified by the floor area-normalized energy need for heating and cooling, and the daylight potential, quantified by the spatial daylight autonomy (sDA) (IESNA, 2012). These performance metrics are obtained through full climate-based simulations run in EnergyPlus (Crawley et al., 2000) and Radiance (Larson and Shakespeare, 2011)/Daysim<sup>8</sup>, using the EnergyPlus weather file for Geneva. For each performance criterion, a metamodel is developed whose target is to quickly assess the corresponding metric solely using some of the inputs  $X$  described in the preceding section.

#### *Passive solar potential*

The energy need for heating and cooling, used to quantify the passive solar potential, is an indicator of how far we are from comfortable thermal conditions mainly due to solar heat gains and sun shading (caused by surrounding buildings), when looking at different building shapes with certain fixed assumptions e.g. on occupancy. The selection of this performance metric is further justified in previous work (Nault et al., 2015).

The initial EnergyPlus (Crawley et al., 2000) Input Data Files (IDF) were generated via the Grasshopper plug-in Archsim<sup>9</sup>. They were then processed in Matlab before being simulated with EnergyPlus to obtain the annual energy need for heating and cooling, assuming

---

<sup>7</sup><https://energyplus.net/weather> (last accessed on March 4, 2016)

<sup>8</sup><http://daysim.ning.com/> (last accessed on March 20, 2016)

<sup>9</sup>[www.archsim.com/](http://www.archsim.com/) (last accessed on March 20, 2016)

an ideal loads system. The values were summed over all buildings and normalized by the total floor area for each design variant. Each floor (3m high) was defined as a thermal zone, and windows were modeled equally on each facade between a narrow range of window-to-wall ratio (44% and 48%). Additional energy simulation settings are listed in Table 3.

#### *Daylight potential*

The spatial daylight autonomy, as defined by the IESNA (2012), describes the annual sufficiency of daylight levels inside a building. The  $sDA_{300/50\%}$  is used here, corresponding to the percent space that exceeds a target illuminance of 300 lux over 50% of the occupied hours over the year. Our evaluation of the sDA however differs slightly from the recommendations of IESNA (2012), in particular the occupancy period which we have set to 8am-10pm (instead of 8am-6pm), so as to cover most daylight hours as opposed to representing a clear building function (e.g. office). This allows comparing between design variants without assuming any fixed function. The simulations were conducted using DIVA-for-Grasshopper with the settings listed in Table 4.

#### *3.3. Data analysis*

The above-mentioned steps of design generation, descriptor extraction and performance simulation yield a fully populated table with inputs and targeted outputs for the metamodels. Prior to training the metamodels, the data was examined for potential anomalies or outliers. Furthermore, an analysis was conducted to verify the correlation level between inputs and outputs, as well as among inputs. These checks respectively help identify the relevant and redundant inputs, information that can be useful for analyzing the regression results.

#### *3.4. Metamodel training and testing*

The procedures for training metamodels using two distinct methods - multiple linear regression (MLR) and Gaussian Processes (GPs) - are detailed below. The training process is iterative and involves specific model validity checks ('internal testing'). The final metamodels are subsequently confronted to 'external' test data in order to investigate their prediction accuracy and generalizability.

##### *3.4.1. Multiple linear regression (MLR)*

From the initial set of geometry- and irradiation-based parameters, trials were first conducted to identify the ones with the most statistical significance in explaining each output

Table 3: Settings for the energy need simulations. SHGC: solar heat gain coefficient.

Building function	Office
Heating/cooling set point	20/26 °C
Loads	
Equipment	12 W/m <sup>2</sup>
Lighting	3 W/m <sup>2</sup>
Occupancy	0.05 people/m <sup>2</sup>
Ventilation	0.0125 m <sup>3</sup> /s person
Infiltration	0.1 ach
Surface conductivity	
Wall	U = 1.3 W/m <sup>2</sup> K
Roof	U = 1.84 W/m <sup>2</sup> K
Ground floor	U = 1.25 W/m <sup>2</sup> K
Windows (double low e argon)	U = 1.5 W/m <sup>2</sup> K, SHGC = 0.597
Solar protections	
Type	Venetian blinds
Activated when	Incident irradiation $\geq$ 180 W/m <sup>2</sup>

Table 4: Settings for the daylight simulations. SHGC: solar heat gain coefficient.

Windows	
Type	Double-argon
SHGC	0.27
Visual transmittance	65%
Occupancy	8am - 10pm
Illuminance target	300 lux
Sensor grid resolution	2 m
Sensor grid height above ground floor	0.8 m
Number of ambient bounces	2
Simulated floor	Ground floor of all buildings
Shades	None

using the stepwise fitting algorithm<sup>10</sup>. This method proceeds by adding and removing terms in an iterative fitting process, looking at the sum of squared errors (SSE) as a performance criterion. Applying this algorithm allow us to more confidently reduce the size of the inputs dataset by retaining only the most influential ones. Despite this reduction, a risk of collinearity between the stepwise-selected inputs remains and is addressed by applying ridge regression<sup>11</sup> within the algorithm described below and through which the MLR-metamodels were individually trained and tested (see also sequence illustrated in Fig. 1).

<sup>10</sup>[www.mathworks.com/help/stats/stepwiselm.html](http://www.mathworks.com/help/stats/stepwiselm.html) (last accessed on October 20, 2016)

<sup>11</sup>[www.mathworks.com/help/stats/ridge.html](http://www.mathworks.com/help/stats/ridge.html) (last accessed on October 20, 2016)

The data acquired is equally split into a training and testing set. Multiple linear regression is applied on the training data to estimate a metamodel of the form:

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon = \beta_0 + \sum_{i=1}^P \beta_i x_i + \varepsilon \quad (1)$$

where  $y(\mathbf{x})$  is a response (ground truth),  $f(\mathbf{x})$  the model output,  $\varepsilon$  the approximation and measurement error,  $\mathbf{x} \in \mathbb{R}^P$  are the inputs values with  $P$  the number of inputs, and  $\beta$  the unknown coefficients. To estimate the model coefficients, we minimize:

$$\sum_{j=1}^N (f(\mathbf{x}_j) - y_j)^2 + \lambda \sum_{i=1}^P \beta_i^2 \quad (2)$$

where  $N$  is the number of training samples and  $\lambda$  a regularization parameter which penalizes large coefficient values likely to appear if there are many correlated inputs, which is particularly relevant to our problem. Indeed, since we start with an extensive list of geometry- and irradiation-based parameters as potential inputs (see Fig. 3), and whose values are not directly controlled but rather computed within the modeling and simulation process, we may end up with correlated series of inputs. In such cases, a large positive coefficient on one input can be canceled by an equally large negative coefficient on a correlated input (given the same order of magnitude of values in the two sets) (Hastie et al., 2009). The  $\lambda$  penalization parameter is a feature of ridge regression, which falls in the category of shrinkage methods (Hastie et al., 2009).

The  $\beta$  minimizing equation 2 can be obtained as:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (3)$$

where  $\mathbf{X} \in \mathbb{R}^{N \times P}$ . Starting with 1000 values of  $\lambda$  between 0 and 1 (interval defined following previous trials), we use a technique called k-fold cross-validation to find the best  $\lambda$  (Hastie et al., 2009). We split the data in  $k(=10)$  sets, take  $k-1$  sets for training, then test the fitted model on the remaining set. The best  $\lambda$  corresponds to the value yielding a minimum root mean square error (RMSE) averaged over the  $k$ -fold iterations, with RMSE:

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (y_j - f(\mathbf{x}_j))^2}{N}} \quad (4)$$

Once we have fitted a model over the entire training set using  $\lambda_{\text{best}}$ , two-fold cross-validation is used to estimate the prediction error, which represents the performance of the model over unseen data (testing set) (Hastie et al., 2009). This performance is quantified by the RMSE, along with the fraction of predictions falling within a factor two of the reference values (FAC2):

$$0.5 \leq \frac{f(\mathbf{x}_j)}{y_j} \leq 2 \quad (5)$$

as well as the coefficient of correlation<sup>12</sup>  $R$ :

$$R = \frac{\text{cov}(f(\mathbf{x})y)}{\sigma_{f(\mathbf{x})}\sigma_y} = \frac{1}{N-1} \sum_{j=1}^N \left( \frac{f(\mathbf{x}_j) - \mu_{f(\mathbf{x})}}{\sigma_{f(\mathbf{x})}} \right) \left( \frac{y_j - \mu_y}{\sigma_y} \right) \quad (6)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation.

