

Article

An Octahedric Regression Model of Energy Efficiency on Residential Buildings

Francisco J. Navarro-Gonzalez  and Yolanda Villacampa *

Department of Applied Mathematics, University of Alicante, 03690 Alicante, Spain; francisco.navarro@ua.es

* Correspondence: villacampa@ua.es; Tel.: +34-96-5903-691

Received: 23 October 2019; Accepted: 13 November 2019; Published: 19 November 2019



Featured Application: The use of this new regression method increases the available toolkit for system modeling and machine learning techniques. Their main advantages are simplicity and easy geometric meaning, allowing the treatment of a big class of modelling problems. The application to the study of a problem related with energy efficiency on buildings reviews a widely used dataset introducing some additional considerations that serve as a benchmark of the proposed methodology.

Abstract: System modeling is a main task in several research fields. The development of numerical models is of crucial importance at the present because of its wide use in the applications of the generically named machine learning technology, including different kinds of neural networks, random field models, and kernel-based methodologies. However, some problems involving the reliability of their predictions are common to their use in the real world. Octahedric regression is a kernel averaged methodology developed by the authors that tries to simplify the entire process from raw data acquisition to model generation. A discussion about the treatment and prevention of overfitting is presented and, as a result, models are obtained that allow for the measurement of this effect. In this paper, this methodology is applied to the problem of estimating the energetic needs of different buildings according to their principal characteristics, a problem that has importance in architecture and civil and environmental engineering due to increasing concerns about energetic efficiency and ecological footprint.

Keywords: octahedric regression; kernel estimator average; overfitting prevention; building energy efficiency

1. Introduction

Energy consumption rates are increasing up to the point of causing environmental problems associated to the use of non-renewable energy sources. Two main aspects should be considered in order to improve the situation. First, the development and improvement of renewable energy sources is receiving growing interest from researchers. Second, the reduction of waste and increments in the efficiency of energetic uses is a crucial point to take into account. Buildings contribute to this situation in an important form (between 20% to 40% of the overall energy consumption, depending on the country), so the number of studies devoted to its reduction have increased in the last few years.

Some papers focus on working with real data obtained at different sampling rate, trying to estimate and schedule a more sustainable energy consumption pattern [1,2] or physical and statistical models [3–6].

However, the diversity of buildings and the randomness of weather conditions difficult the problem of finding and modeling the influence of the different factors on the overall consumption to give advices addressed to reduce the environmental impact of existing or new constructions. One of

the biggest problems in this case is the availability of data to work with, because the difficulty of its collection. The existence of public free data at the UCI Machine Learning repository [7] may explain the number of studies based on the dataset of reference [8], where a variety of machine learning (ML) techniques have been applied, as Table 1 presents.

Table 1. Machine learning (ML) techniques used to model the reference [8] dataset.

Machine Learning Techniques	Papers
Artificial Neural Networks (ANN)	[2,9–19]
Decision Trees (DT)	[20]
Support Vector Regression (SVR)	[11,13,14,19,21–23]
Random Forest (RF) and Trees Ensemble	[8,13,14,20–29]
Multi-Layer Perceptron (MLP)	[14,20,23,27]
Gaussian Mixture Model (GMM)	[30]
Gradient Boosted Regression Trees (GBRT)	[24,31]
Extreme Learning Machine (ELM)	[10,32,33]
Linear Regression (LR)	[10,13,21,23,27,34,35]
Radial Basis Function Networks (RBFN)	[10,12,19,23,27]
Hybrid models	[10,11,19,22,28,36,37]
Multivariate Adaptive Regression Splines (MARS)	[10,19,23]
Gaussian Processes Regression (GPR)	[18,21,23,27,38]
k-Nearest Neighbor (k-NN)	[12,23]
Genetic and Evolutionary Techniques (GT)	[14,17,18,39,40]
Others	[11,17,18,21,27,28,35,41–45]

Studying the relationships between data and determining the dependence between them is a critical point in research in a wide number of fields, such as engineering, medicine, economics, sociology, etc. The problem can be stated as follows: given an experimental dataset,

$$\left\{ \left(x_{[k]}^1, x_{[k]}^2, x_{[k]}^3, \dots, x_{[k]}^d, y_{[k]} \right) \right\}_{k=1,2,\dots,p}, \quad (1)$$

corresponding to some standardized variables,

$$(x, y) = (x^1, \dots, x^d, y) \in [0, 1]^{d+1} \subset \mathbb{R}^{d+1}, \quad (2)$$

the goal is to determine the relationship between them through a function f with

$$y = f(x^1, \dots, x^d), \quad (3)$$

Two main approaches can be taken at this step, depending on the knowledge of the analytical expression corresponding to Equation (3). Where this analytical form is known (linear or non-linear), the problem is determining the values of the parameters of the estimator $z(x_{[k]}^1, \dots, x_{[k]}^d, \lambda_1, \lambda_2, \dots, \lambda_Q)$ that best approach the function $f(x^1, \dots, x^d)$, an operation that can be done by minimizing some class of error function $e(\lambda)$ defined on the set of parameters (for example, example the mean square error).

$$e(\lambda_1, \lambda_2, \dots, \lambda_Q) = \sum_{k=1}^p \left(y_{[k]} - z(x_{[k]}^1, \dots, x_{[k]}^d, \lambda_1, \lambda_2, \dots, \lambda_Q) \right)^2, \quad (4)$$

If $(\lambda_1^*, \lambda_2^*, \dots, \lambda_Q^*)$ is that minimum, the proposed approximation is $f(x^1, \dots, x^d) \cong z(x_{[k]}^1, \dots, x_{[k]}^d, \lambda_1^*, \lambda_2^*, \dots, \lambda_Q^*)$. If no analytical relationship can be supposed, the only option is using numerical methodologies. Many are available, and the use of one or another depends on several factors, including the preference of the researchers.

The finite element method (FEM) is a numerical method for finding numerical solutions to differential equations and boundary problems in partial differential equations, initially used for solving structural problems in civil and aeronautical engineering [46–51] and can be described as follows:

Given a differential equation defined by a differential operator, D ,

$$D(f) = v, \quad (5)$$

where $f, v \in V$, V being a function space, the finite element method replaces V by a finite dimensional subspace $V_h \subset V$, which is composed by continuous piecewise polynomial functions of degree K (Sobolev space), associated with a division of the domain, $[0, 1]^d$, where the problem is defined in N_e parts called elements Ω_i .

$$[0, 1]^d = \cup_{i=1}^{N_e} \Omega_i, \quad (6)$$

The problem to solve is now

$$D(f_h) = v_h \text{ where } f_h, v_h \in V_h, \quad (7)$$

Let us consider a basis for the functions of V_h , with $\dim V_h = Q$

$$V_h = \langle \varphi_1(x), \varphi_2(x), \dots, \varphi_Q(x) \rangle, \quad (8)$$

satisfying the next conditions at a set of Q points called nodes $(\zeta_1, \zeta_2, \dots, \zeta_Q)$

$$\varphi_i(\zeta_j) = \delta_{ij}, \quad (9)$$

The functions of the basis are called shape functions and they are used to interpolate in points different from the nodes. Thus, any function f can be developed in terms of the basis with components u_i taking the form

$$f_h(x) = \sum_{i=1}^Q u_i \cdot \varphi_i(x), \quad (10)$$

According to Equation (9), the sum is restricted to the nodes that determine the element which contains the point x . Several conditions can be defined to get the optimum values of u_i as an approximation for the solution to the differential equation.

The selection of the nodes is a crucial point for the precision of the results. The operation that constructs the set of nodes and elements is called meshing. The resulting mesh can be locally characterized by a parameter related to the size of each element and the error of its approximated solution.

The numerical regression model for the relation Equation (3) consists in the set formed by the nodal values $\{u_i\}$, which is called a representation model for the system. Then, the value of the relationship at any point can be estimated using the expression

$$f(x) \cong f_h(x) = \sum_{i=1}^Q u_i \cdot \varphi_i(x), \quad (11)$$

For simplicity in the subsequent calculations, let us use a mesh formed by regular hyper-cubic elements with edge length h . Introducing the complexity c as the number of elements in each dimension, $h = 1/c$, the total number of elements is c^d , and the total number of nodes is $(c + 1)^d$.

The key point in FEM-based methods is the determination of the values $\{u_i\}$ considering the definition of an error function and its minimization. That problem is equivalent to solve a linear system of size, $(c + 1)^d$.

Authors have been developing regression techniques based on the properties of the finite elements as method of approximation [52–55]. To guide the reader through the methods presented ahead, a short summary is included at this point:

1. Minimization of least squared error defined on the dataset.
 - a. Problems: Derived linear system is usually underdetermined.

2. Minimization of an energy associated with the deformation of the mesh defined on the problem domain.
 - b. Advantages: Associated linear system results always determined.
 - c. Problems: Computational time order of solving a symmetric linear system of size $(c + 1)^d$.
3. Introduction of multi-indexes dimensional decomposition over a Galerkin optimization scheme.
 - d. Advantages: Computational time order is $O((c + 1)^d)$.
 - e. Problems: Dependence of computational time as a dimensional power of complexity.
4. Local average of simple radial functions based estimators calculated on the finite element meshing nodes.
 - f. Advantages: Independence between the time of computation and the complexity.
 - g. Problems: Computational time order of $O(d \cdot 2^d)$ with a power dependence on the dimension.

The problem of numeric model estimation given the experimental dataset $\left\{ \left(x_{[k]}^1, x_{[k]}^2, x_{[k]}^3, \dots, x_{[k]}^d, y_{[k]} \right) \right\}_{k=1,2,\dots,p}$, which comes from measurements of a system determined by the relationship $y = f(x^1, \dots, x^d)$ that has an unknown expression, is equivalent to determining a function $z(x^1, \dots, x^d)$ defined in terms of an algorithm to its calculation that results in a minimum over some kind of global error E obtained as a function $H(\vec{e})$ of individual errors $e_{[k]}$.

$$E = H(e_{[1]}, e_{[2]}, \dots, e_{[p]}), \quad (12)$$

If Equation (12) is calculated at each point of the dataset, it results in

$$e_{[k]} = \Xi \left(y_{[k]} - z \left(x_{[k]}^1, x_{[k]}^2, x_{[k]}^3, \dots, x_{[k]}^d \right) \right), \quad (13)$$

where Ξ represents a function acting on the local error (usually the absolute value or squared). Presented in that form, the modeling problem is in some way equivalent to the problem of solving a differential equation, considered in Equation (5), and the same techniques based on the reduction of the dimension of the functional space of the solutions can be developed.

One natural form to do this is applying the finite element methods in which the problem is reduced to the calculation on the nodal values. So, the model is determined by the value of the Q nodes

$$y(x^1, \dots, x^d) = z(x^1, \dots, x^d, u^1, \dots, u^Q), \quad (14)$$

This basic idea has been developed by the authors in references [52,53]. The error of a model is defined as

$$E(u^1, \dots, u^Q) = \sum_{k=1}^p \left(y_{[k]} - z \left(x_{[k]}^1, \dots, x_{[k]}^d, u^1, \dots, u^Q \right) \right)^2, \quad (15)$$

Usually, the obtained linear system is underdetermined, so additional conditions over the nodal values are required. The cause of these degrees of freedom is the absence of sample points in some elements of the discretized domain. These additional constraints come from the “minimum deformation” or “rigidization” condition, where the value of an undetermined node is calculated as the average on the values of the adjacent nodes,

$$u_{unknown}^r = \frac{1}{\eta_r} \cdot \sum_{v=1}^{\eta_r} u^v, \quad (16)$$

A deeper use of the minimum deformation principle is used in reference [54], where the point of view is a bit different. The geometric image of the mesh deformation introduced in the previous paragraph has been developed until the definition of an energy associated with the model. Furthermore, the energy U is directly inspired by the elastic energy of a bi-dimensional mesh composed by vertex joined by springs (with elongation V_{elong} and flexion V_{flex} components for the energy of the mesh) and an interaction term $V_{interact}$ accounting for the attractive effect between the surface of the defined mesh and the experimental points.

$$U(u^1, \dots, u^Q) = V_{interact}(u^1, \dots, u^Q) + c_f \cdot V_{flex}(u^1, \dots, u^Q) + c_e \cdot V_{elong}(u^1, \dots, u^Q), \quad (17)$$

To simplify, under the assumption of a general principle of smoothness of the model, the last expression approximates as

$$U(u^1, \dots, u^Q) = V_{interact}(u^1, \dots, u^Q) + c_f \cdot V_{flex}(u^1, \dots, u^Q), \quad (18)$$

where c_f corresponds to the relative coupling between the different energy components. Using this global energy for the system, the equilibrium point can be obtained as a minimization problem whose solution takes the form of a linear system,

$$(N + c_f \cdot \Theta) \cdot U = Y, \quad (19)$$

The main advantage over the first methodology is that the system obtained has a unique solution and no additional time-wasting processes, like rigidization, are needed.

Both methods suffer of a common bottleneck related with the algorithms available to solve the linear system. However, given that the discretization is composed of hyper-cubes, it is possible to introduce a multi-index system $(i_1, \dots, i_d)$ for each node and element. Using this notation, the shape function in any dimension can be calculated as the products of shape functions of dimension one $\varphi_{i_j}^{[1]}(x^j)$,

$$\varphi_{i_1, \dots, i_d}(x^1, \dots, x^d) = \prod_{j=1}^d \varphi_{i_j}^{[1]}(x^j), \quad (20)$$

Then, the system Equation (19) and that derived from Equation (15) can be built following a structure that allows for a faster solution. In particular, in reference [55], the system is obtained from Galerkin's method using a definition of the error given by

$$z_{\{E(x)\}} = \sum_{i_1, \dots, i_d} u_{i_1, \dots, i_d} \cdot \varphi_{i_1, \dots, i_d}(x^1, \dots, x^d), \quad (21)$$

where $z_{\{E(x)\}}$ is an approximation of the unknown function as a weighted average of the experimental points for each element defined in terms of a radial function ϕ ,

$$z_{\{E(x)\}} = \frac{\sum_{k=1}^P y_{[k]} \cdot \phi(\|x_{[k]} - \eta_E\|)}{\sum_{k=1}^P \phi(\|x_{[k]} - \eta_E\|)}, \quad (22)$$

where η_E is the geometric center of the corresponding element.

The method can then be seen as a two-step process. In the first part, the use of element averages approximates the model by a piecewise function that is afterward smoothed.

The use of a radial weighted average is in fact a simple model that is used to compose the joined structure of the linear system through the Galerkin's error minimization formula. The linear system appears because the model is considered in terms of the determination of all the nodes of the domain in the FEM discretization. However, in fact, obtaining the best estimation for a point is more a local than a global problem as it is suggested by Equation (22). This global character causes that all the

presented methods have computing times depending dramatically on the complexity c , that is, related to the good of fitness of the model.

