# A new hybrid model to foretell thermal power efficiency from energy performance certificates at residential dwellings applying a Gaussian process regression

Paulino José García Nieto[a,*], Esperanza García-Gonzalo[a], José Pablo Paredes-Sánchez[b], Antonio Bernardo Sánchez[c]

[a]Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain

[b]Department of Energy, College of Mining, Energy and Materials Engineering, University of Oviedo, 33004 Oviedo, Spain

[c]Department of Mining Technology, Topography and Structures, University of León, 24071 León, Spain

**Abstract** An energy performance certificate (EPC) provides information on the energy performance of an energy system. The objective of this research aimed at obtaining a predictive model for early detection of thermal power efficiency (TPE) for energy conversion and preservation in buildings. This article expounds a sound and solid non-parametric Bayesian technique known as Gaussian Process Regression (GPR) approach, based on a set of data collected from different dwellings in an oceanic climate. Firstly, this model introduces the relevance of each predictive variable on energy performance in residential buildings. The second result refers to the statement that we can predict successfully the TPE by using this model. A coefficient of determination equal to 0.9687 was thus established in order to predict the TPE from the observed data, using the GPR approach in combination with the differential evolution optimiser (DE). The concordance between experimental observed data and the predicted data from the best-proposed novel hybrid DE/GPR–relied model demonstrated here the adequate efficiency of this innovative approach.

[*]Corresponding author. Tel.: +34-985103417; fax: +34-985103354.
*E-mail address*: lato@orion.ciencias.uniovi.es (P.J. García Nieto).

**Keywords** Energy performance certificate (EPC); Gaussian process regression (GPR); Differential evolution (DE); Residential buildings

## 1 Introduction

The analysis of energy management and energy consumption in buildings is essential to address the energy efficiency challenge in buildings and to meet the current demands for societal comfort, urban development and the resulting increase in energy consumption [1]. This sector demands about 40% of total energy consumption worldwide, consequently, is also responsible for the corresponding carbon emissions [2]. Energy Performance Certificates (EPC) are a rating system aimed at determining the energy performance of buildings in the European Union [3]. The requirements for energy certification of buildings are laid down in Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 and are transposed in each European country in the form of a certification procedure for the energy performance of new buildings. Many factors determine a building performance based on energy demand such as weather conditions, building structure and the energy systems used to meet that demand for energy in order to maintain thermal comfort in the building itself. This complexity makes it very difficult to accurately predict building performance, which is required for proper energy management. In order to produce an EPC, an energy study is carried out by an assessing agent who visits the property, analyses its location and examines key elements of the building such as insulation, energy systems (e.g. domestic boiler, storage tank and radiators, etc.) and/or other structural characteristics. The assessing agent then enters the findings into a software program that performs the energy efficiency calculation, e.g. CE3, CE3X, HULC or CERMA. The software provides a numerical value for the energy efficiency rating as well as a recommended value for the improvement potential based on the building characteristics. In this respect, the building is given a rating certificate of between A (very efficient) and G (inefficient). However, the EPC will also include advice on the most cost-effective ways to improve the rating and use of energy in the building. Energy systems such as heating and cooling systems, play a central role for energy efficiency and the implementation of local actions towards better energy management. Understanding the characteristics and performance of the existing systems is fundamental in devising realistic energy-saving strategies [4]. Energy system analysis stands out as one of the main challenges in thermal analysis in district heating systems for buildings [5]. EPC analysis based on advanced modeling tools is expected to be an important driver in promoting energy efficiency improvements in buildings [6]. On this matter, more developed methods such as artificial neural networks (ANN) [7] and support vector machines (SVM) [8] have been used for predicting the thermal power efficiency (TPE) value – an indicator of the energy efficiency – at residential buildings. Recently, Melo and coworkers [9] have used multiple linear regressions for estimating the energy performance of the building stock (surrogate models). Energy conversion through thermal energy systems needs to be studied from a new interdisciplinary perspective based on efficiency management with analysis models. This is where most advanced energy systems call for performance data analysis in order to determine appropriate energy conversion and management in future cities. However, TPE prediction at residential buildings has not been totally successful so far [10]. TPE is a parameter related to energy efficiency testing of a single dwelling. TPE is a non-dimensional measure of the performance of a device that uses thermal energy, such as a heating/cooling system. As such, it defines the performance of advanced thermal energy systems. An accurate TPE forecast has become very attractive, as some guidelines for the use of this essential parameter in residential buildings (e.g. de Wilde [11]) have limit values for energy efficiency.