This process is repeated for 100 splitting iterations, in order to verify the potential of our proposed approach and function type (MLR) by looking at the prediction accuracy for different training and testing subsets of our reference dataset.

The last step consists in generating the final function that will represent the predictive model to be subsequently used for estimating the performance of new neighborhood designs. To do so, ridge regression is applied on the entire dataset, with a two-fold cross-validation phase for defining the best ridge parameter  $\lambda_{\text{best}}$ .

It is to note that during both the training-testing and final fitting phases, the input data is centered and scaled to have a mean of 0 and a standard deviation of 1, when applying ridge regression through Matlab’s ridge function. The coefficients defined are then restored to the scale of the initial data.

### 3.4.2. Gaussian Processes (GPs)

The preceding section described the steps for building a multiple linear regression meta-model. Herein, the form of the functional relationship between inputs and outputs was by construction “multi-linear”. Gaussian Processes take a different approach in that they assume that the functional relationship between inputs and outputs is unknown. GPs essentially define a (Gaussian) probability distribution over all possible functions. Hereby those functions which are more likely to describe the data will receive a higher probability.

There are several ways to construct a metamodel using Gaussian Processes. We follow the

---

<sup>12</sup>[www.mathworks.com/help/matlab/ref/corrcoef.html](http://www.mathworks.com/help/matlab/ref/corrcoef.html) (last accessed on August 15, 2016)

mathematical formulation described in Moonen and Allegrini (2015), as the predecessor of this approach has shown to yield promising results in the field of building energy simulation (Evins et al., 2014). In this approach, the relation between the simulated responses  $y(\mathbf{x})$  and the output of the metamodel  $f(\mathbf{x})$  can be expressed as:

$$\mathbf{Y}(\mathbf{X}) = f(\mathbf{X}) + \varepsilon \quad (7)$$

where  $\mathbf{X} \in \mathbb{R}^P$  are the normalized input values (i.e. the input values  $\mathbf{x}$  linearly mapped onto the interval  $[0 \ 1]$ ) with  $P$  the number of inputs,  $\mathbf{Y}$  are the standardized output values (i.e. the output values  $\mathbf{y}$  linearly mapped as such that they have zero mean and unit variance) and  $\varepsilon$  is an error term accounting amongst others for the imperfection of the metamodel. The scaling, under the form of standardization and normalization, is applied to rule out bias (e.g. because inputs and outputs concern quantities given in different units and orders of magnitude) and to facilitate metamodel construction. Formally, a similarity between equations 7 and 1 can be noted. Apart from the applied scaling, the main difference is that both right hand side terms in equation 7 are modeled in a stochastic way.

The response of the metamodel  $f(\mathbf{X})$  is modeled as a zero-mean Gaussian process. This means that for every given set of inputs, the possible responses are assumed to be normally distributed.

$$f(\mathbf{X}) \propto GP(\mathbf{0}, \boldsymbol{\Sigma}_f) \quad (8)$$

where  $\boldsymbol{\Sigma}_f$  is a covariance matrix. We employ the classical squared exponential covariance function (see e.g. Ch. 4 of Rasmussen and Williams (2006)), which expresses that two configurations with similar inputs should yield a similar performance. The strength of this correlation is expressed by  $P$  unknown hyperparameters, one hyperparameter per input parameter, and supplemented by one additional hyperparameter expressing the possible variance in the performance of different configurations characterized by identical inputs.

The error  $\varepsilon$  is modeled as a zero-mean Gaussian distribution with a yet unknown variance, expressed by the hyperparameter  $\lambda_e$ :

$$\varepsilon \propto N(\mathbf{0}, \lambda_e^{-1} \mathbf{I}) \quad (9)$$

where  $\mathbf{I} \in \mathbb{R}^{P \times P}$  is a unit matrix. The proposed covariance structure reflects that random errors are uncorrelated.



A Bayesian approach is employed to determine the most ‘likely’ values for the  $P + 2$  unknown hyperparameters. This is called metamodel calibration. The term ‘likely’ is to be understood in a stochastic sense: the metamodel provides an output for each combination of values of the hyperparameters, yet some of these outputs are closer to the ground truth than others. Consequently, some combinations of values are more likely than others. The ‘initial guess’ for each hyperparameter is expressed by a probability distribution, the so-called prior distribution. The latter is iteratively updated in the Bayesian framework to obtain the so-called posterior distribution. We employed `fminbnd`, a standard Matlab minimization algorithm based on golden section search and parabolic interpolation, to determine that combination of hyperparameters for which the posterior probability is maximal.

Once the hyperparameters determined, predictions for unknown input parameters  $\mathbf{x}$  can be obtained post-hoc as drawn from

$$N(\mathbf{V}\lambda_e\mathbf{y}, \mathbf{V}) \tag{10}$$

where  $\mathbf{V}^{-1} = \lambda_e\mathbf{I} + \boldsymbol{\Sigma}_f^{-1}$ .

As for MLR, the quality of predictions can be evaluated by calculating error metrics such as RMSE, FAC2 and R. To ensure the validity of the resulting metamodel, different validation efforts were conducted. As for MLR, data were randomly split into two equal groups, where one group was used to train the model, and one for its validation. This was repeated for 100 different splits and it was verified that the resulting hyperparameter-distributions had a low variance. A second type of validation relates to the iterative procedure to find the most likely hyperparameters. Different prior-distributions were tried and it was verified that the algorithm converged to similar posterior hyperparameters.

For a stepwise treatment of the construction of a metamodel based on GPs and its validation, the interested reader is referred to Moonen and Allegrini (2015). For a fundamental treatment of the subject, additionally accounting for systematic bias between inputs and outputs, we refer to Higdon et al. (2008) and Kennedy and O’Hagan (2001).

#### *3.4.3. External testing on additional data*

The previous sections demonstrated how metamodels can be constructed based on the initial dataset containing variations of the base case designs described in §3.2.1. To deepen the validation test of the metamodels, they were applied on a newly acquired set of data

ensuing from workshops conducted with practitioners and further detailed in Nault et al. (2016b); Nault (2016).

Design variants for the area identified in Fig. 2 (top) were generated in preparation for the event while others, designed by the participants, were collected during the workshops, for a total of 30 cases. Examples are shown in Fig. 2 (bottom). Designs differed in the number, typology (shape) and dimensions of buildings.

The required geometry- and irradiation-based inputs were computed and fed to the meta-models to obtain the predictions. To measure the predictive performance of the metamodels, simulations were conducted with the same settings (§3.2.3) and error metrics RMSE, FAC2 and R were calculated. Residuals, computed as the difference between simulated and predicted values, were also analyzed.



Figure 2: Top: context and area (framed) on which workshop participants had to create a design (adapted from the Vernand materplan in Ville de Lausanne (2013)). Bottom: examples of designs generated by the participants.

## 4. Results & Discussion

In this section, we look at the two performance criteria, to which we will refer as the ‘energy’ and ‘daylight’ metrics. The data analysis and metamodel development are done for both metrics separately following the methodology explained in §3.3 and 3.4, respectively.

### 4.1. Data analysis

The extensive list of geometry- and irradiation-based parameter values recorded for each design variant is listed in Fig. 3, which quantifies their linear correlation degree, measured by Pearson’s coefficient<sup>13</sup>, for the energy dataset. Light and dark squares represent a perfect positive and negative correlation respectively. A similar image was obtained for the daylight dataset, not included here for that reason.

We observe a significant degree of collinearity among the input parameters. At least part of these correlation results stems from the parametric modeling, i.e. the way variants were generated and the fact that they were duplicated with a 90° rotation. As such, we observe some correlation among the North-East-South-West parameters and between the plot ratio, roof ratio and mean height. Both metamodeling approaches handle collinearity in a different way. As explained in §3.4.1, MLR reduces the dimensionality of the input dataset by a stepwise algorithm, while GPs automatically ignore redundant inputs (see §4.2.2).

Figure 4 contains scatter plots showing the relationship between the simulated energy need and the values for each input parameter, color- and marker-coded for each case series  $M\alpha$ . The same plots for the spatial daylight autonomy are shown in Fig. 5. The Pearson’s correlation coefficient is shown in the top right corner of each scatter plot.