To escape from this global requirement, one possibility is to construct only local models as Equation (22). Nevertheless, the calculation of the average is influenced by the experimental errors, and it is desirable to reduce the impact of the individual errors at the experimental points. Therefore, if the model is estimated in different points, the influence of these errors will vary from one point to another.

Given that the finite element model introduces a structure composed by points separated by a distance parametrized by the model complexity $h = 1/c$ and a method for interpolate the values defined on the nodes of an element, a natural option is using radial functions $\Phi(x)$ in the calculation of simple models on each node,

$$c_*(\zeta_r) = \frac{\sum_{k=1}^P y_k \cdot \Phi(\|x_k - \zeta_r\|)}{\sum_{k=1}^P \Phi(\|x_k - \zeta_r\|)}, \quad (23)$$

Afterward, the interpolation of these values through the form functions

$$\hat{z}(x_0) = \sum_{r=1}^Q \varphi_r(x_0) \cdot c_*(\zeta_r), \quad (24)$$

Although this methodology improves some of the problems commented previously, it has also some points that should be considered.

First, the effectiveness of the error smoothing process Equation (24) is optimum for points near the center of the element, but the points beside a node are mainly determined by the value of the radial average on this node. Second, the use of hyper-cubic elements presents an advantage from the point of view of shape function calculations as considered in Equation (20), but the number of nodes is 2^d , introducing a dependence with the dimension in the computational complexity of the algorithm of $O(d \cdot 2^d)$.

These problems are solved with the methodology presented in the following point, called Octahedric regression. It represents an improvement in the computational order $O(d \cdot 2^d)$ of the algorithm, a feature that will allow the modelling of systems with higher dimensions.

The presented methodology (Octahedric regression) is a hybrid methodology, including the main characteristics of the finite element method, radial basis function, and nearest neighbors. It can be explained as a two-step algorithm where the result is calculated after a FEM-like interpolation process acting on a set of simple estimators obtained with a weighted version of a $(1+\epsilon)$ -approximate nearest neighbor.

Octahedric regression presents some degree of similitude with the technique of kriging. The main difference is that the set of points used to calculate the final result has a fixed structure (forming an octahedral around the objective point), and the quantities involved in the interpolation are not the sample values, but predictors obtained using a radial-based weighted average of the experimental data.

The rest of the paper is organized as follows: Section 2 introduces the octahedric regression methodology and its computational algorithm. The last part of Section 2 is devoted to the study of the treatment of overfitting from the point of view of the new methodology. Section 3 presents the application of the octahedric regression to the problem of determining heating and cooling loads using the characteristics of a group of buildings. Finally, Section 4 exposes the conclusions and future investigations.

2. Materials and Methods

Definition 1. A parametrised radial function [56,57], is a function $\Phi : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ characterized by a parameter ω that accomplishes the conditions

$$\left. \begin{array}{l} \forall \omega \neq 0 \lim_{r \rightarrow \infty} \Phi(r, \omega) = 0 \\ \forall r \neq 0 \lim_{\omega \rightarrow 0} \Phi(r, \omega) = 0 \end{array} \right\}, \quad (25)$$

An example of parametrised radial function is the exponential

$$\Phi(r, \omega) = e^{-\frac{r}{\omega}}, \quad (26)$$

The following definitions introduce the basic tools used afterwards to develop the methodology.

Definition 2. Given a function $f(x)$ defined on a domain, $\Omega = [0, 1]^d$, the weighted average regression of $f(x)$ in a point x_0 , $c_*(x_0)$, is a number implicitly defined as

$$\int_{\Omega} [c_*(x_0) - f(x)] \Phi(\|x - x_0\|, \omega) dx = 0, \quad (27)$$

where $\Phi(., .)$ is a parametrised radial function.

Definition 3. The J -th moment of the radial function Φ is defined as

$$W^J(x_0) = \int_{\Omega} \|x - x_0\|^J \cdot \Phi(\|x - x_0\|, \omega) dx, \quad (28)$$

Using this expression, Equation (27) can be written as

$$c_*(x_0) \cdot W^0(x_0) = \int_{\Omega} f(x) \cdot \Phi(\|x - x_0\|, \omega) dx, \quad (29)$$

Definitions 4 and 5 introduce the concept of support of an interpolation to be used in the following definitions.

Definition 4. Given a point $x_0 \in \Omega$ and a set of Q d -dimensional vectors $\{\zeta_r\}_{r=1}^Q$, with $Q \geq d = \dim(\Omega)$, the set of points obtained by the combinations $\{x_0 + \zeta_r\}_{r=1}^Q \subset \Omega$ is called a support for interpolation around x_0 .

Definition 5. An octahedric support for interpolation of size h around x_0 is the support for interpolation around x_0 given by the $2 \cdot d$ vectors

$$\left\{ \frac{h}{2} \cdot e_1, \frac{h}{2} \cdot e_2, \dots, \frac{h}{2} \cdot e_d, -\frac{h}{2} \cdot e_1, -\frac{h}{2} \cdot e_2, \dots, -\frac{h}{2} \cdot e_d \right\}, \quad (30)$$

Definitions 6 and 7 present the main objective of the research in form of an interpolation of several radial function-based estimations.

Definition 6. Given an octahedric support for interpolation of size h around a point x_0 and a function $f(x)$, the interpolated function $\hat{f}(x)$ is defined as

$$\hat{f}(x) = \frac{1}{2d} \cdot \sum_{i=1}^d \left[f\left(x_0 + \frac{h}{2} \cdot e_i\right) + f\left(x_0 - \frac{h}{2} \cdot e_i\right) \right], \quad (31)$$

Definition 7. Given a function defined on Ω , and a point $x_0 \in \Omega$, the octahedric estimation $\hat{z}(x)$ of width h of $f(x)$ at x_0 is defined as the interpolation of the weighted averages on its octahedric support of size h , given by

$$\hat{z}(x_0) = \frac{1}{2d} \cdot \sum_{i=1}^d \left[c_*\left(x_0 + \frac{h}{2} \cdot e_i\right) + c_*\left(x_0 - \frac{h}{2} \cdot e_i\right) \right], \quad (32)$$

where $c_*(.)$ are the numbers defined in Definition 2. By similitude with the methods introduced in Section 1, the value of $\frac{1}{h}$ is called pseudo-complexity.

To study the meaning of the definition 7, let consider the case when h is small. From Equation (29),

$$c_*\left(x_o \pm \frac{h}{2} \cdot e_i\right) \cdot W^0\left(x_o \pm \frac{h}{2} \cdot e_i\right) = \int_{\Omega} f(x) \cdot \Phi\left(\|x - (x_o \pm \frac{h}{2} \cdot e_i)\|, \omega\right) dx, \quad (33)$$

Developing W^0 and Φ around x_o up to second order in h ,

$$W^0\left(x_o \pm \frac{h}{2} \cdot e_i\right) = W^0(x_o) \pm \partial_i W^0(x_o) \cdot \frac{h}{2} + \frac{h^2}{8} \cdot \partial_i^2 W^0(x_o) + O(h^3), \quad (34)$$

$$\Phi\left(\|x - x_o \mp \frac{h}{2} \cdot e_i\|, \omega\right) = \Phi(\|x - x_o\|, \omega) \mp \partial_i \Phi(\|x - x_o\|, \omega) \cdot \frac{h}{2} + \frac{h^2}{8} \cdot \partial_i^2 \Phi(\|x - x_o\|, \omega) + O(h^3), \quad (35)$$

Equation (33) takes the form

$$\begin{aligned} & \left[W^0(x_o) \pm \partial_i W^0(x_o) \cdot \frac{h}{2} + \frac{h^2}{8} \cdot \partial_i^2 W^0(x_o) \right] \cdot c_*\left(x_o \pm \frac{h}{2} \cdot e_i\right) = \\ & = \int_{\Omega} f(x) \left[\Phi(\|x - x_o\|, \omega) \mp \partial_i \Phi(\|x - x_o\|, \omega) \cdot \frac{h}{2} + \frac{h^2}{8} \cdot \partial_i^2 \Phi(\|x - x_o\|, \omega) \right] dx, \end{aligned} \quad (36)$$

That is,

$$\begin{aligned} & \left[W^0(x_o) \pm \partial_i W^0(x_o) \cdot \frac{h}{2} + \frac{h^2}{8} \cdot \partial_i^2 W^0(x_o) \right] \cdot c_*\left(x_o \pm \frac{h}{2} \cdot e_i\right) = \int_{\Omega} f(x) \cdot \Phi(\|x - x_o\|, \omega) dx \mp \frac{h}{2} \cdot \\ & \int_{\Omega} f(x) \cdot \partial_i \Phi(\|x - x_o\|, \omega) dx + \frac{h^2}{8} \cdot \int_{\Omega} f(x) \cdot \partial_i^2 \Phi(\|x - x_o\|, \omega) dx, \end{aligned} \quad (37)$$

Summing negative and positive components of the support on dimension i ,

$$\begin{aligned} & W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) + c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] + \frac{h}{2} \cdot \partial_i W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) - c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] \\ & + \frac{h^2}{8} \cdot \partial_i^2 W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) + c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] = \\ & = 2 \cdot c_*(x_o) \cdot W^0(x_o) + \frac{h^2}{4} \cdot \int_{\Omega} f(x) \cdot \partial_i^2 \Phi(\|x - x_o\|, \omega) dx, \end{aligned} \quad (38)$$

Summing now for every dimension index i and dividing by $2 \cdot d$:

$$\begin{aligned} & W^0(x_o) \cdot \hat{z}(x_o) + \frac{h}{4 \cdot d} \cdot \sum_i^d \partial_i W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) - c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] + \frac{h^2}{16 \cdot d} \\ & \cdot \sum_i^d \partial_i^2 W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) + c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] \\ & = c_*(x_o) \cdot W^0(x_o) + \frac{h^2}{8 \cdot d} \cdot \sum_i^d \int_{\Omega} f(x) \cdot \partial_i^2 \Phi(\|x - x_o\|, \omega) dx \end{aligned} \quad (39)$$

Therefore, the octahedric regression is

$$\begin{aligned} \hat{z}(x_o) = & c_*(x_o) - \frac{1}{4 \cdot d \cdot W^0(x_o)} \cdot \left[h \cdot \sum_i^d \partial_i W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) - c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] + \frac{h^2}{4} \right. \\ & \cdot \left. \sum_i^d \partial_i^2 W^0(x_o) \cdot \left[c_*\left(x_o + \frac{h}{2} \cdot e_i\right) + c_*\left(x_o - \frac{h}{2} \cdot e_i\right) \right] - \frac{h^2}{2} \cdot \int_{\Omega} f(x) \cdot \sum_i^d \partial_i^2 \Phi(\|x - x_o\|, \omega) dx \right], \end{aligned} \quad (40)$$

Taking as radial function given by Equation (26),

$$\Phi(\|x\|, \omega) = e^{-\frac{\|x\|}{\omega}} \rightarrow \partial_i \Phi(\|x\|, \omega) = -\frac{1}{\omega} \cdot \Phi(\|x\|, \omega) \cdot \partial_i \|x\| = -\frac{1}{\omega} \cdot \Phi(\|x\|, \omega) \cdot \frac{x^i}{\|x\|}, \quad (41)$$

So, for points far from hypercube's boundary, by symmetry,

$$\partial_i W^0(x) = \partial_i \int_{\Omega} \Phi(\|t-x\|, \omega) dt = \int_{\Omega} \partial_i \Phi(\|t-x\|, \omega) dt = \frac{1}{\omega} \cdot \int_{\Omega} \Phi(\|t-x\|, \omega) \cdot \frac{(t-x)^i}{\|t-x\|} dt = 0, \quad (42)$$

Equation (40) becomes

$$\hat{z}(x_0) = c_*(x_0) + \frac{h^2}{16 \cdot d \cdot W^0(x_0)} \cdot \left[2 \cdot \int_{\Omega} f(x) \cdot \sum_i^d \partial_i^2 \Phi(\|x-x_0\|, \omega) dx - \sum_i^d \partial_i^2 W^0(x_0) \cdot \left[c_* \left(x_0 + \frac{h}{2} \cdot e_i \right) + c_* \left(x_0 - \frac{h}{2} \cdot e_i \right) \right] \right], \quad (43)$$

That is, octahedric regression is a correction of order h^2 to the weighed average for central objective points. As a consequence, when $h \rightarrow 0$, both values tend to coincide $\hat{z}(x_0) \rightarrow c_*^s(x_0)$.

The following definition and propositions are related with the behaviour of the estimations with relation to the experimental error distributions, considered as normal and uncorrelated over the problem domain.

Definition 8. Given a function $f(x)$ defined on a domain Ω and a random field $e(x) \sim N(0, \sigma(x))$ defined on Ω , an experimental realisation of f associated with a sample $e^s(x)$ of $e(x)$ is a function $y^s : \Omega \rightarrow \mathbb{R}$, given by

$$y^s(x) = f(x) + e^s(x). \quad (44)$$

Proposition 1. The expected value (denoted as $E[\dots]$) of the weighted averaged regression corresponding to the values y^s in any point x_0 is

$$E[c_*^s(x_0)] = c_*(x_0). \quad (45)$$

Proof. Following the Equation (27), the regression of the experimental realisation is

$$\int_{\Omega} [c_*^s(x_0) - y^s(x)] \cdot \Phi(\|x-x_0\|, \omega) dx = 0, \quad (46)$$

Using Equation (44), the last expression can be written as

$$\int_{\Omega} [c_*^s(x_0) - f(x)] \cdot \Phi(\|x-x_0\|, \omega) dx = \int_{\Omega} e^s(x) \cdot \Phi(\|x-x_0\|, \omega) dx. \quad (47)$$

Therefore, by Equation (27) and using the definition (Equation (29)) of $W^0(x_0)$,

$$(c_*^s(x_0) - c_*(x_0)) \cdot W^0(x_0) = \int_{\Omega} e^s(x) \cdot \Phi(\|x-x_0\|, \omega) dx. \quad (48)$$

Under the conditions of definition 8, $E[e^s(x)] = 0$, and the expected value of Equation (48) is

$$E[c_*^s(x_0) - c_*(x_0)] \cdot W^0(x_0) = 0. \quad (49)$$

So,

$$E[c_*^s(x_o)] = c_*(x_o). \quad (50)$$

□

Proposition 2. Let $e(x)$ a distribution not correlated for different points of Ω , $E[e^s(x) \cdot e^s(y)] = \sigma^2(x) \cdot \delta(x - y)$. Then, the variance of the weighted averaged regression corresponding to the values y^s is given by

$$Var(c_*^s(x_o) - c_*(x_o)) \cdot (W^0(x_o))^2 = \int_{\Omega} \sigma^2(x) \Phi^2(\|x - x_o\|, \omega) dx. \quad (51)$$

Proof. Taking Equation (48) to the square and calculating expected values,

$$\begin{aligned} & E[(c_*^s(x_o) - c_*(x_o))^2] \cdot (W^0(x_o))^2 \\ &= \int_{\Omega} \int_{\Omega} E[e^s(x) \cdot e^s(y)] \cdot \Phi(\|x - x_o\|, \omega) \cdot \Phi(\|y - x_o\|, \omega) dx dy \\ &= \int_{\Omega} \sigma^2(x) \cdot \Phi^2(\|x - x_o\|, \omega) dx. \end{aligned} \quad (52)$$

That is,

$$Var(c_*^s(x_o) - c_*(x_o)) \cdot (W^0(x_o))^2 = \int_{\Omega} \sigma^2(x) \cdot \Phi^2(\|x - x_o\|, \omega) dx. \quad (53)$$

□

Now, the computational algorithm is presented with some considerations relative to its implementation.