As a result, the implementation of the innovative technique that combines the Gaussian process regression (GPR) approximation [12–14] with the optimisation algorithm Differential Evolution (DE) [15–18] to foretell the outlet TPE used in energy performance certificates at residential buildings could be an attractive methodology since, as far as authors know, it has never been tackled in prior researches before. Additionally, genetic algorithms (GA) [19–23] and limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [24–30] optimizers have been used for comparison purposes along with the DE optimizer, that are three of the most common used methods to optimize the hyperparameters, reaching the conclusion that, for the problem at hand, the more consistent and robust method is the chosen one to build the model, that is, DE optimizer. On the other hand, the GPR technique is a statistical learning methodology developed by statistics and Bayesian analysis, which is capable of dealing with non-linearities, including interactions among variables [12–14]. If we compare it with other classical and metaheuristic regression techniques, GPR approximation presents some benefits [31]: (1) GPR has a remarkable ability to be widespread; (2) the GPR optimal parameters can be determined using heuristic optimizers; (3) the GPR results show an evident probabilistic significance; (4) the GPR works well on small datasets; (5) it makes use of the whole information available; and (6) moreover, it has the ability to provide uncertainty measurements on the predictions, which is the main characteristic that differentiates it from other regression methods. In effect, the DE metaheuristic optimiser has been used here satisfactorily to calculate the optimal GPR hyperparameter. Furthermore, former researches indicate that GPR is a very appropriate tool in a large number of real applications as computational fluid dynamics (CFD) [32], geometrical characteristics of cladding tracks [33], modeling of systems relied on solar/waste energy [34], blood pressure measurement [35], state of health for lithium battery [36], building energy use [37], time series analysis [38], wind speed prediction [39], gravity field modeling [40], sunspot cycle prediction [41], wave energy converter arrays prediction [42], hydrodynamic forces prediction between multiple floating bodies [43] and so on. However, it has never been used for evaluating TPE from energy performance certificates at residential dwellings to define energy performance contribution from modern energy systems.

The principal goal of the current research was to foretell the TPE, as a function of predictive variables, in terms of building performance in residential properties with thermal systems. This was determined from energy performance certificates employing Gaussian process regression (GPR) along with the differential evolution (DE) optimizer. This model defines a new algorithm to analyse energy conversion and management in buildings. This is a key element for the future development of more advanced thermal engines and machines in the technological industrial context.

The framework of this article is arranged in the following way: Section 2 presents the experimental arrangement, all the variables included in this research and GPR methodology; Section 3 draws up the findings gathered with this novel technique by collating the GPR results with the observed values as well as the significance ranking of the predictive variables; finally, Section 4 concludes with a list of the principal results of research.

## 2 Materials and methods

### 2.1 Methodology: machine learning techniques

There are different approaches to the challenge of making energy predictions [44]. One of these approaches to the problem lies in the use of automatic learning methods (black box methods), in the context of the fourth industrial revolution. We are becoming smarter, more sustainable and more efficient in our use of energy, which will bring about more future changes. The purpose of using machine learning to predict problems is to simulate a data-based model to estimate new outcomes. A series of samples are required for predicting the results of energy models using machine learning. Such samples represent individual observations of the expected energy model. Each sample consists of a number of inputs (conversion and management operation parameters) and at least one output. Samples are used in order to train the learning model of the machine to match the inputs to the targets.  Such an approach allows sophisticated models to be predicted within a reasonable time.

Studies have shown that it is possible to reproduce almost any desired model by applying a correct machine learning algorithm, regardless of its complexity level [45]. Applying machine learning to predict building energy performance has been extensively researched. The ANN is a non-linear computer model inspired by the human brain. A typical ANN includes sequential layers. ANNs have previously been used to simulate energy efficiency in buildings. Studies in this field have sought to predict heating [46], cooling [47], heating and cooling [48] and electrical energy consumption in buildings [49].

The studies mostly focused on the machine learning algorithms used in previous research efforts. A number of machine-learning algorithms have been used to develop energy prediction problems to model end-use energy consumption in the residential sector [50]. A review of state-of-the-art studies, both on energy modeling of buildings and on approaches to building energy, reveals how important the new hybrid models are as regards building types (e.g. residential and non-residential), time forecasting (e.g. hourly or annual), and expected types of energy consumption (e.g. heating or cooling) [51]. Certification and energy efficiency are a driving force in physics-based models towards the future development of tools based on self-learning algorithms.

ANNs show multiple advantages over physics-based models since this avoids having to simulate the physical processes that drive the energy transfer of the entire energy model in homes. However one of the common challenges of ANNs is the training required in the conversion and operation of a real-world building as an energy system. These disadvantages of ANNs are overcome by using the GPR technique for lazy learning as well as a similarity measure between parameters to predict the value of an unobserved result from training data. Some earlier works involving the GPR technique suggest a model for energy consumption based on data on energy use factors such as user habits, operating conditions and equipment, focusing in particular on occupants of buildings [51], the probabilistic forecast of building energy demand [52], or determining the thermal performance of the building [53] and energy savings [54].