As can be seen, the energy and daylight datasets are composed of different amounts of data; respectively 508 and 560 points each corresponding to a design variant within the  $M\alpha$  series. This is due to factors linked to available computational resources and modeling errors that required removing some cases, particularly in the energy dataset (e.g. M0 and M2 completely removed).

In both figures, we observe clusters defined not only by the case, but by the two orientation configurations, e.g. in plots of the last row. For the energy metric (Fig. 4), the clearest patterns and highest correlation values are observed for the *FormFactor* (-0.90, decreasing energy need with higher compactness) and *EnvelopeIrradPerFA* (0.88, increasing energy need

---

<sup>13</sup>[www.mathworks.com/help/stats/corr.html](http://www.mathworks.com/help/stats/corr.html) (last accessed on August 15, 2016)

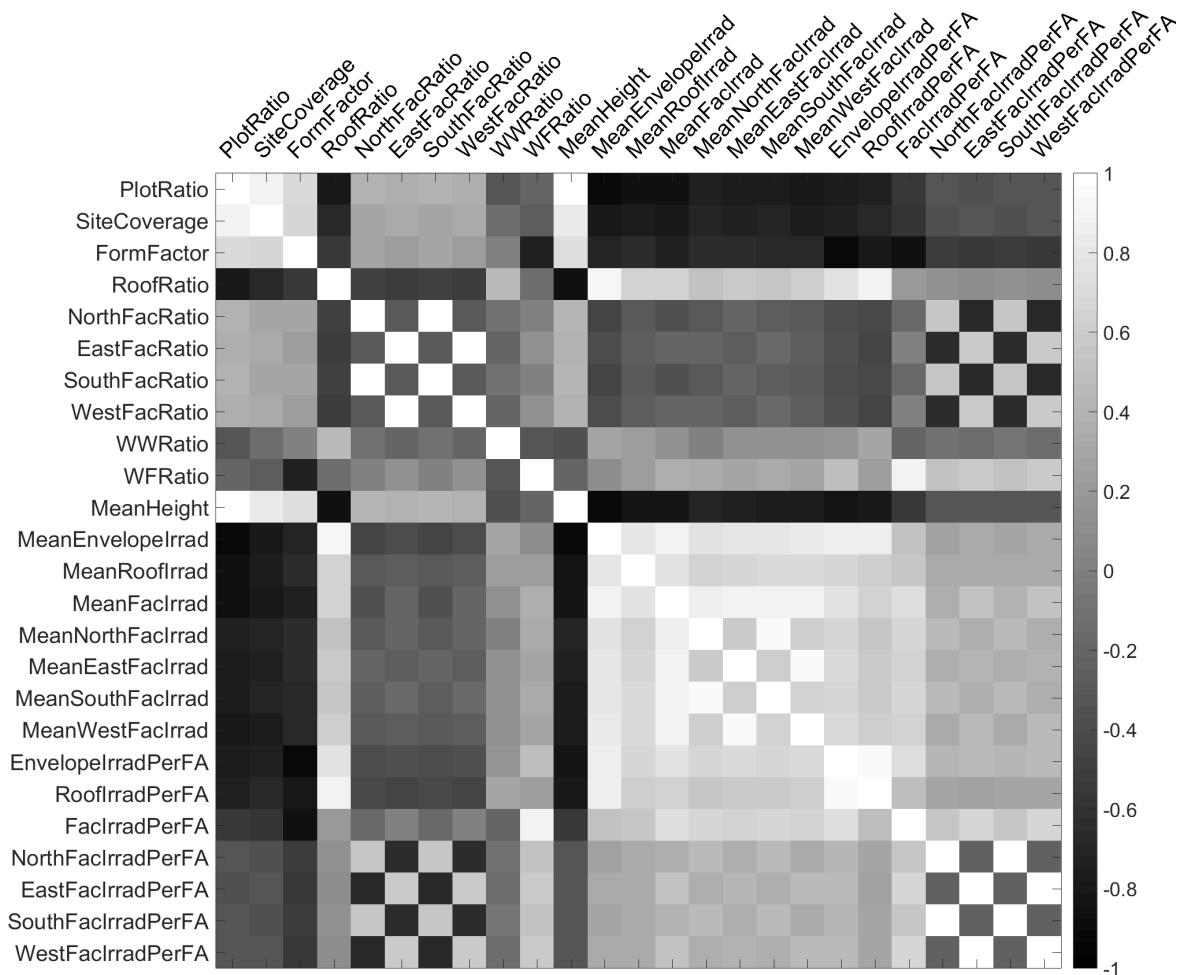


Figure 3: Correlation degree between each pair of potential inputs from the energy dataset. A value of 1 (white) indicates a perfect positive correlation, while a value of -1 (black) represents a perfect negative correlation. Fac: facade; WW/F: window-to-wall/floor; Irrad: irradiation; FA: floor area. Similar results were found for the daylight dataset.

with more solar exposure per floor area). While the latter trend may appear surprising, it is linked to the fact that higher envelope irradiation levels, either averaged or normalized by the floor area, are negatively correlated to a certain degree to the compactness (*FormFactor*), as seen in Fig. 3. More compact designs have less exposed surface areas, yet they are associated to a better performance as the heat exchanges linked to the shape are more favorable than the solar exposure contributions for the overall yearly energy need for heating and cooling for the climate of Geneva.

Similar trends are observed for the daylight dataset (Fig. 5). The strongest correlations are found for the *EnvelopeIrradPerFA* (0.81) and *FacIrradPerFA* (0.83), both positive; while

the irradiation levels increase, so does the daylight metric, as one would expect. Similarly for the *WFRatio*, which however shows a weaker correlation degree (0.72).

It is important to note that these results, as well as the inputs correlation levels found in Fig. 3, are dependent on the way the different design variants were generated and the ranges of each variable. For instance, the window-to-wall ratio was set to be almost constant during the parametric modeling process. Because of its narrow range (*WWRatio* between 44% and 48%), we don't see a trend in both Fig. 4 and 5, which we would however expect to appear if the range was extended. This particular parameter, which is the only one in the list that was so constrained, was excluded in the subsequent phases of fitting the metamodels.