2.1. Computational Algorithm

In real cases, complete sets of values of $y^s(x)$ in the domain Ω are not available, so one must use only a subset of P points. For this sample, the integrals are calculated using finite sums, so

$$\int_{\Omega} \Phi(\|x - x_o\|, \omega) dx \cong \frac{1}{P} \cdot \sum_{k=1}^P \Phi(\|x_k - x_o\|, \omega), \quad (54)$$

$$\int_{\Omega} y^s(x) \cdot \Phi(\|x - x_o\|, \omega) dx \cong \frac{1}{P} \cdot \sum_{k=1}^P y_k^s \cdot \Phi(\|x_k - x_o\|, \omega). \quad (55)$$

Now, the finite sample version of Equation (46) is

$$\sum_{k=1}^P [c_*^s(x_o) - y_k^s] \cdot \Phi(\|x_k - x_o\|, \omega) = 0. \quad (56)$$

Following (Equation (56)), the proposed algorithm (Algorithm 1) can be condensed in the next schema:

Let us introduce the size of the problem data as $= P \cdot d$. Given that the calculation of the radial kernel function has a computational cost of $O(d)$, where d is the dimension, the number of operations of the algorithm can be calculated as $O(2)$, while the memory requirements will be $O()$.

Algorithm 1. Octahedric regression algorithm

```

1: for i in Integers{1 .. P}
2:   Real estimation[i] = 0
3:   for k in Integers{1 .. dimension}
4:     Real estim_plus = 0, estim_minus = 0
5:     Real dist_plus = 0, dist_minus = 0
6:     for j in Integers{1 .. P}
7:       dist_plus = dist_plus + Radial_Kernel(X[i] + h × ei, X[j])
8:       dist_minus = dist_minus + Radial_Kernel(X[i] - h × ei, X[j])
9:       estim_plus = estim_plus + Z[j] × Radial_Kernel(X[i] + h × ei, X[j])
10:      estim_minus = estim_minus + Z[j] × Radial_Kernel(X[i] - h × ei, X[j])
11:    estim_plus = estim_plus/dist_plus
12:    estim_minus = estim_minus/dist_minus
13:    estimation[i] = estimation[i] + (estim_plus + estim_minus)/(2 × dimension);

```

2.2. Full and Restricted Models as a Prevention of Overfitting

Equation (43) shows the dependence of the estimator on the parameter h . However, the algorithm includes one additional parameter, ω , used in the radial function. Let us order the points depending on the distance to x_0 . Taking Equation (54) of the function $W^0(x)$ at the support points $x_0 \pm \frac{h}{2} \cdot e_i$,

$$W^0\left(x_0 \pm \frac{h}{2} \cdot e_i\right) \approx \Phi\left(\frac{h}{2}, \omega\right) \cdot \left\{1 + \frac{\Phi\left(\|x_1 - x_0 \mp \frac{h}{2} \cdot e_i\|, \omega\right)}{\Phi\left(\frac{h}{2}, \omega\right)} + \dots + \frac{\Phi\left(\|x_P - x_0 \mp \frac{h}{2} \cdot e_i\|, \omega\right)}{\Phi\left(\frac{h}{2}, \omega\right)}\right\}, \quad (57)$$

The term on brackets represent the relative weight of each sample point in the weighted average. Summing up the contributions on each support point, the result can be written as

$$1 + \Psi(x_1 - x_0, \omega) + \dots + \Psi(x_P - x_0, \omega), \quad (58)$$

where

$$\Psi(x_k - x_l, \omega) = \frac{\sum_{i=1}^d \left\{ \Phi\left(\|x_k - x_l - \frac{h}{2} \cdot e_i\|, \omega\right) + \Phi\left(\|x_k - x_l + \frac{h}{2} \cdot e_i\|, \omega\right) \right\}}{2 \cdot d \cdot \Phi\left(\frac{h}{2}, \omega\right)}, \quad (59)$$

The value of

$$\Pi(x_r) = 1 + \sum_{k \neq r}^P \Psi(x_k - x_r, \omega), \quad (60)$$

represents the weight of all the points in the estimation of the model calculated at x_r . At zero order in h , Equation (59) implies that Equation (60) is approximately

$$\Pi(x_r) \cong 1 + \frac{\sum_{k \neq r}^P \Phi(\|x_k - x_r\|, \omega)}{\Phi\left(\frac{h}{2}, \omega\right)}, \quad (61)$$

By Equation (25), the fractions converge very fast to 0 as the distance to x_r grows, and the only contribution depends on the points that are at a distance similar to the nearest point, that is, when $\Phi(x_k - x_r, \omega) \approx \Phi\left(\frac{h}{2}, \omega\right)$. The number of q points involved are determined by the value of the parameter ω , and $c_S(x_0)$ is a weighted mean of the q -nearest points. So, the octahedric regression presented in the present paper corresponds to a mean of a simpler estimator calculated on the support of size h defined around x_0 . These simple estimators correspond to the weighted average of the $q(x_0, \omega)$ -nearest neighbours. In the limit, $\omega \rightarrow 0$, $c_S(x_0 \pm h \cdot e_i)$ are obtained from the nearest point. In the case of considering all the experimental points, the model is called full. In that case, when $h \ll 1$, the nearest to the support points will be very frequently x_0 , and then: $c_S(x_0 \pm h \cdot e_i) \rightarrow y_0$, and according to Equation (43), $\hat{z}(x_0) \rightarrow y_0$, confirming the trend to overfitting.

The overfitting is caused by the incorporation of noise to the model. If the point that is being calculated is not included in the estimation, the points that are used to obtain the values of $c_S(x_0 \pm h \cdot e_i)$ have a greater probability of presenting independent noise influence, diminishing in that form the overfitting. The new model calculated in this form is called restricted and corresponds to having a cross-validation for each experimental point where the test set is formed by itself.

3. Results

The fit of a model can be measured using different parameters. If y_i and $\tilde{y}_i$ are the observed and estimated values,

1. Mean squared error: $MSE = \frac{1}{P} \cdot \sum_{i=1}^P (y_i - \tilde{y}_i)^2$
2. Mean absolute error: $MAE = \frac{1}{P} \cdot \sum_{i=1}^P |y_i - \tilde{y}_i|$
3. Mean absolute percentage error: $MAPE = \frac{100}{P} \cdot \sum_{i=1}^P \left| \frac{y_i - \tilde{y}_i}{y_i} \right|$
4. Regression error characteristic (REC) curve: It is a curve obtained plotting the error tolerance on the X-axis versus the percentage of points predicted within the tolerance on the Y-axis.

As a case study, the energy efficiency dataset [8] from the UCI Machine Learning Repository [7] has been selected. The data corresponds to the input values of two simulations generated by the software Ecotect [58] from Autodesk, San Rafael (CA) US, representing the heating and cooling loads necessary to achieve comfortable indoor conditions. The dataset variables are represented in Table 2.

Table 2. Input and output variables for the energy efficiency problem.

Input Variables	Output Variables
Relative compactness	Heating load
Surface area	Cooling load
Wall area	-
Roof area	-
Overall height	-
Orientation	-
Glazing area	-
Glazing area distribution	-

The studied buildings are generated using 18 cubes with a side length of 3.5 m, forming 12 different building forms with equal volume but different areas and dimensions. Each side can act as different elements (wall, floor, roof, or window) with different thermodynamic properties. The different combinations of values allowed for the model can be consulted with more detail in reference [8]. As a result of the process, 768 buildings are simulated, and the results for both simulations (heating and cooling cases) are presented in the dataset.

The quality parameters obtained for the learning techniques used in reference [8] are shown in Table 3. Iteratively reweighted least squares (IRLS) is a method used to diminish the effect of outliers in classical linear regression [59], while RF stands for random forest.

Table 3. Parameters of quality for models of energy efficiency dataset in reference [8].

Variable	Model	MAE	MSE	MAPE
Heating load	IRLS	2.14 ± 0.24	9.87 ± 2.41	10.09 ± 1.01
	RF	0.51 ± 0.11	1.03 ± 0.54	2.18 ± 0.64
Cooling load	IRLS	2.21 ± 0.28	11.46 ± 3.63	9.41 ± 0.80
	RF	1.42 ± 0.25	6.59 ± 1.56	4.62 ± 0.70

MAE: Mean Absolute Error; MSE: Mean Square Error; MAPE: Mean Absolute Percentage Error.

Given the huge development of research references based on this dataset, the median and minimum value for each applied algorithm of machine learning is shown on Table 4, where the bold text represents the best case of each quality parameter.

Table 4. Median and minimum MAE and MAPE for different Machine Learning techniques. Bold value represent the smallest value of each model's quality parameter.

Model Type	Heating Load		Cooling Load	
	MAE (kW)	MAPE (%)	MAE	MAPE
LR	2.09(0.122)	21.43(3.9)	2.27(1.64)	6.94(4.6)
MLP	0.35(0.25)	1.62(1.20)	0.72(0.39)	2.34 (1.65)
ANN	0.56(0.085)	8.05(2.36)	1.47(0.57)	6.37(3.82)
k-NN	1.94(1.75)	-	1.88(1.17)	4.95(1.55)
SVR	0.56(0.236)	2.26(1.13)	1.76(0.59)	3.47(2.65)
RBFN	0.51(0.35)	21.5(2.76)	1.15(0.99)	-
RF and TE	0.51(0.19)	2.18(1.35)	1.3(0.41)	4.19(2.3)
ELM	0.19(0.018)	13.65(9.56)	0.98(0.24)	17.87(8.02)
GBRT	0.21(0.18)	-	0.4 (0.31)	-
MARS	0.53(0.08)	2.57(2.20)	1.12(0.15)	11.66(4.09)
GPR	0.35(0.18)	11.97(1.40)	0.93(0.45)	-
Hybrid	0.62(0.037)	2.28(1.56)	0.68(0.13)	3.1(2.45)
Genetic	0.83(0.38)	0.96(0.43)	0.97(0.52)	3.4(1.92)
Other	-	1.51(1.1)	-	-

3.1. Heating Load Models

3.1.1. Initial Dataset Variables

The variables in reference [8] have been used in previous studies, but some considerations can be done with respect them previously to their inclusion in the proposed model. A more detailed analysis will be done in Section 3.1.2. The relationships between the variables can be seen in Figure 1.

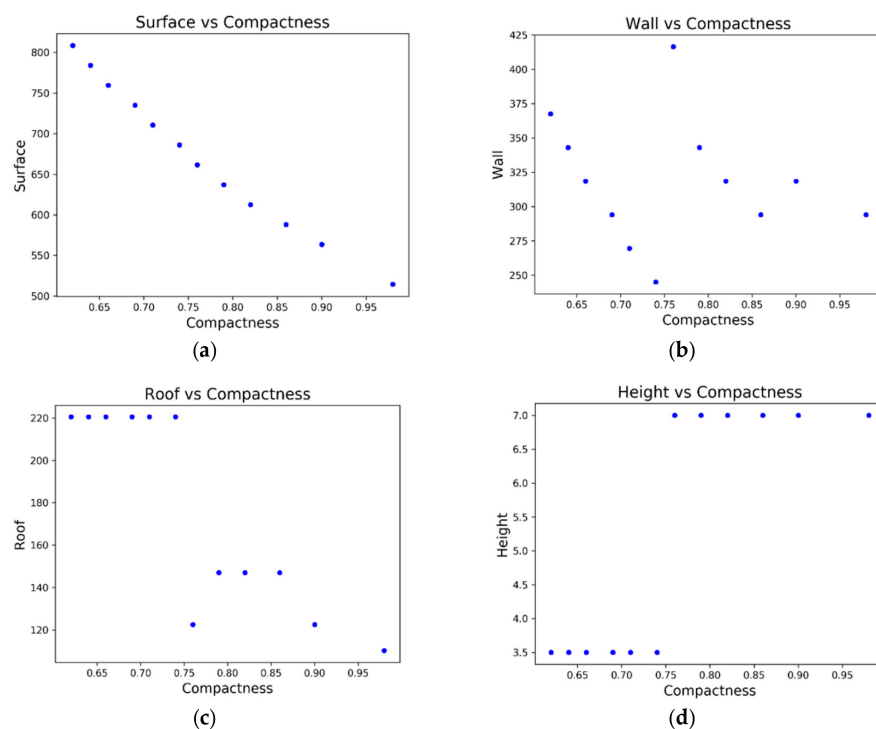


Figure 1. Cont.