In this context, this work offers a novel hybrid model that could be an interesting approach since, at the knowledge of the authors, it has not been yet addressed in previous investigations to foretell the energy performance of the buildings by the use of TPE parameter. Unlike many popular supervised machine learning algorithms that learn exact values for every parameter in a function, the Bayesian approach infers a probability distribution over all possible values. Gaussian process regression is nonparametric (i.e. not limited by a functional form), so rather than calculating the probability distribution of parameters of a specific function, GPR calculates the probability distribution over all admissible functions that fit the data.

## 2.2 Experimental setup in the study area

The EPC includes objective information as data about the management of the energy systems and environmental conditions of the building. The study area is located in the oceanic climate area, which is considered a widespread environment condition and stretches across Northwest Europe. To promote energy efficiency in building, European regulations in force have established mandatory energy efficiency certification of buildings. For this purpose, standardized systems have been established so as to obtain and register such energy certificates. In this regard, a series of standardized parameters are recorded, in order to guarantee the quality of the information in these certifications. This paper is based on parameters from the study area as listed in the registry of the Spanish regulatory authority in energy certifications [55]. The data collection is based on the evaluation of energy efficiency from buildings in Spain by software tools as CE3, CE3X, HULC and CERMA. Such parameters are obtained by means of standardized software that accredits the validity of the energy certification of buildings in accordance with the reference [55]. The available data were collected from 137 samples of different residential houses in the North of Spain [55]. This location was chosen because it is a representative area with performance conditions in buildings of oceanic climate and modern energy systems based on the recent advances in the heating and cooling sector in buildings in Spain, one of the most representative of Europe. They are examples of standardised software used for energy certification of advanced energy systems, depending on the characteristics of each building [56]. The diagram of the analysis is shown in Fig. 1.

**Fig. 1** Diagram of data collection

2.3 Variables of the concerning hybrid model

The quality of this model depends on the quality of the input data, being this one of the most valuable elements in the analysis model. The obtained energy performance certificate (EPC) was established to serve as the basis to calculate the TPE based on different operation variables that the GPR–relied methodology requires as data entry. In this case, the output variable is the TPE, which is an indicator of the minimisation of the energy demand through the design and efficient use of renewable energy integration in the buildings. TPE is the ratio between the useful output of a device or installation and the input, in energy terms, i.e., hot/cold thermal energy. In the particular case of a refrigeration or heating pump system, thermal efficiency is the ratio between the net heat output for heating, or the extraction for cooling, and the energy input (i.e., the coefficient of performance). In this case, very high ratios (even greater than 100%) are obtained from advanced energy systems. The TPE is calculated by the certification software in accordance with European Directives [56].

**Fig. 2** Example of a building and the parameters analysed in it as an energy system

Some predictive variables have been selected based on the certification software in accordance with the Spanish construction standards. These parameters are defined as main elements of the energy system in certification databases. The regulatory authority establishes the minimum content of the energy certificate, which affects aspects related to the thermal enclosure and the thermal installations of the enclosure, also known as energy systems [55]. Fig. 2 shows the predictive operation variables in the energy system as follows:

- Useful surface ($m^2$): it is the floor area demarcated by the inner perimeter of the external enclosures of the building structure or an element of a building, including half of the floor area of its covered exterior spaces (such as terraces, balconies and clotheslines, porches, loading docks, overhangs, etc.), as measured on its horizontal projection of its roof.
- Thermal power (kW): it is the power generated by the energy system in order to provide comfortable thermal conditions in a building. It is defined in terms of the energy system of the building to cover the thermal demand under the existing environmental conditions of the building.
- $CO_2$ emissions (kg $CO_2/m^2$ year): they are the potential emissions caused by the energy system due to the operating conditions and the energy conversion to produce thermal energy in the energy system from primary energy.
- Primary energy consumption (kWh/$m^2$ year): it is the energy supplied to the building, which may come from either renewable or non-renewable sources, and which has not undergone any previous conversion or transformation process. It is the energy found in fuels and other energy sources, and includes the energy

necessary to generate the final energy consumed by the building, including losses due to transport to the building, storage, etc.

- Opaque enclosures ($m^2$): they are the coverings of  the sides of the building which are not translucent. These elements act as a barrier to heat transfer and energy loss.
- Holes and skylights ($m^2$): they are constructive elements at the top of the building which are translucent and could contribute to the increase of energy losses by heat transfer, e.g. windows areas.

TPE is the output variable (dependent variable), while the above discribed varibables are input (independent variables) from our machine learning models.