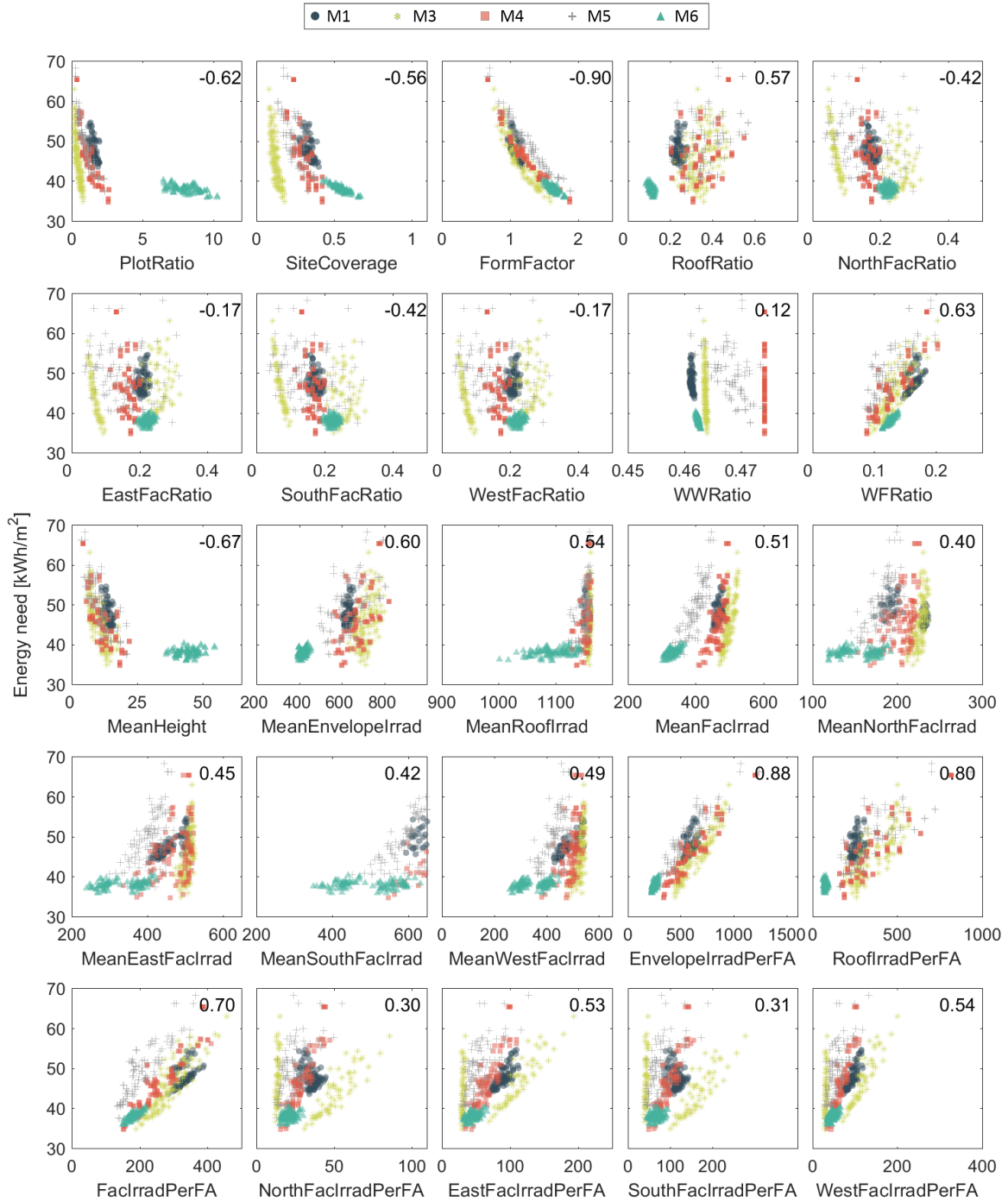


Figure 4: Simulated energy need ( $\text{kWh/m}^2$ ) versus each input parameter value for the 508 design variants in the  $M\alpha$  series, with the associated Pearson's linear correlation coefficient (corner of each graph). Fac: facade; WW/F: window-to-wall/floor; Irrad: irradiation; FA: floor area. Best viewed in color.

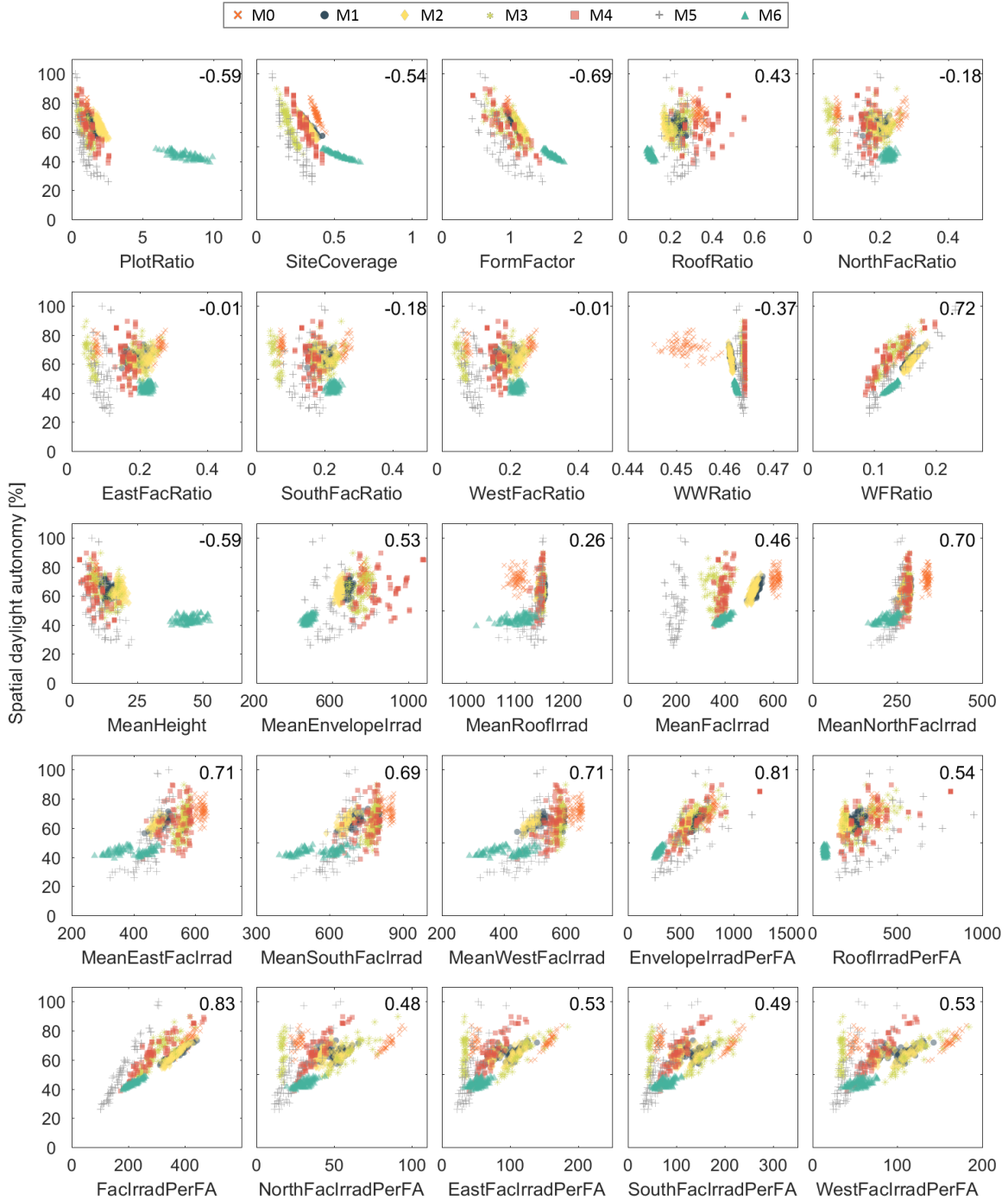


Figure 5: Simulated spatial daylight autonomy (%) versus each input parameter value for the 560 design variants in the  $M\alpha$  series, with the associated Pearson's linear correlation coefficient (corner of each graph). Fac: facade; WW/F: window-to-wall/floor; Irrad: irradiation; FA: floor area. Best viewed in color.

#### 4.2. Metamodel training and testing

In this section, the metamodel training and (‘internal’) testing results are analyzed individually for the MLR and GPs approach, followed by a comparison of the two methods when applied to an ‘external’ test set.

##### 4.2.1. Multiple linear regression (MLR)

Applying the stepwise algorithm on the entire dataset for each metric separately resulted in the selection of inputs listed in Fig. 6. These graphs show the main effect of each predictor on the output, when averaging out the effects of other predictors. The line represents a confidence interval. For example, changing the *FormFactor* from 0.67 to 1.89 causes a decrease in the energy need of about  $10 \pm 2$  kWh/m<sup>2</sup>. Given our set of inputs, caution must be taken when interpreting this information, as some predictors are correlated. This statement is supported by contradictions that can be observed between some of the main effects values for the energy metric and their corresponding correlation coefficient shown in Fig. 4. For instance, the *MeanEnvelopeIrrad* parameter presents a negative main effect (around -18), while a positive correlation of 0.60 was found when looking at the raw data.

It is therefore important to note that the MLR regression approach does not capture a physical representation of the problem. Yet, and independently from this interpretation limitation, it allows us to exploit the acquired information (reference datasets) in a way that can allow predicting the outputs to some level of accuracy.