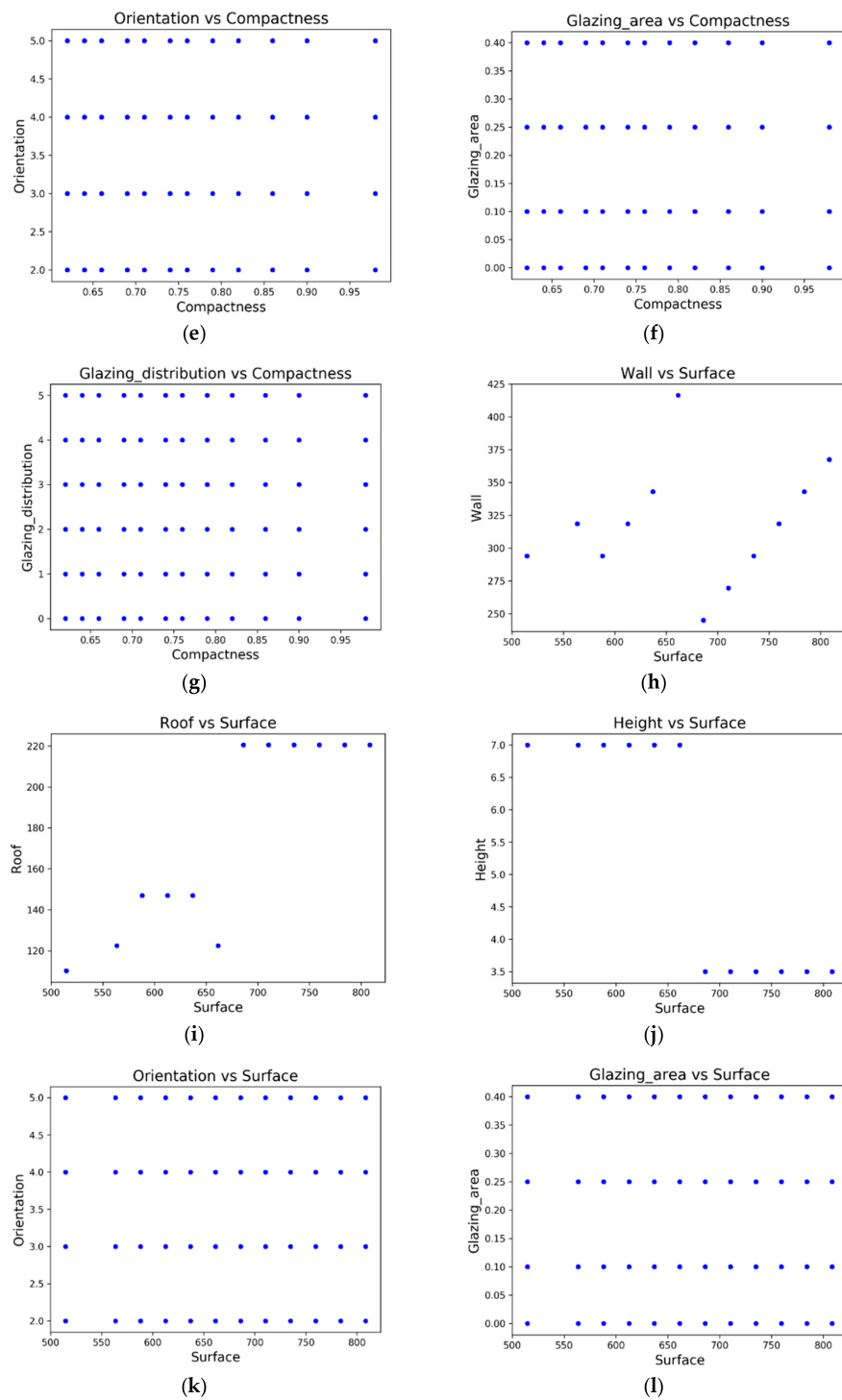


Figure 1. Cont.

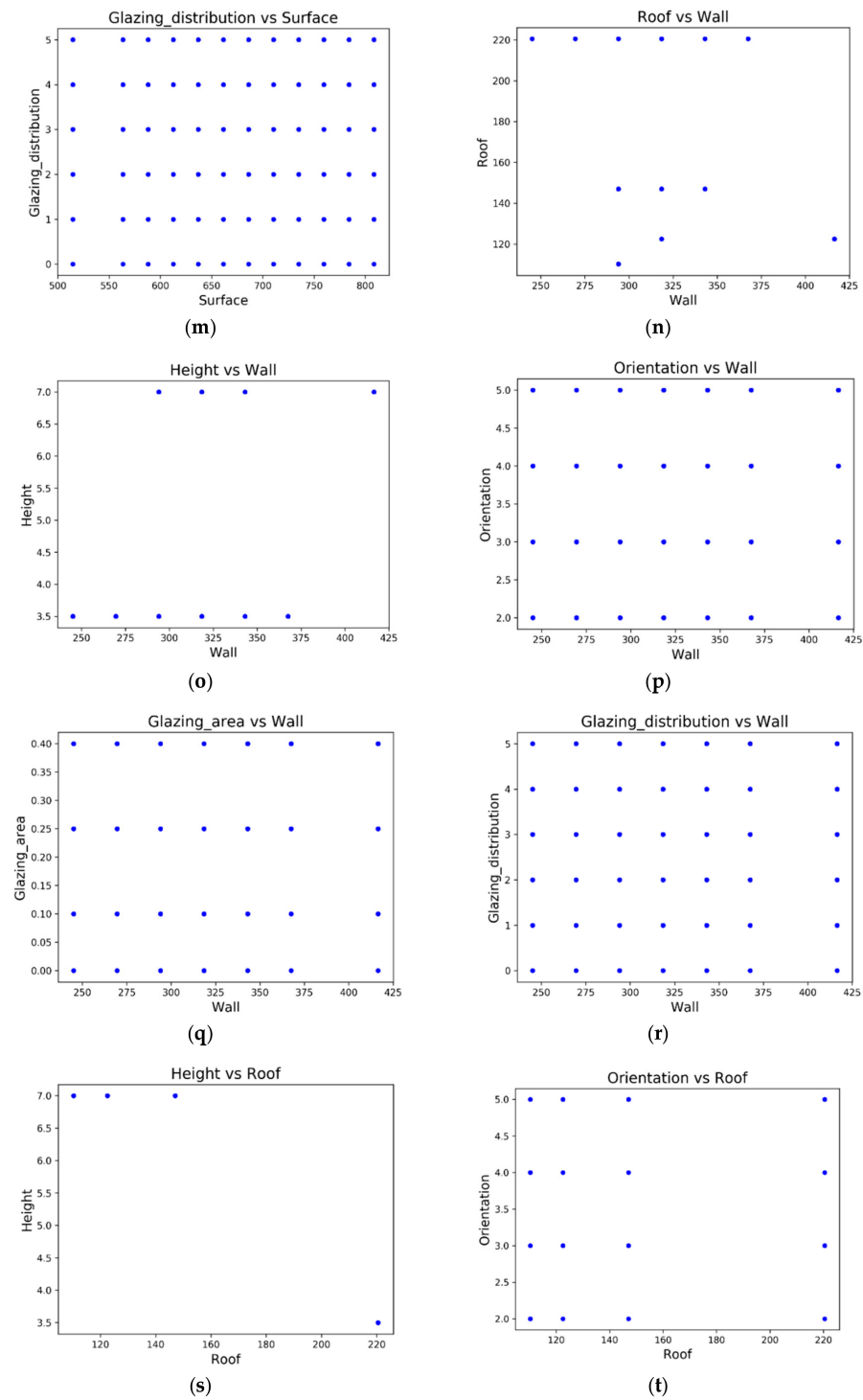


Figure 1. Cont.

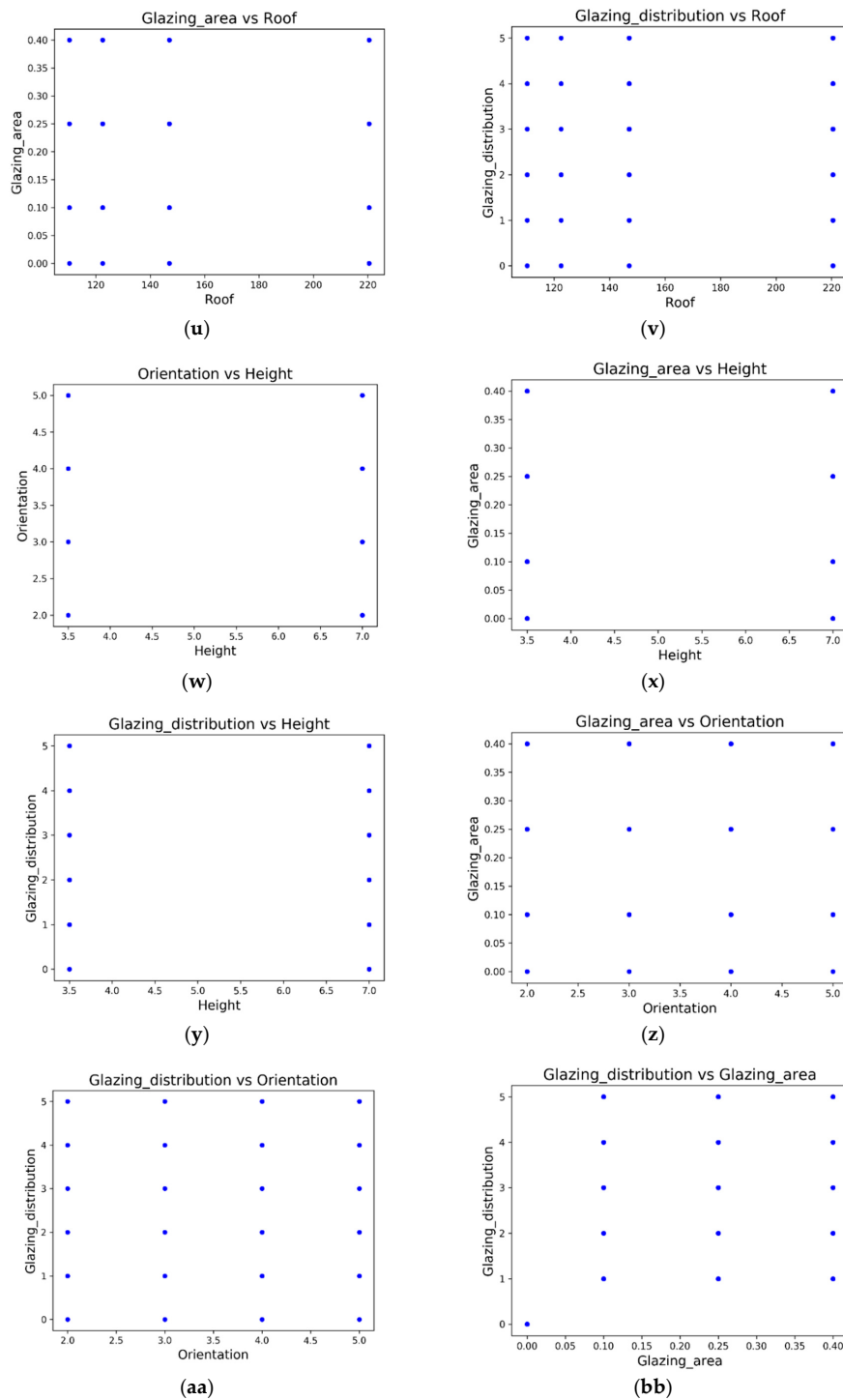


Figure 1. Relationship between the independent variables for the cooling load problem. (a) Surface vs. Compactness; (b) Wall vs. Compactness; (c) Roof vs. Compactness; (d) Height vs. Compactness; (e) Orientation vs. Compactness; (f) Glazing area vs. Compactness; (g) Glazing distribution vs. Compactness; (h) Wall vs. Surface; (i) Roof vs. Surface; (j) Height vs. Surface; (k) Orientation vs. Surface; (l) Glazing area vs. Surface; (m) Glazing distribution vs. Surface; (n) Roof vs. Wall; (o) Height vs. Wall; (p) Orientation vs. Wall; (q) Glazing area vs. wall; (r) Glazing distribution vs. Wall; (s) Height vs. Roof; (t) Orientation vs. Roof; (u) Glazing area vs. Roof; (v) Glazing distribution vs. Roof; (w) Orientation vs. Height; (x) Glazing area vs. Height; (y) Glazing distribution vs. Height; (z) Glazing area vs. Orientation; (aa) Glazing distribution vs. Orientation and (bb) Glazing distribution vs. Glazing area.

With respect to the dependent variable, the relationship between the heating load and the independent descriptors can be seen in Figure 2.

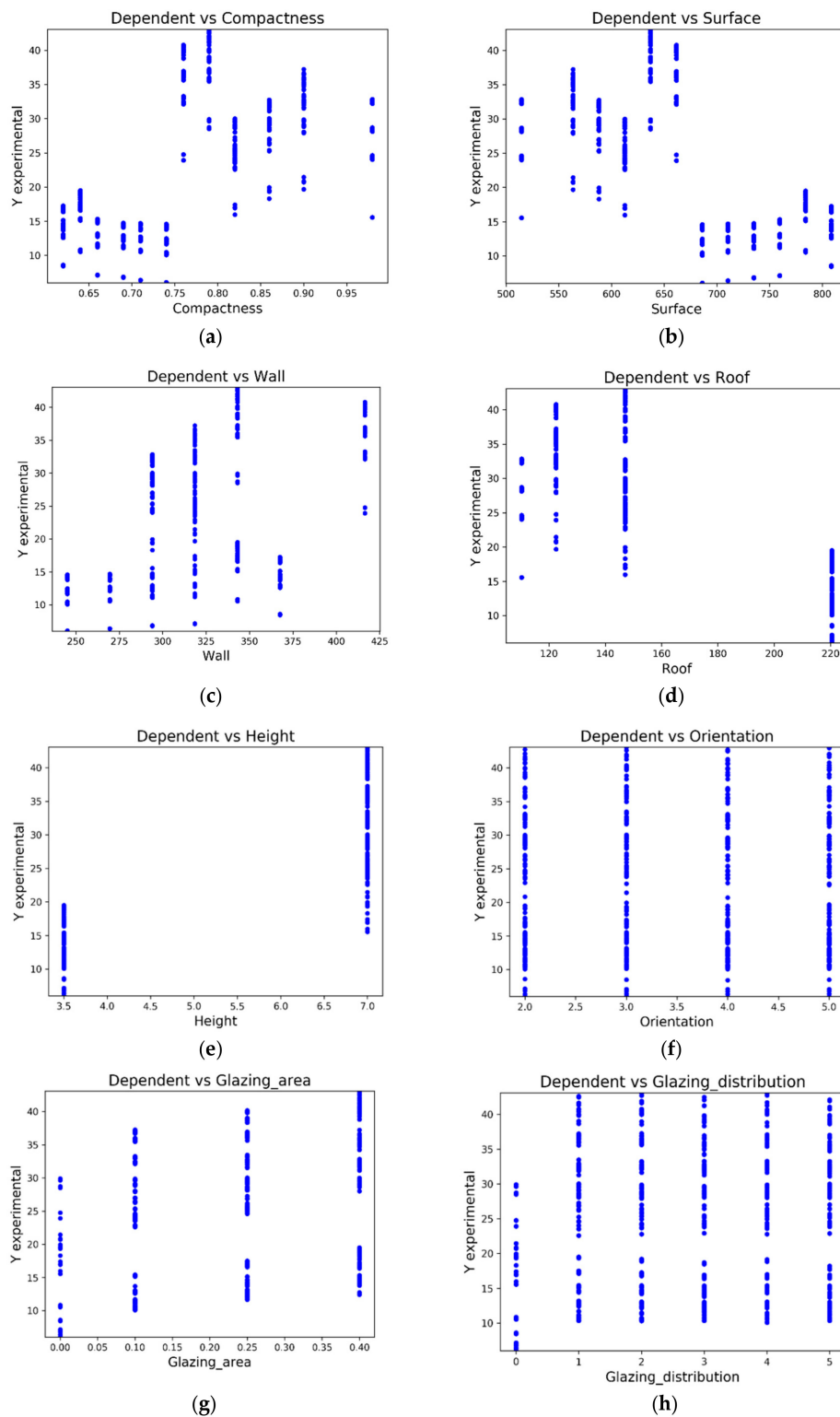


Figure 2. Heating load values depending on (a) relative compactness; (b) surface area; (c) wall area; (d) roof area; (e) overall height; (f) orientation; (g) glazing area; (h) glazing area distribution.