2.4 Gaussian process regression (GPR)

A Gaussian process (GP) is defined as a stochastic process that generates samples over time $\{X_t\}_{t\in\tau}$ in such a way that it does not affect the finiteness of a linear combination $X_t$ having (or more generally any linear function of the sample function $X_t$), a linear combination that will normally be distributed [14,31,57–61]. Let’s assume that $T=\{(\mathbf{x}_i, y_i)/i=1,2,...,N\}$ depicts the training collection data of the Gaussian method. When we approach a regression problem using Gaussian processes (also termed kriging), the following hypothesis is made: for a Gaussian process f observed at the $\mathbf{x}$ coordinates, the vector of values $f(\mathbf{x})$ is only a sample of a multivariate Gaussian distribution of dimension equal to the number of observed coordinates n. It is a well-known fact that Gaussian processes can be utterly established by their second-order statistics. Hence, supposing that a Gaussian process with a zero mean, the definition of the covariance matrix $K$ (positive definite kernel) will completely determine the performance of the Gaussian process. The covariance matrix permits to define Gaussian process concepts like isotropy, stationarity, smoothness and periodicity. The following is a list of some common covariance functions used in many regression problems [14,31,57–61]:

- Linear: $K_L(\mathbf{x},\mathbf{x}') = \mathbf{x}^T\mathbf{x}'$
- Squared exponential: $K_{SE}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \cdot e^{-\frac{|d|^2}{2\ell^2}}$
- Rational quadratic: $K_{RQ}(\mathbf{x},\mathbf{x}') = \left(1+|d|^2\right)^{-\alpha}$, $\alpha \ge 0$
- Periodic: $K_P(\mathbf{x},\mathbf{x}') = e^{-\frac{2\sin^2\left(\frac{d}{2}\right)}{\ell^2}}$
- Ornstein–Uhlenbeck: $K_{OU}(\mathbf{x},\mathbf{x}') = e^{-\frac{2\sin^2\left(\frac{d}{2}\right)}{\ell^2}}$

where $d$ is equal to $d = \mathbf{x} - \mathbf{x}'$, $\ell$ is the characteristic length-scale of the process and $\sigma_f^2$ is the signal variance. In this study, the squared exponential kernel, also called radial basis function (RBF), was used due to its better performance compared to other kernels with a larger volume of training data. The results of this model based on Gaussian processes (GPs) rely on the values of the hyperparameter $\theta$ (for example, ℓ and $\sigma_f^2$ values) since these values determine the behavior of the model. In practice, the real experimental data are noisy observations so that [14,25]:

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon \tag{1}$$

where $\varepsilon$ depicts the white noise as an additive term in Eq. (1). In practice, it is common to consider $\varepsilon \sim N\left(0,\sigma_n^2\right)$, which means that Gaussian noise will be independent and identically distributed, being $\sigma_n$ the standard deviation of this noise. To fix ideas, it is possible to take into account a finite ensemble of the noisy real experimental data as a separate Gaussian procedure expressed as [31,58–60]:

$$\mathbf{y} \sim GP\left(m(\mathbf{x}), K(\mathbf{x},\mathbf{x}') + \sigma_n^2 \delta_{ij}\right) = GP\left(0, K(\mathbf{x},\mathbf{x}') + \sigma_n^2 \delta_{ij}\right) \tag{2}$$

where $\delta_{ij}$ is the Kronecker delta distribution and [14,31]:

$$\begin{aligned} m(\mathbf{x}) &= E\left[ f(\mathbf{x}) \right] \\ K(\mathbf{x},\mathbf{x}') &= Cov(\mathbf{x},\mathbf{x}') = E\left[ \left( f(\mathbf{x}) - m(\mathbf{x}) \right)\left( f(\mathbf{x}') - m(\mathbf{x}') \right)^T \right] \end{aligned} \tag{3}$$

Therefore, according to the hypothesis of a zero-mean distribution, $f(\mathbf{x}) \sim N\left(0, K(\theta,\mathbf{x},\mathbf{x}')\right)$, we have that $K(\theta,\mathbf{x},\mathbf{x}')$ is termed a covariance matrix between all possible pairs $(\mathbf{x},\mathbf{x}')$ for a given set of hyperparameters $\theta$. In this case, the logarithmic marginal probability is given by [25,52–54]:

$$\begin{aligned} \log p\left( f(\mathbf{x}) \middle| \theta, \mathbf{x} \right) = -\frac{1}{2} f(\mathbf{x})^T K(\theta,\mathbf{x},\mathbf{x}') f(\mathbf{x}') - \frac{1}{2} \log\det\left( K(\theta,\mathbf{x},\mathbf{x}') \right) \\ -\frac{n}{2}\log 2\pi \end{aligned} \tag{4}$$

The calculation of the maximum of this marginal likelihood concerning $\theta$ determines the whole requirement of the Gaussian process f. Note in (4) that the first term appertains to a penalty term as a consequence of the failure of the model to fit the experimental or observed values while the second term is a penalty term that grows in proportion to the complexity of the model.