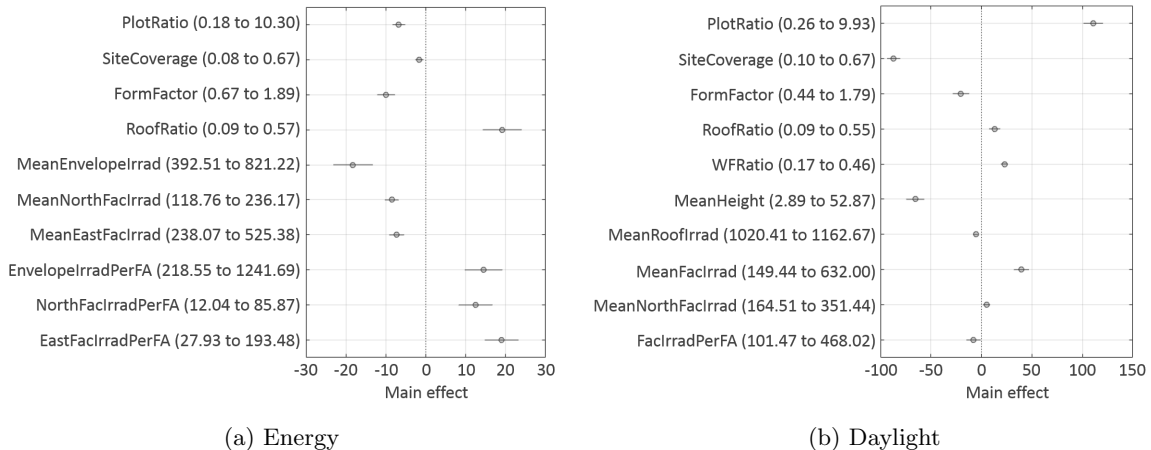


Figure 6: Main effects plot showing the change in the output associated to a change in the corresponding input from the stepwise regression model.



The set of inputs obtained through the stepwise regression were used in the subsequent step of applying ridge regression. Results are shown in the leftmost graphs of Fig. 7, where the value predicted by the ridge regression function are plotted against their corresponding simulated (ground truth) response. The performance statistics - average R, FAC2, and RMSE and mean and standard deviation of residuals - listed in Table 6a show a strong predictive performance. Indeed, the RMSE values represent at most a 3.5% and 10.7% error with respect to the lowest simulated energy need and spatial daylight autonomy respectively.

From the list of stepwise-selected predictors, some inputs were manually discarded and a further reduced version of each function was obtained after applying ridge regression once again. This additional manual step was done in order to investigate how the input coefficients would change, as well as verify if there would be a difference in the ‘generalization’ potential of the metamodels, when applied on the external test set (see §4.2.3). The high prediction accuracy observed during the internal train-test sequence, as mentioned above, leads us to doubt on how well the functions can perform on new, unseen cases. Indeed, these results are close to being ‘too good’; as Hastie et al. (2009) states: “a model with zero training error is overfit to the training data and will typically generalize poorly.” It was therefore of interest to generate a simplified version of the functions, thus the manual reduction, which was done based on a more or less arbitrary logic.

For the energy metric, we decided to cut the orientation-dependent parameters - namely the North and East irradiation inputs - to remove the need for obtaining such information from an eventual design that would not have its facades exactly facing the cardinal points. This meant going from 10 predictors in the ‘full’ version to 6 predictors in the ‘reduced’ one (not counting the constant term). From the daylight full model of 10 inputs, only two parameters were cut on the basis of their presumed physically weak link to the output metric: the *MeanRoofIrrad* and *MeanNorthFacIrrad*.

The center plots in Fig. 7 show the predicted versus simulated values obtained this reduced version of the metamodels. There is a small increase in the deviation of the points from the identity line compared to the full MLR results. The slightly larger error is captured in the numerical results in Table 6a.

As explained in §3.4.1, the final predictive models are obtained by fitting a function using all cases in the datasets. The model coefficients of the resulting MLR-metamodels for energy and daylight predictions are given in Table 5, for both the ‘full’ and ‘reduced’ versions.

These are non-normalized values; the order of magnitude of a coefficient depends on the

Table 5: Predictor coefficients defined with ridge regression based on the most influential parameters identified through the stepwise search ('full') and following a manual withdrawal of some predictors ('reduced').

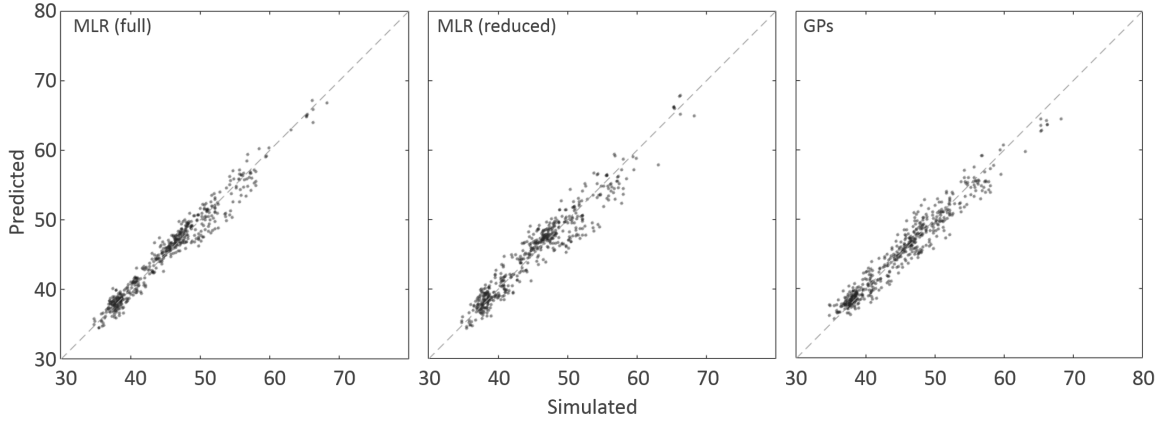
	<b>Energy function</b>		<b>Daylight function</b>	
	full	reduced	full	reduced
Constant ( $\beta_0$ )	78.82	84.81	104.48	70.33
<i>PlotRatio</i>	-0.67	-1.04	11.42	11.45
<i>SiteCoverage</i>	-2.70	1.80	-153.95	-147.64
<i>FormFactor</i>	-8.18	-11.23	-14.96	-21.28
<i>RoofRatio</i>	38.03	29.45	28.48	27.43
<i>WFRatio</i>			77.92	71.30
<i>MeanHeight</i>			-1.31	-1.27
<i>MeanEnvelopeIrrad</i>	-0.041	-0.075		
<i>MeanRoofIrrad</i>			-0.039	
<i>MeanFacIrrad</i>			0.082	0.095
<i>MeanNorthFacIrrad</i>	-0.074		0.027	
<i>MeanEastFacIrrad</i>	-0.026			
<i>EnvelopeIrradPerFA</i>	0.015	0.027		
<i>FacIrradPerFA</i>			-0.021	-0.030
<i>NorthFacIrradPerFA</i>	0.162			
<i>EastFacIrradPerFA</i>	0.112			

order of magnitude of its corresponding predictor value. As such, irradiation-based predictors have a much smaller coefficient due to their generally larger value (in the hundreds) compared to e.g. the plot ratio (max  $\approx 10$  in our dataset).

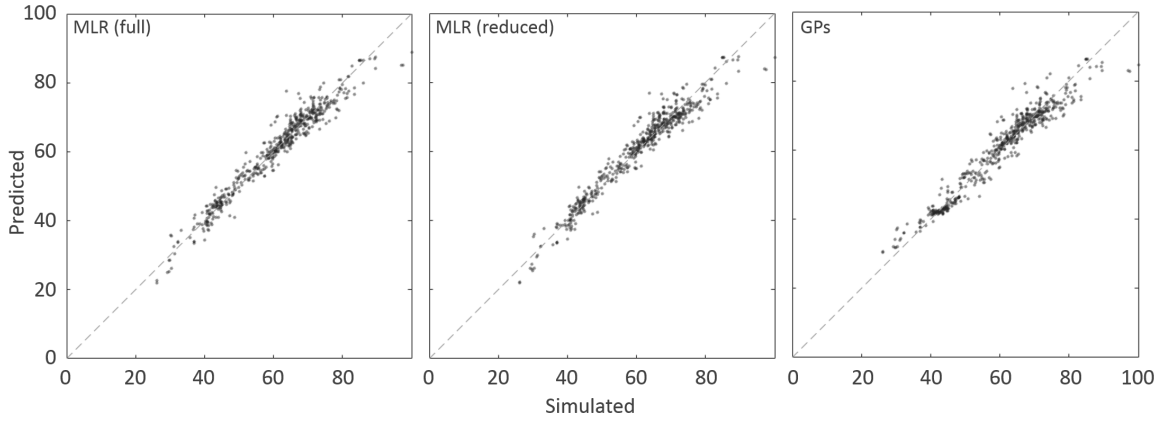
It is important to note that the inputs present in each function as well as their coefficients depend on the initial dataset, in particular the range spanned by each input. Parameters that one might expect to be influential, such as the window-to-floor ratio in the energy model, do not appear because they did not influence the response as much as other parameters present. The interdependency between the inputs present in a model is also highlighted by the differences in their coefficients in the full versus reduced version (Table 5). Moreover, subtle differences in the fitting procedure and training dataset can lead to different model coefficients as was experienced in other work (Nault et al., 2016a). In the same line of thought, cautious interpretation is required not to confuse correlation for causality when using statistical methods.