An analysis of the data using octahedric regression using different values of h and ω parameters gives a result that can be seen in Figure 3:

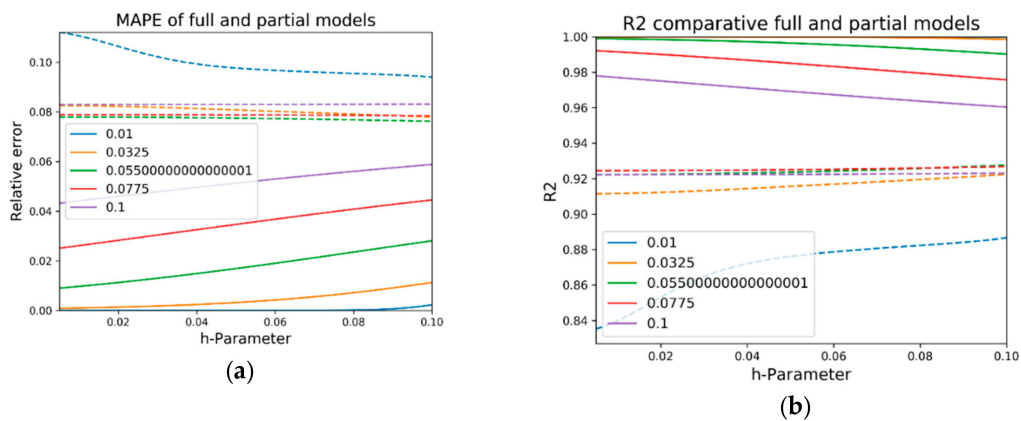


Figure 3. Results for the heating load model depending on the h -parameter. The values of the ω parameter are shown in the legend. Continuous lines represent full models, while discontinuous corresponds to restricted or partial models: (a) MAPE; (b) R2.

The selection of parameters h and d ω can be done from the results shown in Figure 3. However, the total process can be accelerated diminishing the number of models to evaluate using a simple relationship between the parameters ω and h , as $\omega = h$. This model is called normalized octahedric regression and increases the errors in a small magnitude from the best case of free selection for ω and h , but the model selection is easier, as can be seen in Figure 4.

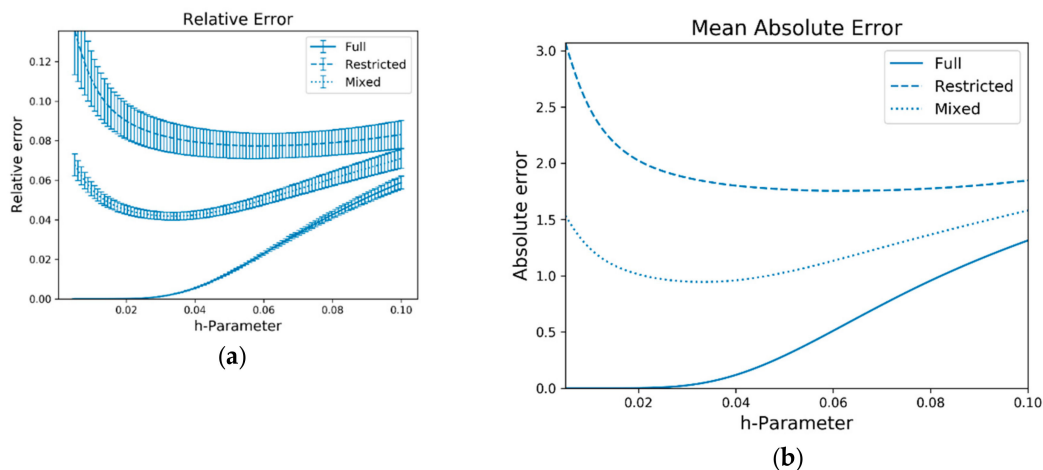


Figure 4. Results of the heating load problem for the full, partial, and mixed models with the parameter selection $h = \omega$. (a) Mean absolute percentage error (MAPE) depending on h ; (b) Mean absolute error (MAE) depending on h .

As was previously observed from Equation (61), full model tends to overfitting when $h \rightarrow 0$, while restricted models tend to overestimate the stochastic errors. For these reasons, a good indicator of the model behavior would be the mixed model, defined as

$$mixed(x) = \frac{1}{2} \cdot (full(x) + restricted(x)), \quad (62)$$

Following the behavior of the mixed model in Figure 4, the selection of $\omega = h = 0.033$ seems a logical option, corresponding to a pseudo-complexity of 30. Figure 5 shows the estimated versus the experimental values.

Figure 6 shows the error over each independent variable. The REC curve for the model is shown in Figure 7.

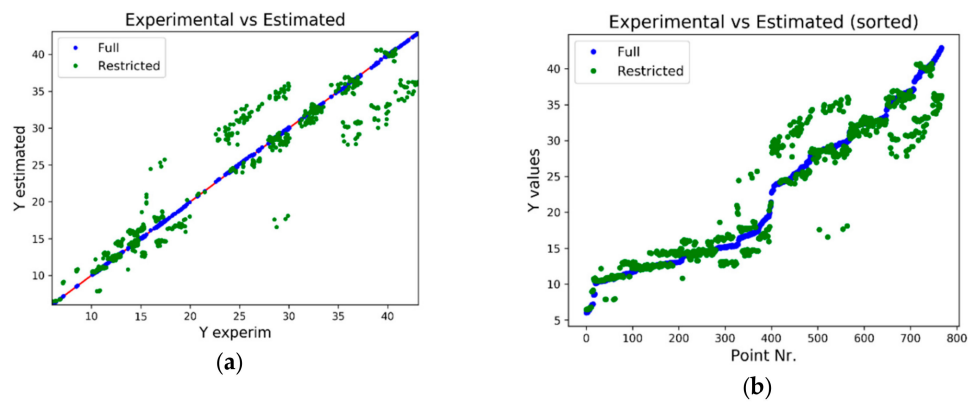


Figure 5. Results of the heating load problem for the full and restricted models with the parameter selection $h = \omega = 0.033$. (a) Estimated vs experimental values; (b) Estimated vs experimental values (sorted).

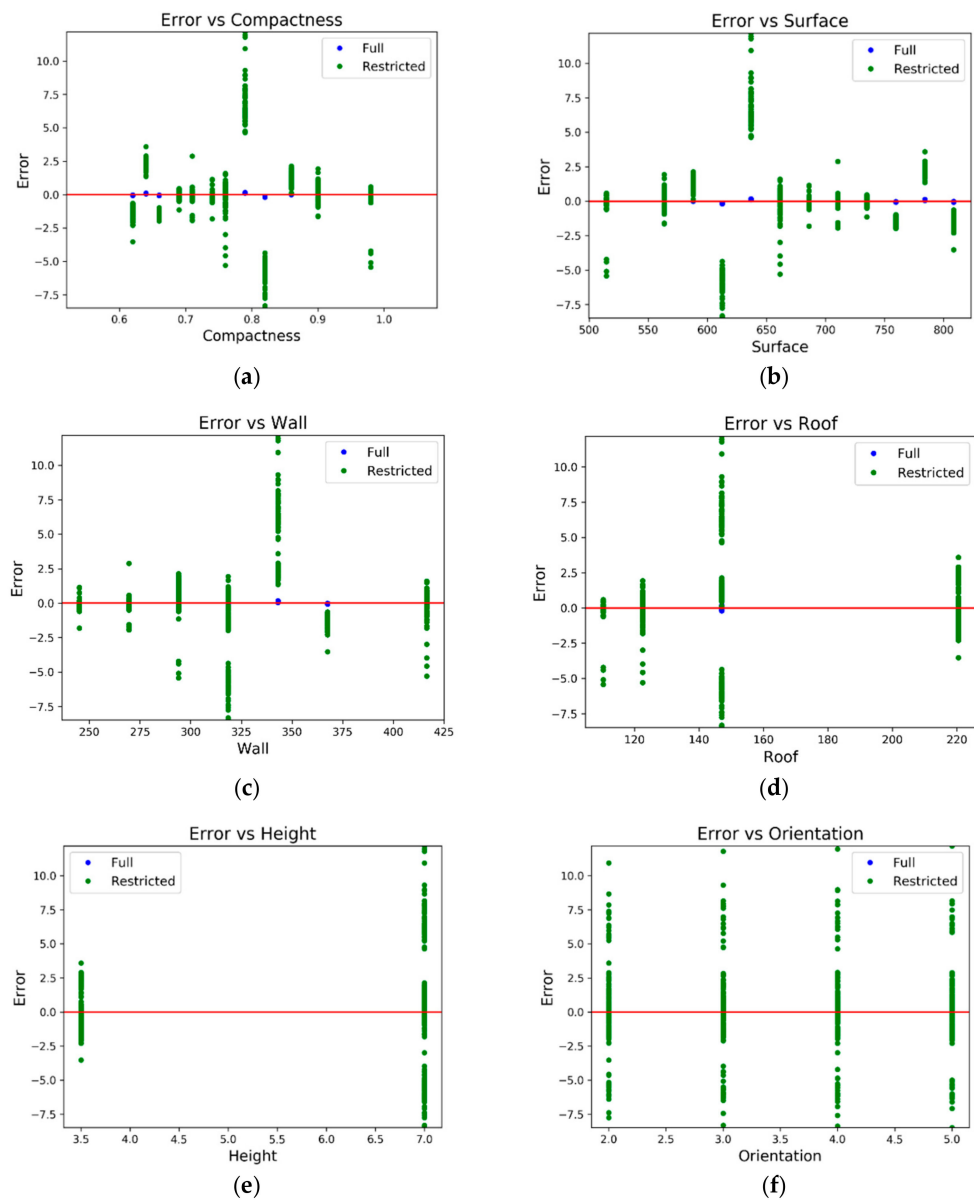


Figure 6. Cont.

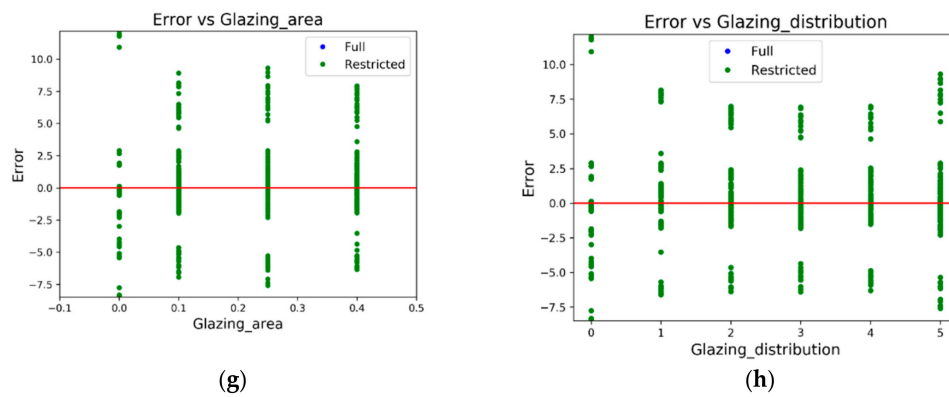


Figure 6. Errors over independent variables. (a) Relative compactness; (b) Surface area; (c) Wall area; (d) Roof area; (e) Overall height; (f) Orientation; (g) Glazing area; (h) Glazing area distribution.

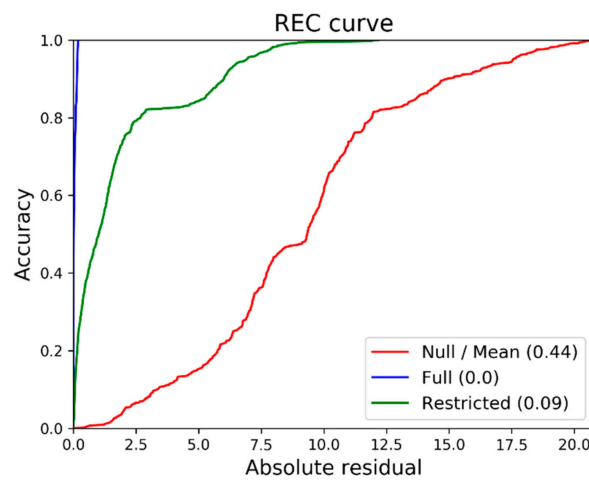


Figure 7. Regression error characteristic (REC) curve for the heating load full restricted and null (Mean) models with $h = \omega = 0.033$.

3.1.2. Reduced Models—Separated Models by Number of Floors

A detailed revision of the model obtained in Section 3.1.1 shows two points that should be considered more in detail. First, the plots of Figure 1a–d,h–j,n,o,s,w–y show some kind of relationship between the involved variables. Considering that the vertical interfaces of the buildings are of floor or roof type (and given that the build unit is a cube, both variables must have the same value, resumed by *Floor*), the total surface can be written as

$$S = Wall + 2 \cdot Floor, \quad (63)$$

Also, given the definition of the relative compactness defined as in reference [60] taking the cube as elemental compactness measure,

$$RC = 6 \cdot \frac{V^{\frac{2}{3}}}{S}, \quad (64)$$

Equation (63) and Equation (64) introduce two constraints that jointly to the constant volume condition imposed by the algorithm of design of the buildings, reduce in two the independent or basic variables related with the building's geometry to be considered. In this study, these two variables will be the wall and the floor areas. Moreover, the overall height can be eliminated from the variable set because, given that all the basic cubic elements will have one lower side, if floor surface equals $18 \cdot 12.25 = 220.5 \text{ m}^2$, overall height must be 3.5 m.

Selection of variables have been done, by example, in [28], where different models are compared for the sets of variables (roof area, overall height), (relative compactness, roof area, overall height), (relative compactness, surface area, wall area, roof area, overall height, glazing area) and the full dataset. In reference [18], the considered variables are (surface area, wall area, roof area, overall height, glazing area), while reference [35] uses (relative compactness, surface area, wall area, overall height, glazing area) to model the cooling problem. The importance of each variable is studied in [9] using ANOVA with the result of (relative compactness, surface area, wall area, overall height, glazing area) and (relative compactness, wall area, roof area, overall height, glazing area) as the most important variables for the heating and cooling problems, respectively.

However, any extension to a more general case should consider if any of the constraints remain or can be ignored, introducing additional independent descriptors.

So, the reduced set of variables considered in the modelling with the new methodology are those presented in Table 5.