Making predictions about an unobserved value $f(\mathbf{x})$ in the $\mathbf{x}$ coordinate, after specifying $\theta$, is just an issue of extracting samples from the predictive distribution $p\left( y^* \middle| \mathbf{x}^*, f(\mathbf{x}), \mathbf{x} \right) = N\left( y^* \middle| A, B \right)$ where the posterior estimation of the mean A is given by [14,31,57–61]:

$$A = K\left(\theta, \mathbf{x}^*, \mathbf{x}\right) K\left(\theta, \mathbf{x}, \mathbf{x}'\right)^{-1} f\left(\mathbf{x}\right) \tag{5}$$

and the posterior estimation of the variance is determined by [14,31,57–61]:

$$B = K\left(\theta, \mathbf{x}^*, \mathbf{x}^*\right) - K\left(\theta, \mathbf{x}^*, \mathbf{x}\right) K\left(\theta, \mathbf{x}, \mathbf{x}'\right)^{-1} K\left(\theta, \mathbf{x}^*, \mathbf{x}\right)^T \tag{6}$$

where [14,31]:

- $K\left(\theta, \mathbf{x}^*, \mathbf{x}^*\right)$: would be the variance matrix at the new unobserved point $\mathbf{x}^*$ for a given vector of $\theta$ hyperparameters;
- $K\left(\theta, \mathbf{x}^*, \mathbf{x}\right)$: would be the covariance matrix between a new unobserved value $\mathbf{x}^*$ and all the remaining observed values of the $\mathbf{x}$ coordinate for a given vector of $\theta$ hyperparameters;
- $K\left(\theta, \mathbf{x}, \mathbf{x}'\right)$: is the covariance matrix between all possible pairs $\left(\mathbf{x}, \mathbf{x}'\right)$, as previously defined.

It is possible to point out that the posterior mean estimation $f\left(\mathbf{x}^*\right)$ at the new unobserved point $\mathbf{x}^*$ is just a linear combination of the observed values of $f\left(\mathbf{x}\right)$. Additionally, the variance of $f\left(\mathbf{x}^*\right)$ is independent of the observed values of $f\left(\mathbf{x}\right)$.

Hence, the GPR technique is relied on a nonparametric methodology since its predictive capacity falls on the observed values $\mathbf{y}$ and on the input data. Following this procedure, the values $\theta = \left\{l, \sigma_f, \sigma_n\right\}$ are called the GPR model hyperparameters [14,31,62]. In order to determine the optimal hyperparameters $\theta' = \arg\max_{\theta} \log p\left(\mathbf{y} | X, \theta\right)$, it is possible to employ any standard optimiser after parameter initialization. In this research work, the metaheuristic optimisation method, termed DE algorithm [15–18,63,64] described below, is successfully applied.

## 2.5 Differential evolution (DE) optimisation algorithm

The differential evolution (DE) metaheuristic algorithm [15–18,63,64] is a simple but powerful tool used to solve global optimisation problems. In this sense, those control parameters involved in the DE are widely related to the problem under consideration and therefore influence its performance.

Storn and Price introduced the DE algorithm in 1997, which is a stochastic optimisation algorithm [15–18]. Let $S \subseteq \Re^{D}$ be the search space for the problem under examination, the DE involves a population of *NP* vectors (candidate solutions) $\mathbf{x}_{i,g} = \left\{x_{1i,g}, x_{2i,g}, ..., x_{Di,g}\right\} \in S,\ i = 1, 2, ..., NP$. Each $x_{ji,g}$ corresponds to a problem decision variable and *g* shows the generation to which the vector belongs. These vectors, after being initialised, are subjected to mutation, recombination and selection operations in each generation *g*. The stages of this optimisation technique is as follows [15–18,63,64]:

- Initialisation: The lower and upper limits for each decision variable are defined in advance: $x_j^i \le x_{ji,1} \le x_j^s$. Subsequently, the initial values of the decision variables on the intervals $\left[x_j^i, x_j^s\right]$ are selected randomly and uniformly.

- Mutation: For each vector $\mathbf{x}_{i,g}$ (target vector) in the generation g, a mutated vector $\mathbf{v}_{i,g} = \{v_{1i,g}, v_{2i,g}, ..., v_{Di,g}\}$ is created using several strategies. To classify these variants, the notation $DE/x/y/z$ is used, where *x* indicates the vector to be mutated (rand or best), *y* the number of subtractions of vectors performed (1 or 2), and *z* denotes the recombination scheme used (bin: binomial or\ exp: exponential). The most commonly used strategies to generate $\mathbf{v}_{i,g}$ are [63,64]:

1. $DE/\text{rand}/1/\text{bin}$ : $\mathbf{v}_{i,g} = \mathbf{x}_{r1,g} + F\left(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g}\right)$
2. $DE/\text{best}/1/\text{bin}$ : $\mathbf{v}_{i,g} = \mathbf{x}_{best,g} + F\left(\mathbf{x}_{r1,g} - \mathbf{x}_{r2,g}\right)$

where the indexes $r1$, $r2$, $r3$ are random integers and mutually different generated in the range $[1, NP]$. *F* is a factor between $[0,\infty)$ to scale the difference of vectors (mutation) and $\mathbf{x}_{best,g}$ is the vector with the best value of fitness of the population in the generation *g*. Parameter *F* is termed *differential weight* whose value is normally found in interval $[0, 2]$.