#### 4.2.2. Gaussian Processes (GPs)

Data analysis (§4.1) revealed that the reference dataset contains some redundant descriptors. MLR explicitly accounted for this redundancy by applying a stepwise algorithm. In



(a) Energy ( $\text{kWh/m}^2$ )



(b) Daylight (%)

Figure 7: Predicted versus simulated values in the metamodel training phase for the (a) energy and (b) daylight full and reduced MLR (left and center) and GPs (right) models. Prediction statistics can be found in Table 6a.

contrast, GPs are generally directly applied to the (scaled) reference dataset, without preliminary reduction, and implicitly account for redundancy. If an input parameter is redundant, the iterative calibration procedure will automatically assign a low (or even zero) posterior probability to fitting functions affected by the corresponding hyperparameter. This feature is known as Automatic Relevance Determination (ARD) (see e.g. Ch. 5 of (Rasmussen and Williams, 2006) for an in-depth explanation).

To facilitate comparison with the results of the MLR-metamodel, Bayesian calibration was separately performed on the energy and the daylight datasets, yielding two distinct GP-metamodels. Alternatively, a single GP-metamodel could be constructed having a vectorial output (Moonen and Allegrini, 2015).

Out of 25 inputs, the Bayesian calibration procedure only retained five. For the energy dataset these are *FormFactor*, *NorthFacRatio*, *WFRatio*, *MeanSouthFacIrrad* and *RoofIrradPerFA*. For the daylight dataset *SiteCoverage*, *SouthFacRatio*, *WestFacRatio*, *MeanEnvelopeIrrad* and *EnvelopeIrradPerFA* were retained. At this point two remarkable observations can be made: firstly, both lists do not have a single parameter in common; and secondly, the ‘relevant parameters’ identified by GP differ greatly from those by MLR, as these only share one common predictor. Although one might be tempted to explain these observations, they mainly illustrate the already mentioned fact that stochastic models do not identify causal relationships, but search for correlations.

The rightmost plots in Fig. 7 show the predicted (GP-metamodel) versus simulated (ground truth) responses for the (a) energy and (b) daylight metric for all cases. Each metamodel is controlled by five inputs, while the remaining 20 have no impact on the metamodel predictions (ARD). The average R, FAC2, and RMSE as well as the residuals mean and standard deviation values are displayed in Table 6a. The predictive performance is high and comparable to that of the MLR models, yet based on only half the number of inputs.

Table 6: Performance of metamodels during (a) training and (b) external testing sequence.  $\mu$ : mean;  $\sigma$ : standard deviation; res.: residuals (simulated - predicted).

(a) Performance of prediction during internal training on reference dataset.						
	<b>Metamodel</b>	<b>R</b>	<b>FAC2</b>	<b>RMSE</b>	$\mu$ <b>res.</b>	$\sigma$ <b>res.</b>
Energy	MLR full	0.98	1	1.21	0	1.21
	MLR reduced	0.96	1	1.71	0	1.71
	GPs	0.97	1	1.48	0.002	1.48
Daylight	MLR full	0.98	1	2.82	0	2.82
	MLR reduced	0.97	1	2.97	0	2.97
	GPs	0.97	1	3.16	0.01	3.17

(b) Performance of prediction during external testing on workshop variants.						
	<b>Metamodel</b>	<b>R</b>	<b>FAC2</b>	<b>RMSE</b>	$\mu$ <b>res.</b>	$\sigma$ <b>res.</b>
Energy	MLR full	0.80	1	4.28	-2.62	3.44
	MLR reduced	0.78	1	2.63	-0.71	2.58
	GPs	0.86	1	9.69	-9.49	2.01
Daylight	MLR full	0.59	1	7.14	1.34	7.14
	MLR reduced	0.64	1	6.28	0.42	6.37
	GPs	0.48	1	21.09	20.23	6.07

#### 4.2.3. Metamodel performance on external data

Metamodels, being trained on a reference dataset consisting of a limited number of design variants from the infinite realm of possible neighborhood designs, intrinsically have a limited validity. The external test serves to assess the predictive power of the metamodels by applying them to the 30 independent designs acquired in the context of workshops with practitioners (§3.4.3).

The predicted versus simulated values are shown in Fig. 8 for all metamodels for the (a) energy and (b) daylight metric. The corresponding prediction accuracy measures are listed in Table 6b.

For both approaches - MLR and GPs - results are better for the energy metric which shows a higher correlation coefficient (R). However, all energy models show a tendency towards over-prediction, as demonstrated by the residuals having a negative mean. The reduced MLR metamodel however seems to mitigate this over-prediction, with a near-zero mean of -0.71.

The predictions for the daylight dataset are more scattered regardless the type of metamodel (MLR or GPs). This is an indication that the training dataset is imperfect. Ideally, the reference dataset should homogeneously sample the entire design space, which can be ensured by relying on algorithms such as Latin Hypercube Sampling.

Comparing the performance of the full and reduced MLR models, we observe that the reduced version seems to perform better for both metrics, based on the RMSE and the  $\mu$  and  $\sigma$  of the residuals. In the case of the energy metric, the R value is however slightly higher for the full MLR model. These results highlight the risk of being misled when evaluating only one prediction performance measure such as the correlation coefficient (R).

The fact that the reduced MLR metamodels perform better tends to support our hypothesis that ‘simpler’ metamodels - i.e. with fewer predictors - may generalize better to a more diverse population of designs. Of course the structure of the model and the nature of the predictors are also important and not only the amount of inputs present in the functions. Indeed, the GPs models - which have the smallest number of predictors - did not perform as well on the external test set.

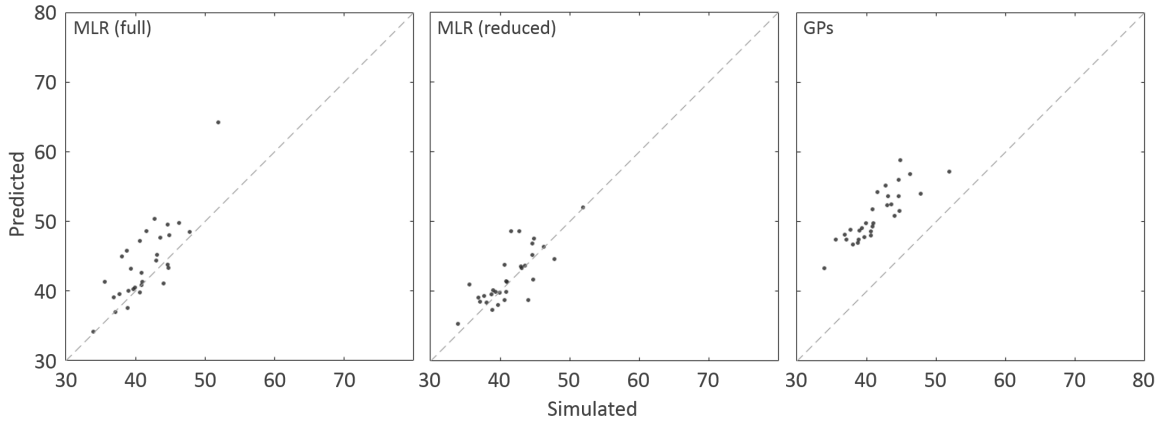
Figure 9 illustrates some of the workshop variants along with their rank out of the 30 variants and their associated prediction errors, where the percent error is calculated as:

$$\text{Percent error} = 100 \times \left| \frac{f(\mathbf{x}_j) - y_j}{y_j} \right| \quad (11)$$