Table 5. Input and output variables considered for the modelling of the energy efficiency problem with octahedric regression.

Input Variables	Output Variables
Wall area	Heating load
Roof area	Cooling load
Orientation	-
Glazing area	-
Glazing area distribution	-

A second detail to consider is the results of Figure 5, especially the plot (e), which shows different behavior for the two values of the height, that physically would correspond to buildings with one and two floors.

Moreover, from Figure 2, plots (d) and (e) show two groups for the values of the dependent variable, being this behaviour more remarkable in the case of the dependent variable height, corresponding to buildings with one and two floors.

So, a separate study of two different models called 1_floor and 2_floors defined by the number of floors will be carried out (Table 6). In the case of one floor buildings, the roof area corresponds to 220.5 m^2 , as was commented previously, so this variable can also be omitted.

Table 6. Input variables for the 1_floor and 2_floors models of the heating load problem.

1_floor	2_floors
Wall area	Wall area
Orientation	Roof area
Glazing area	Orientation
Glazing area distribution	Glazing area
-	Glazing area distribution

The results for the 1_floor normalised model are shown in Figure 8, for different values of h-parameters.

The different behaviors of the relative error and mean absolute error (MAE) recommends the use of an intermediate value $h = 0.02$ for the selected model, whose results are shown on Figure 9.

The 2_floors model results can be seen in Figure 10.

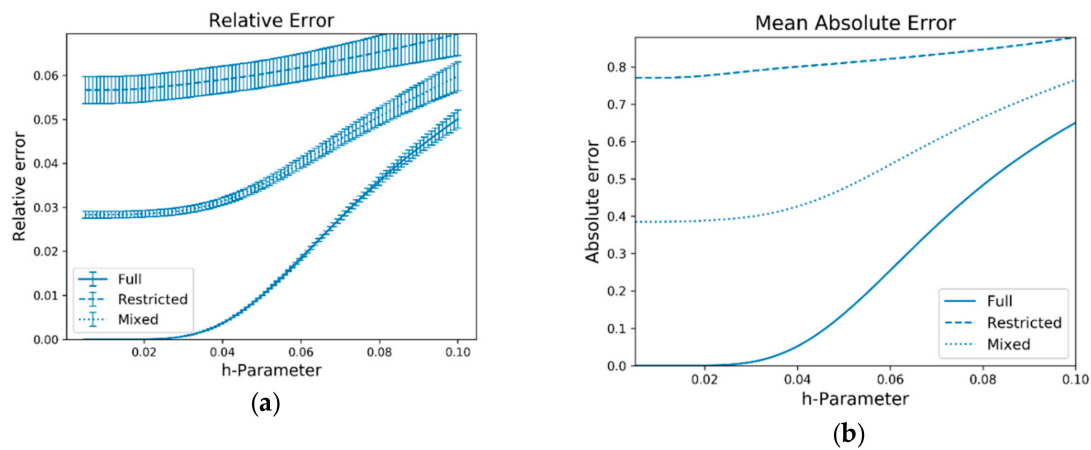


Figure 8. Results of the heating load 1_floor problem for the full, restricted and mixed models with the parameter selection $h = \omega$. (a) MAPE depending on h ; (b) MAE depending h .

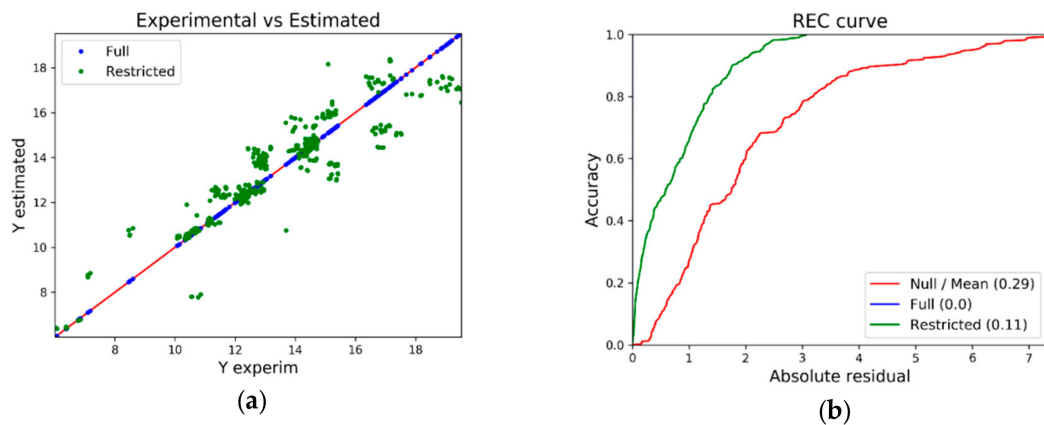


Figure 9. Results of the Heating load 1_floor problem for the full and restricted models with the parameter selection $h = \omega = 0.02$. (a) Estimated vs experimental values; (b) REC curve for the full, restricted and null (mean) models.

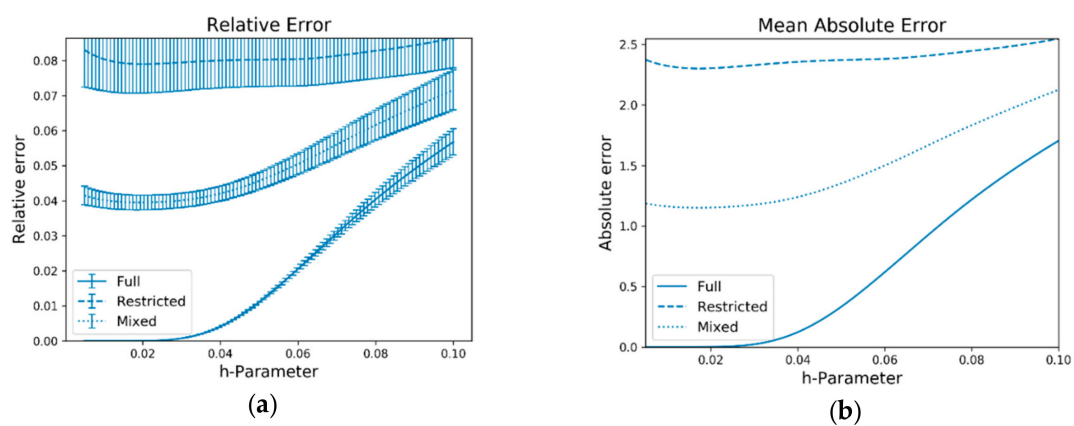


Figure 10. Results of the Heating load 2_floors problem for the full, restricted and mixed models with the parameter selection $h = \omega$. (a) MAPE depending on h ; (b) MAE depending on h .

A selection of $h = 0.02$ gives as results the models shown in Figure 11:

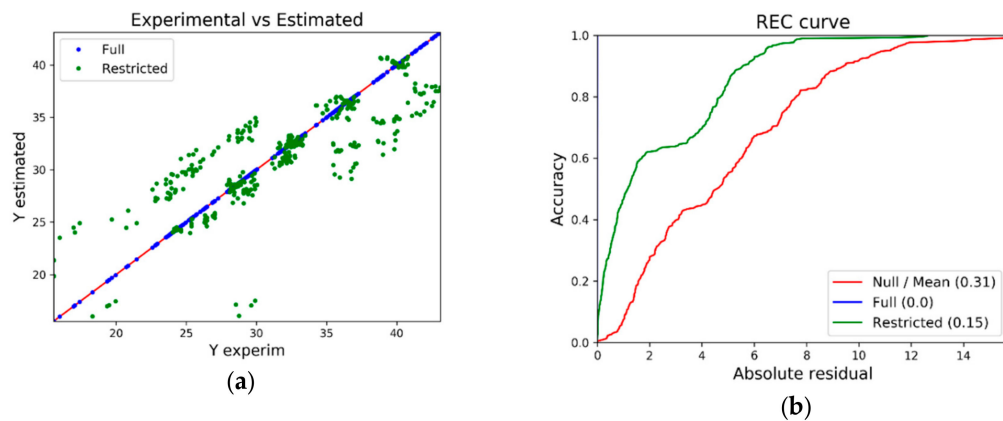


Figure 11. Results of the Heating load 2_floors problem for the full and restricted models with the parameter selection $h = \omega = 0.02$. (a) Estimated vs experimental values; (b) REC curve for the full, restricted, and null (mean) models.

To compare with the join dataset introduced in Table 5, a group of models has been generated to see their behaviors.

A selection of $h=0.04$ gives the minimum error value for the models represented in Figure 12. The summary of the resulting model is shown in Figure 13.

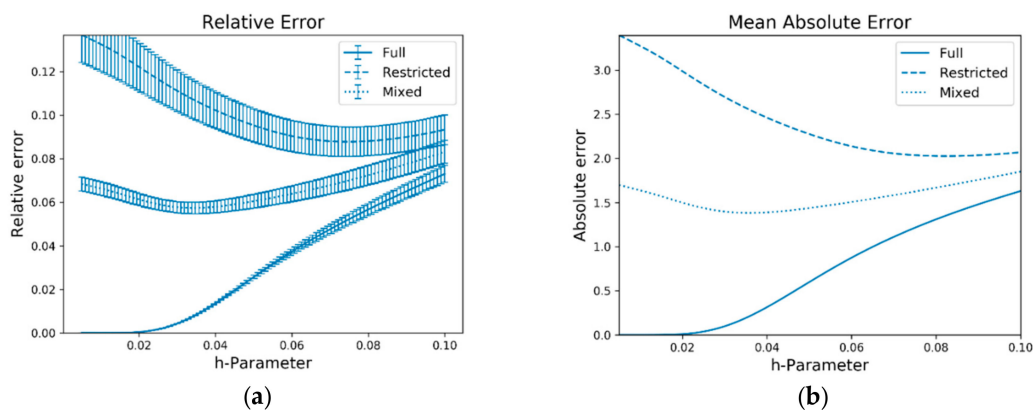


Figure 12. Results of the Heating load (reduced variable set) problem for the full, restricted and mixed models with the parameter selection $h = \omega$. (a) MAPE depending on h ; (b) MAE depending on h .

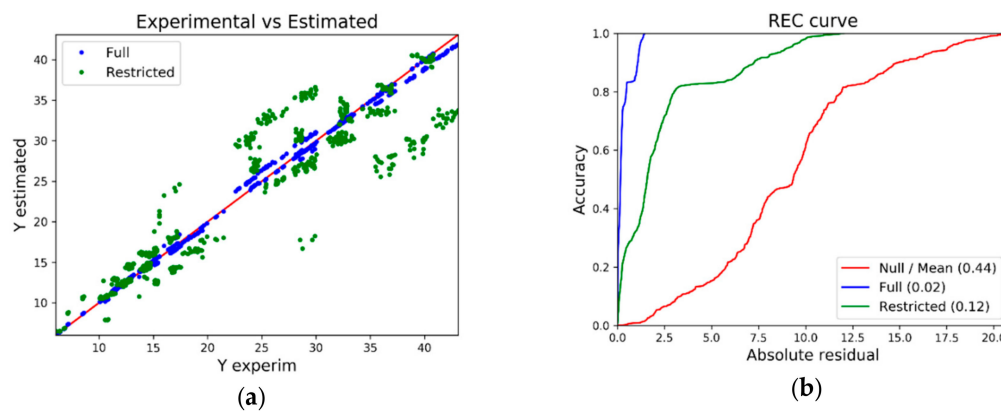


Figure 13. Results of the heating load (reduced variable set) problem for the full and restricted models with the parameter selection $h = \omega = 0.04$. (a) Estimated vs experimental values; (b) REC curve for the full, restricted, and null (mean) models.

A resume of the Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) obtained by the octahedric regression models for the heating load model is shown in Table 7.

Table 7. Quality parameters for the models of the heating load problem.

Parameter	All Variables	Reduced Variables	1_floor	2_floors
Number Points	768	768	384	384
MAE	0.945 (1.847)	1.382 (2.556)	0.388 (0.777)	1.152 (2.304)
MSE	2.289 (8.773)	4.293 (14.644)	0.282 (1.125)	2.803 (11.202)
MAPE	4.182% (8.163%)	5.744% (10.587%)	2.855% (5.707%)	3.953% (7.903%)

3.2. Cooling Load Models

A similar study to the heating load case can be done for the cooling load problem.

3.2.1. Complete and Reduced Datasets

Including the constraints introduced by Equations (63) and (64), the reduced variable model as introduced in Table 5 is applied to the cooling load problem.

The models for the complete and reduced variable datasets are shown on Figure 14.

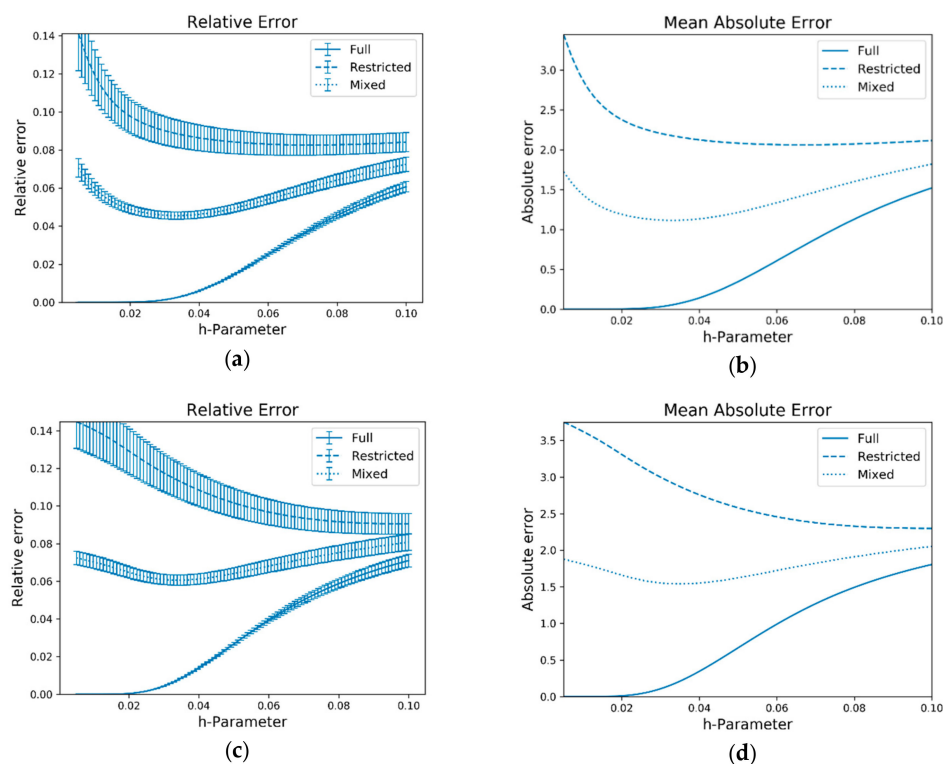


Figure 14. Results of the Cooling load problem for the complete and reduced variable dataset models with the parameter selection $h = \omega$. (a) MAPE depending on h for the model with 8 variables; (b) MAE depending on h for the model with 8 variables; (c) MAPE depending on h for the model with 5 variables; (d) MAE depending on h for the model with 5 variables.