- Recombination: The recombination operation (crossover) is applied to each part of the mutated vector generated $\mathbf{v}_{i,g}$ and its corresponding target vector $\mathbf{x}_{i,g}$ to generate a test vector $\mathbf{u}_{i,g} = \{u_{1i,g}, u_{2i,g}, ..., u_{Di,g}\}$ [15–18,63,64]:

$$u_{ji,g} = \left\{ \begin{array}{ll} v_{ji,g} & \text{if } \left(rand_j\,[0,1] \le CR\right) or \left(j = j_{rand}\right) \\ x_{ji,g} & \text{otherwise} \end{array} \right\} \tag{7}$$

where $j = 1, 2, ..., D$, CR is a constant that indicates the probability of recombination in the range $[0,1)$, called the crossover probability, and $j_{rand}$ is a random integer chosen in the range $\left[1, NP\right]$ to ensure that the test vector is different from the corresponding target vector. The given operator in Eq. (7) corresponds to the binomial crossover.

- Selection: The fitness value of each test vector $f\left(\mathbf{u}_{i,g}\right)$ is compared with its corresponding target vector $f\left(\mathbf{x}_{i,g}\right)$ in the current population. The vector with the best fitness value is the one to enter the population of the next generation [15–18,63,64]:

$$\mathbf{x}_{i,g+1} = \begin{Bmatrix} \mathbf{u}_{i,g} & \text{if } f\left(\mathbf{u}_{i,g}\right) < f\left(\mathbf{x}_{i,g}\right) \\ \mathbf{x}_{i,g} & \text{otherwise} \end{Bmatrix} \tag{8}$$

The last three operations are repeated from generation to generation until a specific detention criterion is satisfied. More specifically, the stopping criterion is fulfilled here if there is no improvement in the $R^2$ after ten iterations, along with a maximum number of iterations equal to 500.

## 2.6 Genetic algorithm (GA) optimisation

A genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimisation and search problems by relying on biologically inspired operators such as *mutation*, *crossover* and *selection* [19–23]. In a genetic algorithm, a population of candidate solutions (called individuals or phenotypes) to an optimisation problem develops into better solutions. Each candidate solution has a set of properties (its chromosomes or genotype) which can be mutated and altered [19–23]. The evolution usually starts from a population of randomly generated individuals, and is an *iterative process*, with the population in each iteration called a *generation*. In each generation, the *fitness* of every individual in the population is evaluated. The fitness is usually the value of the *objective function* in the optimisation problem being solved. The fittest individuals are *stochastically* selected from the current population, and each individual's genome is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

## 2.7 Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimisation algorithm

Limited-memory BFGS (L-BFGS) is an optimisation algorithm in the family of quasi-Newton methods that approximates the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) using a limited amount of computer memory [24–28]. It is a popular algorithm for parameter estimation in machine learning. The algorithm's target problem is to minimise $f\left(\mathbf{x}\right)$ over unconstrained values of the real-vector $\mathbf{x}$ where $f$ is a differentiable scalar function. Like the original BFGS, L-BFGS uses an estimate of the inverse Hessian matrix to steer its search through variable space, but where BFGS stores a dense $n \times n$ approximation to the inverse Hessian (*n* being the number of variables in the problem), L-BFGS stores only a few vectors that represent the approximation implicitly [27–30]. Due to its resulting linear memory requirement, the L-BFGS method is particularly well suited for optimisation problems with many variables. Instead of the inverse Hessian $\mathbf{H}_k^{-1}$, L-BFGS maintains a history of the past *m* updates of the position

$\mathbf{x}$ and gradient $\nabla f\left(\mathbf{x}\right)$, where generally the history size $m$ can be small (often $m<10$) [27–30]. These updates are used to implicitly do operations requiring the $\mathbf{H}_k$ −vector product. The algorithm starts with an initial estimate of the optimal value, $\mathbf{x}_0$, and proceeds iteratively to refine that estimate with a sequence of better estimates $\mathbf{x}_1$, $\mathbf{x}_2$, etc. The derivatives of the function $g_k = \nabla f\left(\mathbf{x}_k\right)$ are used as a key driver of the algorithm to identify the direction of the steepest descent, and also to form an estimate of the Hessian matrix (second derivative) of $f\left(\mathbf{x}\right)$ [24–30]. L-BFGS algorithm shares many features with other quasi-Newton algorithms, but is very different in how the matrix-vector multiplication $d_k = -H_k^{-1} g_k$ is carried out, where $d_k$ is the approximate Newton's direction, $g_k$ is the current gradient, and $H_k^{-1}$ is the inverse of the Hessian matrix [27–30].