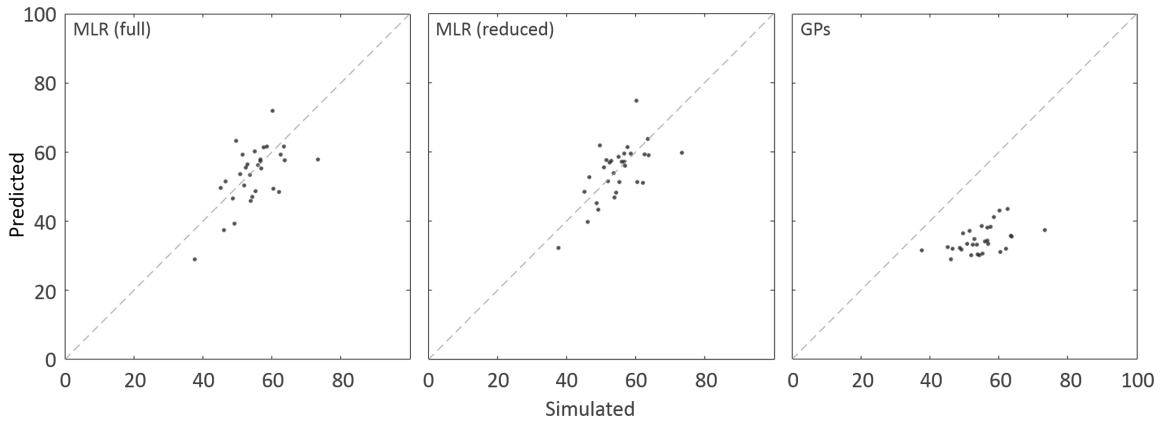
We observe that the ‘high’ error variants shown for both the (a) energy and (b) daylight metric are visually quite different from any of the cases ( $M\alpha$ ) present in the reference dataset. This discrepancy is likely to be reflected in the inputs values, in turn causing the metamodels to predict relatively poorly for those not only unseen but also peculiar designs.

The ‘low’ error examples have simpler geometries and fewer buildings. However, they still show a difference with respect to the reference dataset as they include mixed typologies, a feature not found in our case series where each  $M\alpha$  holds a unique, replicated building shape.

Given the fact that some of the design variants in this external test set significantly differ from what the metamodels have been trained on, and that only a limited number of descriptors are used, the tremendous potential of metamodels can be appreciated.



(a) Energy (kWh/m<sup>2</sup>)



(b) Daylight (%)

Figure 8: Predicted versus simulated values for the external test set, for the (a) energy and (b) daylight full and reduced MLR (left and center) and GPs (right) models. Prediction statistics can be found in Table 6b.

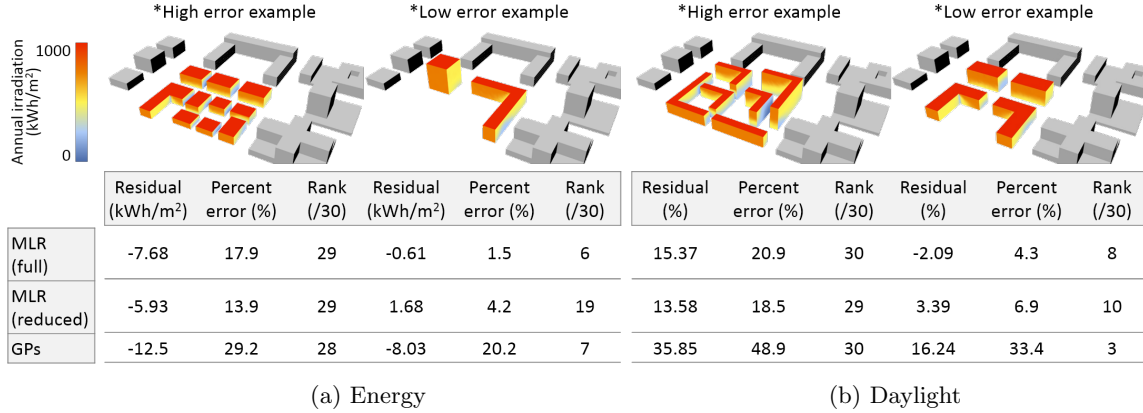


Figure 9: Irradiation map of example variants from the external test set, with associated prediction errors and rank for the (a) energy and (b) daylight metric. These examples were selected according to their position across all three ranks (one rank per metamodel type) and their prediction error. \*The ‘high’ and ‘low’ error categorization should be interpreted as a relative comparison for each metamodel.

### 4.3. Discussion

#### 4.3.1. Metamodel comparison

Following two metamodeling strategies, multiple linear regression and Gaussian Processes, models were trained on the same reference dataset and benchmarked. Both were capable to perform predictions of the energy and daylight datasets based on a very limited number of input parameters. When confronted to external test data, the prediction performance of all models decreased in a similar fashion, but remained valuable to support the needs in an early design phase. For both metrics, the reduced MLR version appears to have the best generalization potential, while the prediction accuracy of GP-based models seems to place more stringent requirements on the reference dataset.

GPs have the advantage over MLR in that they are based on a fully probabilistic framework, and thus, that uncertainties can be extracted for every prediction (Moonen and Allegrini, 2015). Furthermore, the Bayesian calibration framework automatically identifies the relevant input parameters. MLR has the advantage that the method is conceptually easier to grasp, and - when used with care - the presented results show that the method could outperform GPs.

#### 4.3.2. Limitations and potential for improvement

The metamodels presented in this paper were trained on a reference dataset containing geometrical parameters, solar exposure levels, and simulated energy and daylight per-

formance values for a finite set of neighborhood design variants. To obtain the simulated values, assumptions were made on aspects considered as detailed rather than early-phase design variables, such as U-values of opaque and glazed surfaces. Moreover, the simulations were performed only in one climate (Geneva). All results, including the correlation levels and metamodels structure, are therefore dictated by our reference dataset and underlying assumptions.

Future work will look at extending the dataset in terms of typology and climate as well as revising the assumptions according to location-specific regulations. Improvements will be brought to our sampling technique in order to ensure a better coverage of the possible input parameter values. These refinements shall allow applying the metamodels to a larger and more diverse neighborhood design population.

#### *4.3.3. Computational cost*

While a full energy simulation takes between 1 and 5 minutes per variant, depending on the size and complexity of the neighborhood design, the duration of the metamodel's prediction process is in the order of seconds (max ~1 minute)<sup>14</sup>. The time savings are more important for the daylight performance criterion, due to the particularly expensive daylight simulation which can take over 30 minutes per variant. In the metamodel approach, the main time-consuming task is the irradiation simulation.

Applying the metamodel also requires a simpler 3D model, composed of a volume for each building with no need to define thermal zones or facade composition - further decreasing the time consumption and complexity of setting-up and conducting a performance evaluation.

## **5. Conclusion**

This paper presents a prediction-based approach for assessing the passive solar (energy need for heating and cooling) and daylight (spatial daylight autonomy on ground floor) performance of neighborhood designs. The two metamodeling techniques compared - multiple linear regression and Gaussian Processes - appear as promising for generating functions capable of providing early-phase performance estimates from available design information.

Results show that by carefully selecting the most relevant input parameters when fitting

---

<sup>14</sup>These times were recorded for simulations run on an Intel Quad Core 3.70GHz computer with 16 GB of RAM.



the metamodel, to account for both morphology and solar exposure, we can achieve a prediction with a root mean square error (RMSE) below 2 kWh/m<sup>2</sup> for the energy metric and around 3 % for the daylight metric, when iteratively trained and tested on different subsets of our reference dataset. When confronted to a more diverse set of designs, the prediction accuracy decreases while remaining satisfyingly high - particularly for the energy metric and reduced MLR model - for the intended application.

The outcomes of this study remain limited by the assumptions linked to the metamodeling techniques and intrinsic to the reference dataset used for fitting the functions. However, the metamodels performance highlight their validity within certain boundaries e.g. in terms of typology, climate, and design parameter ranges.

We achieve significant savings in terms of computational time, particularly valuable when assessing multiple variants in the exploratory and comparative early design phase. We conclude that the proposed approach is a promising alternative to complex simulation and has great potential as an underlying engine in a decision-support framework enabling an efficient performance-based comparison of design alternatives.

## 6. Acknowledgement

This work was conducted at the Ecole polytechnique fédérale de Lausanne with additional support from the EuroTech Universities Alliance and the SECURE (Synergistic Energy and Comfort through Urban Resource Effectiveness) project funded by the CCEM (Competence Center Energy and Mobility) (E. Nault). Peter Moonen would like to acknowledge the Swiss National Science Foundation (SNF) for funding a scientific stay at University College London (UCL) under grant IZK0Z2\_150428.