The results are very similar, but comparing Figure 14a,b, it seems a good selection of $h = 0.033$ for the complete dataset model, whereas Figure 14c,d seems to recommend a value $h = 0.033$ for the reduced variable dataset. The results for both models are shown in Figure 15.

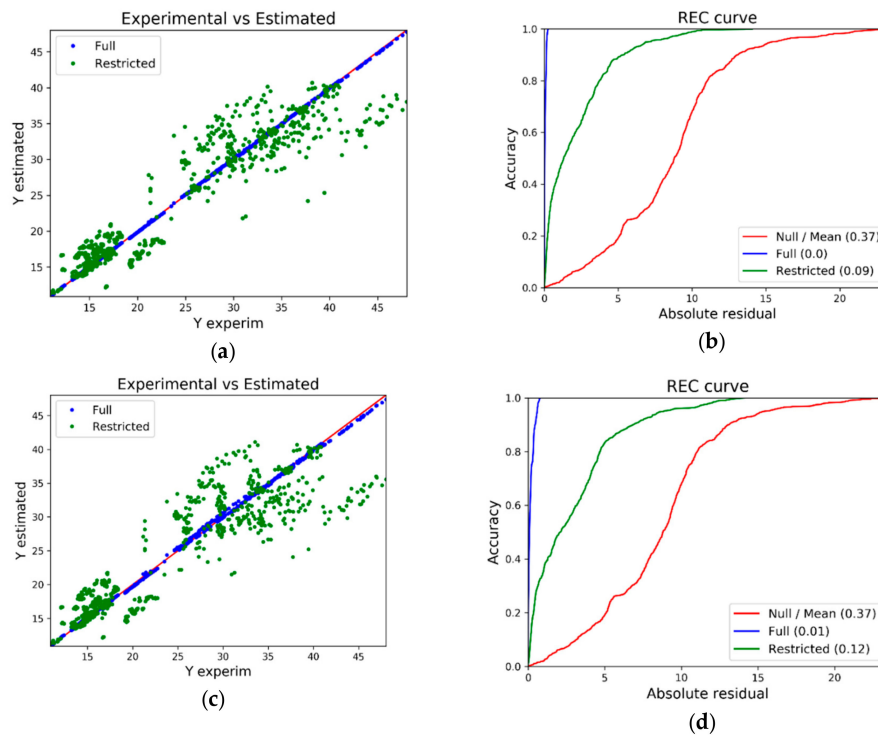


Figure 15. Results of the cooling load problem (a) Estimated vs experimental values for the full and restricted models for the eight-variable dataset with parameter selection $h = \omega = 0.033$; (b) REC curve for the full, restricted and null (mean) models for the full and restricted models for the eight-variable dataset with parameter selection $h = \omega = 0.033$; (c) Estimated vs experimental values for the model with five variables and $h = \omega = 0.033$; (d) REC curve for the model with five variables and $h = \omega = 0.033$.

3.2.2. Separated Models by Number of Floors

An analysis of errors for the complete problem is shown in Figure 16.

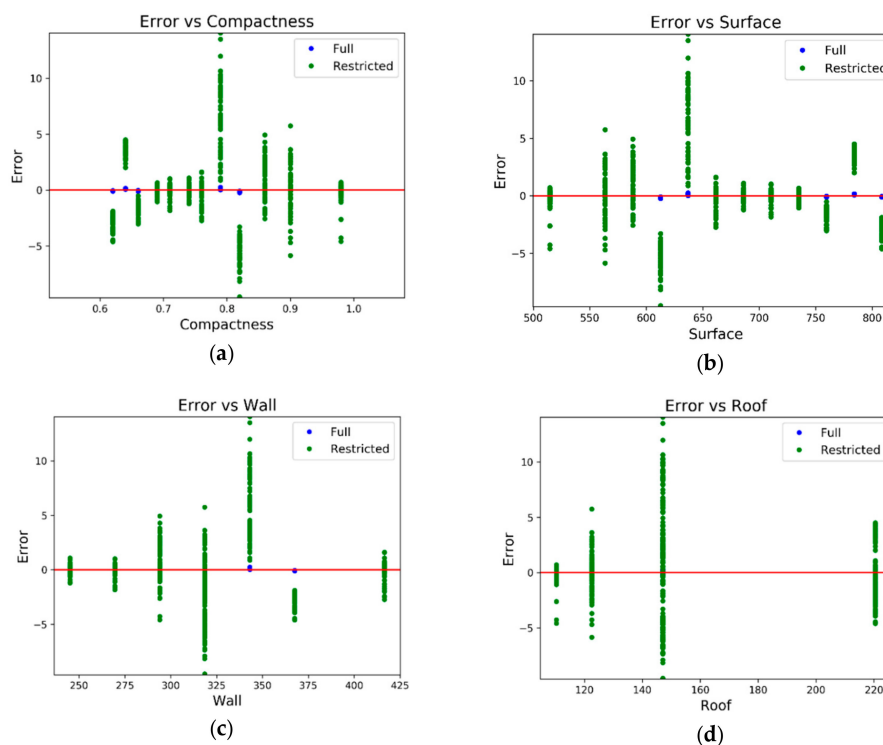


Figure 16. Cont.

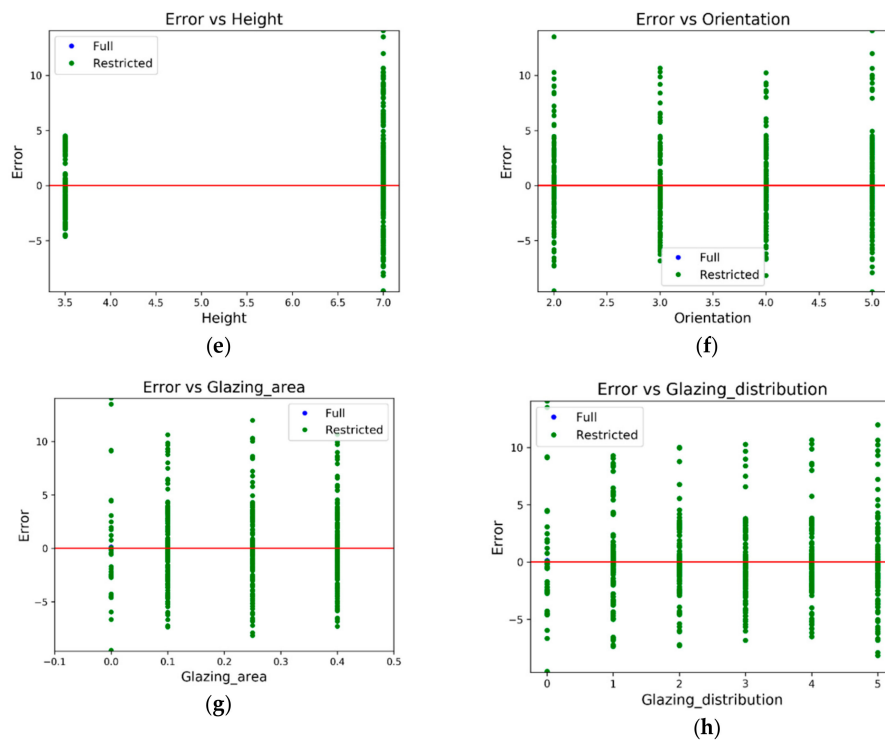


Figure 16. Errors over independent variables for the cooling problem including all the variables. (a) Relative compactness; (b) Surface area; (c) Wall area; (d) Roof area; (e) Overall height; (f) Orientation; (g) Glazing area; (h) Glazing area distribution.

Following the reasoning in Section 3.1.2, the models defined in Table 6 can be studied for the cooling load case, as it is shown in Figure 17:

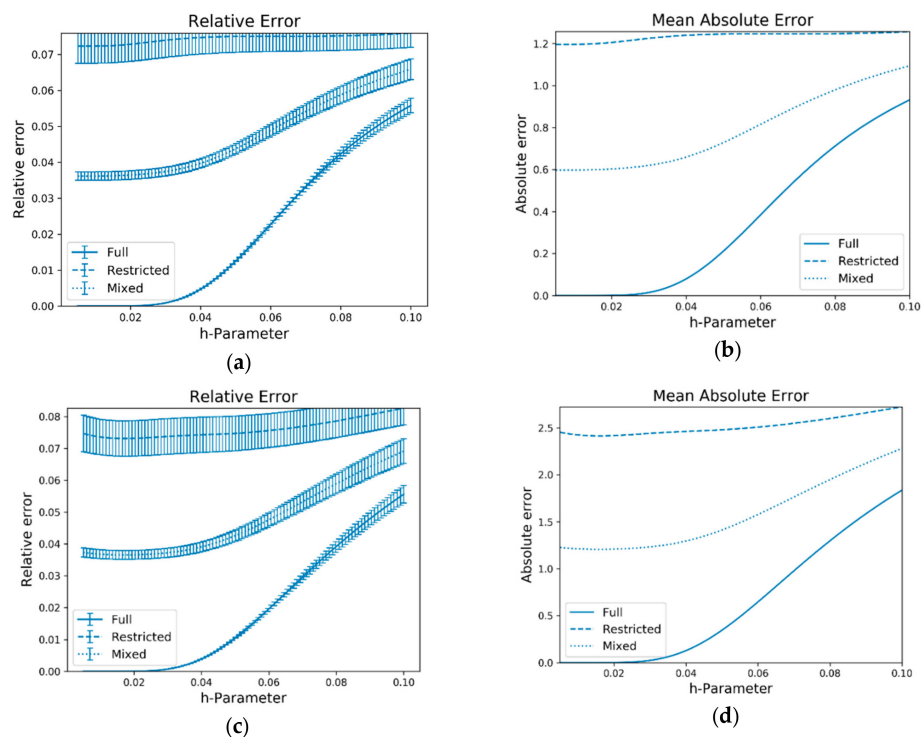


Figure 17. Results of the cooling load problem for the complete and reduced variable dataset models with the parameter selection $h = \omega$. (a) MAPE depending on h for the model one-floor model with four variables; (b) MAE depending on h for the one-floor model; (c) MAPE depending on h for the two-floors model with five variables; (d) MAE depending on h for the two-floors model.

Selection of optimum values of h parameters could be $h = 0.02$ in both cases, obtaining the following corresponding models shown in Figure 18:

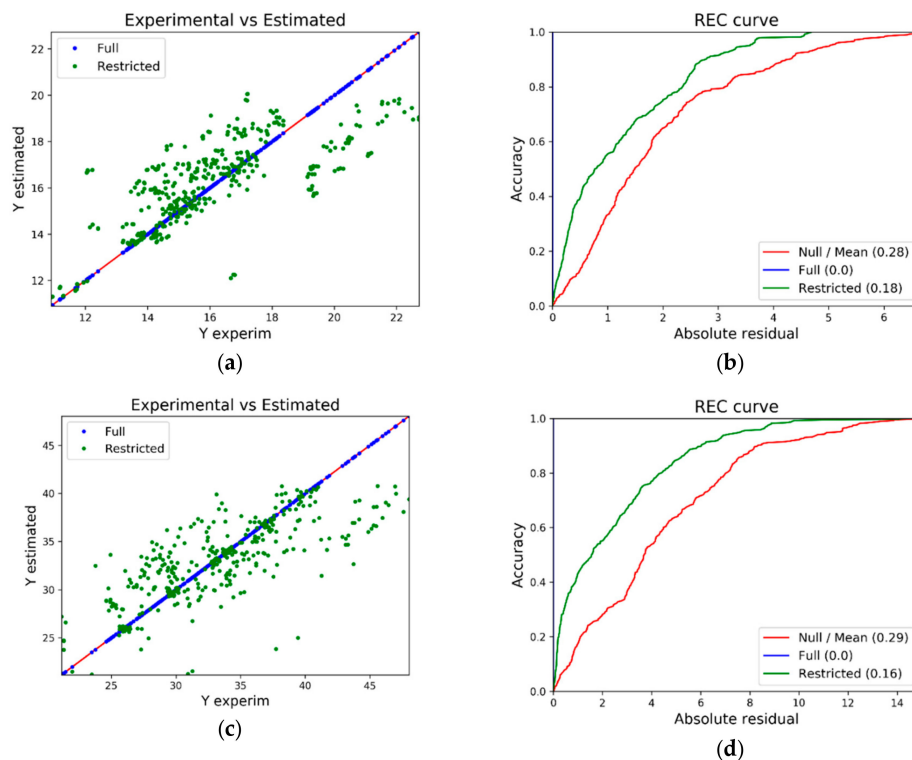


Figure 18. Results of the cooling load reduced problem for the full and restricted models with the parameter selection $h = \omega = 0.02$. (a) Estimated vs experimental values for 1_floor model with five variables; (b) REC curve for the full, restricted and null (mean) 1_floor model; (c) Estimated vs experimental values for 2_floors model; (d) REC curve for the full, restricted, and null (mean) 2_floors model.

A resume of the available models for the cooling load problem can be seen at Table 8.

Table 8. Quality parameters for the mixed and restricted models of the cooling load problem.

Parameter	All Variables	Reduced Variables	1_floor	2_floors
Number Points	768	768	384	384
MAE	1.113 (2.175)	1.538 (2.892)	0.603 (1.206)	1.210 (2.419)
MSE	2.731 (10.453)	4.774 (16.859)	0.688 (2.748)	3.109 (12.426)
MAPE	4.554% (8.885%)	6.058% (11.360%)	3.646% (7.289%)	3.661% (7.320%)

4. Discussion

The proposed methodology has been compared with other machine learning methods applied to the estimation of energy loads in building from dataset in reference [8]. In the heating load problem, the results for the complete datasets are discrete, in the complete and reduced variables cases. For the separated study of building depending on its floors, the result for one floor is between the median value for the different ML techniques (see Table 4). However, the results for the two floors buildings are again discrete. The results are, in general better for MAPE than for MAE indicators.

In the case of the cooling load problem, the results are lightly worse, something that also happens in most of the ML methods presented in Table 4. However, this worsening is relatively minor, obtaining in general results in order with the median of the techniques.

The effect of the additional variables on the results can be seen on Tables 7 and 8, where the quality parameters are better in all the categories for the complete against the reduced models. Using variables correlated between them can deteriorate the performance of ML methods and detecting and treating the associated overfitting is a difficult problem. The proposed methodology shows a relative stability front the effect of these spurious predictors.