2.8 Accuracy of the mentioned approach in energy systems

This novel DE/GPR-based method was developed with six predictive input variables already described in subsection 2.3 above. TPE is, as we know, the dependent variable to be predicted. In order to accurately and reliably forecast TPE from the six remaining input variables, it is mandatory to select the best model that fits the observed dataset. Although several possible statistics can be applied to ascertain the goodness–of–fit, the rule used in this study was the coefficient of determination $R^2$ [66,67], which is a statistic used in the context of a statistical model to foretell future results or to test a hypothesis. Next, we will call the observed values $t_i$ versus the values predicted by model $y_i$. Now we can define the following sums of squares given by [67]:

- $SS_{tot} = \sum_{i=1}^{n} (t_i - \bar{t})^2$ : is the overall sum of squares, proportional to the sample variance.
- $SS_{reg} = \sum_{i=1}^{n} (y_i - \bar{t})^2$ : is the regression sum of squares, also termed the *explained sum of squares*.
- $SS_{err} = \sum_{i=1}^{n} (t_i - y_i)^2$ : is the residual sum of squares.

where $\bar{t}$ is the mean of the $n$ observed data [66,67]:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i \tag{9}$$

Considering the former sums, the coefficient of determination is given by the following equation [66,67]:

$$R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \tag{10}$$

Further criterion considered in this study was the root mean square error (RMSE) [66,67]. Such a statistic is also frequently used to evaluate the predictive capability of a mathematical model. Indeed, the root mean square error (RMSE) [66,67]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (t_i - y_i)^2}{n}} \tag{11}$$

If the root mean square error (RMSE) is zero, then it means that there is no difference between the predicted and the observed data.

Moreover, the GPR methodology relies heavily on three *hyperparameters* [12–14]:

- Variance ($\sigma_f^2$): this parameter refers to the variance of the signal; its purpose is to control the vertical scale of the kernel function.

- Lengthscale ( $\ell$ ): this parameter provides the characteristic scale of the length. It allows controlling the horizontal scale where the kernel function changes.

- Gaussian noise variance ( $\sigma_n^2$ ): this parameter is the variance of a Gaussian additive white noise $\varepsilon \sim N\left(0,\sigma_n^2\right)$.

In this research, a metaheuristic optimisation method, denominated DE algorithm [15–18], along with GA and L-BFGS optimisers for comparative purposes [19–30], is employed because of its characteristics of moderate requirement of computational memory and solid convergence for complex problems. Moreover, it should also be noted that the DE iterates until a convergence is reached (i.e. iterative scheme) to improve a proposed solution concerning a certain quality measure. DE optimiser is used for real-value multidimensional functions, but it does not use the gradient of the problem that is being optimised, which means that DE optimiser does not require for the optimisation problem to be differentiable, like the traditional optimisation methods such as gradient descent and quasi-newton methods. Thus, DE optimizer can also be used in optimisation problems that are not even continuous, change over time, and so on. [18,63,64].

## 3 Results and discussion

All the six operation input variables are shown in Table 1. In this study, the total number of samples used here was 137, that is to say, 137 certificates of energy performance in residential housing were collected and processed.

**Table 1** Set of physical predictive variables of the operation used in this study: names, means and standard deviations

To confront this complex problem here, it is necessary to split the complete set of data in two parts: (1) a training set comprising 80% of the data; (2) and a testing set comprising the remaining 20% of the data. The key idea is to build a GPR-relied model with training data by determining the optimal parameters with the DE, GA or L-BFGS optimisers, take the best model from it and then apply it to the test data to predict values.

As stated above, the output variable (dependent variable) in this study is TPE dealt with by means of the DE/GPR–relied method. The forecasting carried out from the six independent predictive variables was quite good. The selection of the optimal hyperparameters is a key factor in the performance of this method as we see above: (1) the lengthscale $\ell$ and variance $\sigma_f^2$ of the radial basis function (RBF) kernel; and (2) Gaussian noise variance $\sigma_n^2$. The objective function used in the hyperparameter optimisation process is the –log likelihood value (see Eq. 4). In this way, Table 2 indicates the initial intervals of the three hyperparameters of the GPR–based model fitted in this study.

**Table 2** Initial ranges of the three hyperparameters of the DE/GPR–relied model fitted in this study

The GP parameter tuning with DE, L-BFGS and GA algorithms was carried out. Each optimisation process was run 50 times to get the best result and perform a comparative statistical analysis of the three optimisers. The conclusions are as follows: (a) if the three algorithms are allowed enough multiple runs, all three algorithms get approximately the same minimum value of the objective function; (b) the robustness of the algorithm varies widely: DE obtains systematically the same minimum value while GA gets a wide variety of values with a much greater dispersion. L-BFGS performs in an in-between position as the results vary but not as widely as with GA. This can be seen in Table 3 and Fig. 3.