## 7. References

- Asadi, S., Amiri, S. S., Mottahedi, M., 2014. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings* 85.
- Beckers, B., Rodriguez, D., 2009. Helping architects to design their personal daylight. *WSEAS Transactions on Environment and Development* 7 (5).
- Capozzoli, A., Mechri, H. E., Corrado, V., 2009. Impacts of architectural design choices on building energy performance applications of uncertainty and sensitivity techniques. In: *Proc. of IBPSA 2009*, Glasgow, Scotland.
- Cheng, V., Steemers, K., Montavon, M., Compagnon, R., 2006. Urban form, density and solar potential. In: *Proc. of PLEA 2006*, Geneva, Switzerland.

- Crawley, D. B., Pedersen, C. O., Lawrie, L. K., Winkelmann, F. C., 2000. EnergyPlus: Energy Simulation Program. *ASHRAE Journal* 42, 49–56.
- Eisenhower, B., O’Neill, Z., Narayanan, S., Fonoberov, V. A., Mezić, I., 2012. A methodology for meta-model based optimization in building energy models. *Energy and Buildings* 47 (0).
- Ekici, B. B., Aksoy, U. T., 2009. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software* 40 (5).
- Ekici, B. B., Aksoy, U. T., 2011. Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications* 38 (5).
- Evins, R., Allegrini, J., Moonen, P., 2014. Emulating site-specific wind flow information for use in building energy simulations. In: *Proc. of BSO 2014*, UCL, London, UK.
- Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A., 2013. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews* 23.
- Gauthier, R., Atelier Poisson and SDOL, 2012. Malley centre, ouest lausannois, les coulisses de malley; concours d’urbanisme et d’espaces publics à un degré; rapport du jury.
- Hachem, C., Athienitis, A., Fazio, P., 2012. Solar optimized neighbourhood patterns: Evaluation and guidelines. In: *Proc. of eSim 2012*.
- Hall, P., Dean, J., Kabul, I. K., Silva, J., 2014. An Overview of Machine Learning with SAS® Enterprise Miner™. In: *Proc. of SAS Global Forum 2014*.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R., 2009. *The elements of statistical learning*. Vol. 2. Springer.
- Hemsath, T. L., Alagheband Bandhosseini, K., 2015. Sensitivity analysis evaluating basic building geometry’s effect on energy use. *Renewable Energy* 76.
- Hensen, J., Lamberts, R. (Eds.), 2011. *Building performance simulation for design and operation*. Abingdon; Spon Press.
- Heo, Y., Choudhary, R., Augenbroe, G. A., 2012. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings* 47, 550–560.
- Higdon, D., Gattiker, J., Williams, B., Rightley, M., 2008. Computer model calibration using high dimensional output. *Journal of the American Statistical Association* 103, 570–583.
- Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., Modi, V., 2012. Spatial distribution of urban building energy consumption by end use. *Energy and Buildings* 45, 141–151.
- Hygh, J. S., DeCarolis, J. F., Hill, D. B., Ranji Ranjithan, S., 2012. Multivariate regression as an energy assessment tool in early building design. *Building and Environment* 57.
- IESNA, 2012. IES LM-83-12 IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE). Tech. Rep. IES LM-83-12, New York, NY, USA.
- Jakubiec, J. A., Reinhart, C. F., 2011. Diva 2.0: Integrating daylight and thermal simulations using rhinoceros 3d, daysim and energyplus. In: *Proc. of IBPSA 2011*. Sydney.
- Karteris, M., Slini, T., Papadopoulos, A., 2013. Urban solar energy potential in greece: A statistical calculation model of suitable built roof areas for photovoltaics. *Energy and Buildings* 62.

- Kennedy, M. C., O’Hagan, A., 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society (Series B)* 63, 425–464.
- Lam, J. C., Wan, K. K., Liu, D., Tsang, C., 2010. Multiple regression models for energy use in air-conditioned office buildings in different climates. *Energy Conversion and Management* 51 (12), 2692–2697.
- Larson, G., Shakespeare, R., 2011. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Randolph M. Fritz, Seattle.
- Lechner, H., 2009. *Heating, cooling and lighting*. New Jersey; John Wiley and Sons Inc.
- Lobaccaro, G., Fiorito, F., Masera, G., Poli, T., 2012. District geometry simulation: a study for the optimization of solar facades in urban canopy layers. In: *Proc. of SHC 2012*.
- Martins, T. A., Adolphe, L., Bastos, L. E., 2014. From solar constraints to urban design opportunities: Optimization of built form typologies in a Brazilian tropical city. *Energy and Buildings* 76, 43–56.
- Moonen, P., Allegrini, J., 2015. Employing statistical model emulation as a surrogate for cfd. *Environmental Modelling & Software* 72, 77–91.
- Nault, E., 2016. *Solar potential in early neighborhood design*. Ph.D. thesis, EPFL, Lausanne.
- Nault, E., Peronato, G., Rey, E., Andersen, M., 2015. Review and critical analysis of early-design phase evaluation metrics for the solar potential of neighborhood designs. *Building and Environment* 92, 679–691.
- Nault, E., Rey, E., Andersen, M., 2013. Early design phase evaluation of urban solar potential: Insights from the analysis of six projects. In: *Proc. of IBPSA 2013*. Chambéry, France.
- Nault, E., Rey, E., Andersen, M., 2016a. A multi-criteria decision-support workflow for early-stage neighborhood design based on predicted solar performance. In: *Proc. of PLEA 2016*. Los Angeles, USA.
- Nault, E., Rey, E., Andersen, M., 2016b. Urban planning and solar potential: assessing users’ interaction with a novel decision-support workflow for early-stage design. In: *Proc. of SBE16*. Zurich, Switzerland.
- PDL Gare-Lac, 2010. *Plan directeur localisé gare-lac*. Bauart Architectes et Urbanistes SA and Service de l’urbanisme et des bâtiments d’Yverdon-les-Bains.
- Peronato, G., 2014. *Built density, solar potential and daylighting. application of parametric studies and performance simulation tools in urban design*. Master’s Thesis, Università IUAV di Venezia.
- Pessenlehner, W., Mahdavi, A., 2003. Building morphology, transparency, and energy performance. In: *Proc. of IBPSA*, Eindhoven, Netherlands.
- Rasmussen, C. E., Williams, C. K. I. (Eds.), 2006. *Gaussian Processes for Machine Learning*. The MIT Press.
- Rey, E. (Ed.), 2013. *Green Density*. Presses Polytechniques et Universitaires Romandes.
- Ritter, F., Geyer, P., Borrmann, A., 2015. Simulation-based Decision-making in Early Design Stages. In: *Proc. of CIB W78*. Eindhoven, Netherlands.
- Simpson, T. W., Poplinski, J. D., Koch, P. N., Allen, J. K., 2001. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers* 17 (2).

- Sok Ling, C., Hamdan Ahmad, M., Remaz Ossen, D., 2007. The effect of geometric shape and building orientation on minimizing solar insolation on high-rise buildings in hot humid climate. *Journal of Construction in Developing Countries* 12 (1).
- Takebayashi, H., Moriyama, M., 2012. Relationships between the properties of an urban street canyon and its radiant environment: Introduction of appropriate urban heat island mitigation technologies. *Solar Energy* 86 (9).
- Tsanas, A., Xifara, A., 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings* 49.
- van Esch, M., Looman, R., de Bruin-Hordijk, G., 2012. The effects of urban and building design parameters on solar access to the urban canyon and the potential for direct passive solar heating strategies. *Energy and Buildings* 47.
- Ville de Lausanne, 2013. Les Plaines-du-Loup - Plan directeur localisé en vue de la réalisation d'un écoquartier. Tech. rep., Ville de Lausanne - Service d'urbanisme, TRIBU Architecture, CITEC, PAYSAGESTION.
- Zeiler, W., Savanovic, P., Quanjel, E., 2007. Design decision support for the conceptual phase of the design process. In: *Proc. of IASDR 2007*.
- Zhang, J., Heng, C. K., Malone-Lee, L. C., Huang, Y. C., Janssen, P., Hii, D. J. C., Ibrahim, N., 2012. Preliminary evaluation of a daylight performance indicator for urban analysis: Facade vertical daylight factor per unit floor area. In: *Proc. of SimBuild 2012*.
- Zhao, H.-x., Magoulès, F., 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews* 16 (6).