Also, an interesting characteristic of the method is the capability of estimating the overfitting at cheap computational cost through the full and restricted models. A compromise between both values is taken as the output of the predictive process. However, more in-depth research must be done in order to adequately evaluate the behaviour and fitness of the octahedric regression technique. For example, the characteristics of the computational algorithm make it an ideal candidate for the use of parallel computing versions. Measurements of the speedups obtained using some benchmarking datasets will be one of the future developments.

Another promising characteristic of the methodology is that it is derived of the geometrical properties of the octahedral. The spatial distributions of the symmetry axes allow studying the effect of different orientations on the partial calculation of the estimation. The way this property can be implemented to select the principal components of the dataset is another field of research.

With respect to the problem of energy efficiency of buildings, two main lines of study open from this paper. First, the existence of different models far from the considered number of floors could be deduced from the behaviour of the error in Figures 5 and 13. The origin and causes of this different behaviour should be studied in more depth. Related with this problem, the second aspect to consider is the existence of additional variables not related with those in Table 5 that could affect the way the software Ecotect calculates the heating and cooling loads and is not considered in the present study.

Author Contributions: Some of the previous modelling methodologies included in this paper were obtained in the PhD thesis carried out by F.J.N.-G. at the University of Alicante (Spain), under the supervision of Y.V. All the authors wrote the paper. All authors contributed to conceiving and designing the research and to analyzing and discussing the results.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Baldi, S.; Yuan, S.; Endel, P.; Holub, O. Dual estimation: Constructing building energy models from data sampled at low rate. *Appl. Energy* **2016**, *169*, 81–92. [CrossRef]
2. Wang, Z.; Srinivasan, R.S.; Shi, J. Artificial intelligent models for improved prediction of residential space heating. *J. Energy Eng.* **2016**, *142*, 04016006. [CrossRef]
3. Lü, X.; Lu, T.; Kibert, C.J.; Viljanen, M. Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. *Appl. Energy* **2015**, *144*, 261–275. [CrossRef]
4. Chou, J.S.; Pham, A.D. Smart artificial firefly colony algorithm-based support vector regression for enhanced forecasting in civil engineering. *Comput. Aided Civ. Infrastruct. Eng.* **2015**, *30*, 715–732. [CrossRef]
5. Meng, Z.; Zhang, P.J. A Simplified and Scalable Heat-Flow Based Approach for Optimizing the Form, Massing and Orientation for High Performance Building Design. Available online: https://www.researchgate.net/profile/Zhaozhou_Meng/publication/323639680_A_simplified_and_scalable_heat-flow_based_approach_for_optimizing_the_form_massing_and_orientation_for_high_performance_building_design/links/5aa145c00f7e9badd9a42f3a/A-simplified-and-scalable-heat-flow-based-approach-for-optimizing-the-form-massing-and-orientation-for-high-performance-building-design.pdf (accessed on 10 October 2019).
6. Westermann, P.; Evins, R. Surrogate modelling for sustainable building design-A review. *Energy Build.* **2019**, *198*, 170–186. [CrossRef]
7. Dua, D.; Graff, C. *UCI Machine Learning Repository*; School of Information and Computer Science, University of California: Irvine, CA, USA, 2019.

8. Tsanas, A.; Xifara, A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy Build.* **2012**, *49*, 560–567. [[CrossRef](#)]
9. Sholahudin, S.S.S.; Alam, A.G.; Baek, C.I.B.C.I.; Han, H. Prediction and analysis of building energy efficiency using artificial neural networks and design of experiments. *J. Mek.* **2014**, *37*, 37–41. [[CrossRef](#)]
10. Roy, S.S.; Roy, R.; Balas, V.E. Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renew. Sustain. Energy Rev.* **2018**, *82*, 4256–4268.
11. Chou, J.S.; Bui, D.K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy Build.* **2014**, *82*, 437–446. [[CrossRef](#)]
12. Pietrzykowski, M. Application of the Mini-Models Based on n-Dimensional Simplex for Modeling of Buildings Energy Performance. *Int. J. Comput. Technol. Appl.* **2015**, *6*, 695–700.
13. Dan, T.X.; Phuc, P.N.K. Application of Machine Learning in Forecasting Energy Usage of Building Design. In Proceedings of the 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, Vietnam, 23–24 November 2018; pp. 53–59.
14. Jitkongchuen, D.; Pacharawongsakda, E. Prediction Heating and Cooling Loads of Building Using Evolutionary Grey Wolf Algorithms. In Proceedings of the 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), Nan, Thailand, 30 January–2 February 2019; pp. 93–97.
15. Alam, A.G.; Baek, C.I.; Han, H. Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments. In *Applied Mechanics and Materials*; Trans Tech Publications: Baech, Switzerland, 2016; Volume 819, pp. 541–545.
16. Nwulu, N.I. An artificial neural network model for predicting building heating and cooling loads. In Proceedings of the 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 16–17 September 2017; pp. 1–5.
17. Tien Bui, D.; Moayedi, H.; Anastasios, D.; Kok Foong, L. Predicting Heating and Cooling Loads in Energy-Efficient Buildings Using Two Hybrid Intelligent Models. *Appl. Sci.* **2019**, *9*, 3543. [[CrossRef](#)]
18. Le, L.T.; Nguyen, H.; Dou, J.; Zhou, J. A Comparative Study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in Estimating the Heating Load of Buildings' Energy Efficiency for Smart City Planning. *Appl. Sci.* **2019**, *9*, 2630. [[CrossRef](#)]
19. Cheng, M.Y.; Cao, M.T. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Appl. Soft Comput.* **2014**, *22*, 178–188. [[CrossRef](#)]
20. Duarte, G.R.; Fonseca, L.G.D.; Goliatt, P.V.Z.C.; Lemonge, A.C.D.C. Comparison of machine learning techniques for predicting energy loads in buildings. *Ambiente Construído* **2017**, *17*, 103–115. [[CrossRef](#)]
21. Reddy, A.V.R.; Kumar, M.S. A Comparative Analysis of Regression Algorithms for Energy Estimation in Residential Buildings. In *International Conference on Intelligent Computing and Communication Technologies*; Springer: Singapore, 2019; pp. 300–311.
22. Le, L.T.; Nguyen, H.; Zhou, J.; Dou, J.; Moayedi, H. Estimating the heating load of buildings for smart city planning using a Novel Artificial Intelligence Technique PSO-XGBoost. *Appl. Sci.* **2019**, *9*, 2714. [[CrossRef](#)]
23. Yang, L.; Liu, S.; Tsoka, S.; Papageorgiou, L.G. Mathematical programming for piecewise linear regression analysis. *Expert Syst. Appl.* **2016**, *44*, 156–167. [[CrossRef](#)]
24. Papadopoulos, S.; Azar, E.; Woon, W.L.; Kontokosta, C.E. Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *J. Build. Perform. Simul.* **2018**, *11*, 322–332. [[CrossRef](#)]
25. Seyedzadeh, S.; Rahimian, F.P.; Rastogi, P.; Glesk, I. Tuning machine learning models for prediction of building energy loads. *Sustain. Cities Soc.* **2019**, *47*, 101484. [[CrossRef](#)]
26. Gupta, A.; Kohli, M.; Malhotra, N. Classification based on Data Envelopment Analysis and supervised learning: A case study on energy performance of residential buildings. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016; pp. 1–5.
27. Gao, W.; Alsarraf, J.; Moayedi, H.; Shahsavari, A.; Nguyen, H. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. *Appl. Soft Comput.* **2019**, *84*, 105748. [[CrossRef](#)]

28. Toprak, A.; Koklu, N.; Toprak, A.; Ozcan, R. Comparison of Classification Techniques on Energy Efficiency Dataset. *Int. J. Intell. Syst. Appl. Eng.* **2017**, *5*, 81–85. [[CrossRef](#)]
29. Manimaran, S.; AlBastaki, I.; Mangai, J.A. An ensemble model for predicting energy performance in residential buildings using data mining techniques. *ASHRAE Trans.* **2015**, *121*, 402–410.
30. Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling, W.L. Optimized parameter selection for assessing building energy efficiency. In Proceedings of the 7th IEEE Young Researchers Symposium in Electrical Power Engineering (YRS 2014), Ghent, Belgium, 24–25 April 2014.
31. Al-Rakhami, M.; Gumaei, A.; Alsanad, A.; Alamri, A.; Hassan, M.M. An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings. *IEEE Access* **2019**, *7*, 48328–48338. [[CrossRef](#)]
32. Ertugrul, Ö.F.; Kaya, Y. Smart city planning by estimating energy efficiency of buildings by extreme learning machine. In Proceedings of the 2016 4th International Istanbul Smart Grid Congress and Fair (ICSG), Istanbul, Turkey, 20–21 April 2016; pp. 1–5.
33. Kumar, S.; Pal, S.K.; Singh, R.P. Intra ELM variants ensemble based model to predict energy performance in residential buildings. *Sustain. Energy Grids Netw.* **2018**, *16*, 177–187. [[CrossRef](#)]
34. Kavaklioglu, K. Robust modeling of heating and cooling loads using partial least squares towards efficient residential building design. *J. Build. Eng.* **2018**, *18*, 467–475. [[CrossRef](#)]
35. Permai, S.D.; Tanty, H. Linear regression model using bayesian approach for energy performance of residential building. *Procedia Comput. Sci.* **2018**, *135*, 671–677. [[CrossRef](#)]
36. Sonmez, Y.; Guvenc, U.; Kahraman, H.T.; Yilmaz, C. A comparative study on novel machine learning algorithms for estimation of energy performance of residential buildings. In Proceedings of the 2015 3rd International Istanbul Smart Grid Congress and Fair (ICSG), Istanbul, Turkey, 29–30 April 2015; pp. 1–7.
37. Bui, X.N.; Moayedi, H.; Rashid, A.S.A. Developing a predictive method based on optimized M5Rules–GA predicting heating load of an energy-efficient building system. *Eng. Comput.* **2019**, 1–10. [[CrossRef](#)]
38. Goliatt, L.; Capriles, P.V.Z.; Duarte, G.R. Modeling heating and cooling loads in buildings using Gaussian processes. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–6.
39. Castelli, M.; Trujillo, L.; Vanneschi, L.; Popovič, A. Prediction of energy performance of residential buildings: A genetic programming approach. *Energy Build.* **2015**, *102*, 67–74. [[CrossRef](#)]
40. Seyedzadeh, S.; Pour Rahimian, F.; Rastogi, P.; Oliver, S.; Glesk, I.; Kumar, B. Multi-objective optimisation for tuning building heating and cooling loads forecasting models. In Proceedings of the 36th CIB W78 2019 Conference, Northumbria University, Newcastle, UK, 18–20 September 2019.
41. Urdaneta, S.; Juan Contreras, E.Z. Fuzzy Model for Estimation of Energy Performance of Residential Buildings. *Int. J. Appl. Eng. Res.* **2017**, *12*, 2766–2771.
42. Tahmassebi, A.; Gandomi, A.H. Building energy consumption forecast using multi-objective genetic programming. *Measurement* **2018**, *118*, 164–171. [[CrossRef](#)]
43. Nguyen, H.; Moayedi, H.; Jusoh, W.A.W.; Sharifi, A. Proposing a novel predictive technique using M5Rules-PSO model estimating cooling load in energy-efficient building system. *Eng. Comput.* **2019**, 1–10. [[CrossRef](#)]
44. Duarte, G.; Capriles, P.; Goliatt, L.; Lemonge, A. Prediction of energy load of buildings using machine learning methods. In Proceedings of the 4th Conference of Computational Interdisciplinary Science (CCIS), S ao José dos Campos-SP, Brazil, 7–10 November 2016.
45. Nilashi, M.; Dalvi-Esfahani, M.; Ibrahim, O.; Bagherifard, K.; Mardani, A.; Zakuan, N. A soft computing method for the prediction of energy performance of residential buildings. *Measurement* **2017**, *109*, 268–280. [[CrossRef](#)]
46. Reddy, J.N. *Introduction to the Finite Element Method; Mechanical Engineering*, 4th ed.; McGraw Hill Education: New York, NY, USA, 2019; ISBN 9781259861918 1259861910.
47. Brenner, S.; Scott, R. *The Mathematical Theory of Finite Element Methods*; Springer Science & Business Media: Berlin, Germany, 2007; Volume 15, ISBN 0387759336.
48. Aziz, A.K. *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*; Academic Press: Cambridge, MA, USA, 2014; ISBN 1483267989.
49. Babuška, I.; Banerjee, U.; Osborn, J.E. Generalized finite element methods—Main ideas, results and perspective. *Int. J. Comput. Methods* **2004**, *1*, 67–103. [[CrossRef](#)]

50. Whiteman, J.R. *The Mathematics of Finite Elements and Applications: Proceedings of the Brunel University Conference of the Institute of Mathematics and Its Applications Held in April 1972*; Academic Press: London, UK, 1973; ISBN 0127472509 9780127472508. Available online: <https://ua.on.worldcat.org/oclc/764065> (accessed on 3 October 2019).
51. Zienkiewicz, O.C.; Taylor, R.L.; Taylor, R.L. *The Finite Element Method for Solid and Structural Mechanics*; Butterworth-Heinemann: Oxford, UK, 2005; ISBN 0750663219.
52. Villacampa, Y.; Navarro-González, F.J.; Llorens, J. A geometric model for the generation of models defined in Complex Systems. *WIT Trans. Ecol. Environ.* **2009**, *122*, 71–82.
53. Navarro-González, F.J.; Villacampa, Y. A new methodology for complex systems using n-dimensional finite elements. *Adv. Eng. Softw.* **2012**, *48*, 52–57. [[CrossRef](#)]
54. Navarro-González, F.J.; Villacampa, Y. Generation of representation models for complex systems using Lagrangian functions. *Adv. Eng. Softw.* **2013**, *64*, 33–37. [[CrossRef](#)]
55. Navarro-González, F.J.; Villacampa, Y. A finite element numerical algorithm for modelling and data fitting in complex systems. *Int. J. Comput. Methods Exp. Meas.* **2016**, *4*, 100–113. [[CrossRef](#)]
56. Buhmann, M.D. Radial Basis Functions. *Acta Numer.* **2000**, *9*, 1–38. [[CrossRef](#)]
57. Carlson, R.E.; Foley, T.A. Interpolation of track data with radial basis methods. *Comput. Math. Appl.* **1992**, *24*, 27–34. [[CrossRef](#)]
58. Autodesk. Available online: www.autodesk.com/ecotect-analysis (accessed on 10 September 2011).
59. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006.
60. Werner, P.; Ardeshir, M. Building Morphology, Transparency, and Energy Performance. In Proceedings of the 8th International IBPSA Conference, Eindhoven, The Netherlands, 11–14 August 2003; pp. 1025–1032.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).