**Table 3** Minimum, mean, maximum and standard deviation values for the objective function (–log likelihood) for the 50 runs of the optimisation stage for the DE, GA and L-BFGS algorithms

**Fig. 3** Boxplots for the values of the objective function (-log likelihood) obtained from running each of three optimisers fifty times

Thus, the results show DE as the more consistent and robust algorithm. Moreover, the variation between results is so small that it would not be necessary to run it multiple times, making the whole process more efficient.

As DE, L-BFGS and GA algorithms are stochastic, the same final results will not be obtained in different runs. In order to compare results, each process was carried out fifty times. As the sets of results do now follow a normal distribution with a Shapiro-Wilk normality test, a Mann-Whitney rank test was performed to verify that the values are significantly different. The results are, indeed, different for each optimisation algorithm as can be seen by the *p*-values in Table 4, all well under 0.05.

**Table 4** *p*-values of the Mann-Whitney rank test comparing the sets 50 objective functions values obtained for each of the different hybrid models using DE, GA and L-BFGS as parameter tuning optimisation algorithms

We picked up the set of parameters obtained with the DE optimizer. According to this methodology, Table 5 identifies the optimal parameters of the best fitted GPR–relied approach encountered with the DE optimizer.

**Table 5** Optimal hyperparameters of the best-fitted GPR–relied model encountered with the DE, optimizer: variance $\sigma_f^2$ and lengthscale $\ell$ for the RBF kernel, the Gaussian noise variance $\sigma_n^2$, and the corresponding objective function (-log likelihood) value for the optimized model for the training set

The value of $R^2$ was determined by employing the optimised model processing the testing dataset. The unit Gpy, used to implement the Gaussian process in python [68,69], along with the DE optimizer [15–18], were employed to build the definitive regression approach.

Considering the calculations accomplished, the DE/GPR–relied technique has permitted to construct an innovative model with high allowances to assess the thermal power efficiency (TPE) through the test dataset. Most certainly, the value of $R^2$ of the best GPR approach was 0.9687 with a correlation coefficient $(r)$ of 0.9867 for the dependent TPE variable. Table 6 shows the coefficients of determination for the three different models of regression and for the testing dataset.

**Table 6** Coefficient of determination ( $R^2$ ), coefficient of correlation $r$ and root mean square error *RMSE* for the testing dataset for the DE/GPR regression model

3.1 The importance of variables in the energy system

The importance of variables was studied removing a variable and evaluating the new model performance and comparing it with the full model's. The greater the drop in the goodness of fit parameter, the greater the importance of the independent variable. Both in our experience, and also in the literature [70] the traditional assessment of the relevance of the variables through automatic relevance determination (ARD) does not furnish an appropriate method as it automatically undervalues the effect of linear input variables compared to nonlinear ones that have a similar effect in the calculation of the squared error [71]. An alternative method to estimate the importance of the variables would be an analysis of sensitivity. That is, the value of an independent variable is slightly modified and the later latent average is evaluated [72]. If a single independent variable is changed resulting in a large alteration of the variance of the latent mean (VLM), such variable will be outstanding. Also, it must be taken into account that, even though either quantitative estimations or the relevance are given, the nature of the importance study remains qualitative.

A further relevant finding of this study is the importance of the input variables in order to foretell TPE for this complicated nonlinear research (see Table 7 and Fig. 4). Therefore, $CO_2$ emissions is the most relevant predictive variable according to DE/GPR approach in the TPE forecasting followed by opaque enclosures, and then holes and skylights, primary energy consumption, useful surface and thermal power.

**Table 7** Relative importance of the input variables as stated in the DE/GPR–relied approach for the TPE according to VLM method

**Fig. 4** Relative importance of the input variables as stated in the DE/GPR–relied approach for the TPE according to VLM method

$CO_2$ emissions are the most representative variable in predicting TPE for the VLM method, followed by opaque enclosures, and holes and skylights. Primary energy consumption, useful surface and thermal power are not very significant.

TPE, in the simplest terms, represents the difference between the energy input (primary energy) and energy output (useful energy). $CO_2$ emissions are related to the energy efficiency of the building's heating and cooling systems, as demonstrated by Ye et al. [73] in office buildings. The results show a direct relationship between energy efficiency and $CO_2$ emissions based on the conditions of efficiency systems or renewable energy use, i.e. solar thermal systems. Opaque enclosures and holes, and skylights are related to the building design and construction. Wide-open walls or high ceilings in the building design could lead to heat losses.

The model shown entails the information necessary to reproduce the main parameters of the certification, the construction and the energy system [6]. This situation defines a new challenge in the energy performance in buildings an efficient design to avoid thermal flows or losses, which reduce useful energy for thermal comfort. The principle underlying the TPE model is to make the energy efficiency of buildings be accompanied by recommendations or how to improve it. On the one hand, it shows the potential improvements in the behaviour of the energy system and on the other hand, it acts on the building structure.

Nearly zero-energy buildings (NZEB) have very high energy performance with the contribution of advanced energy systems. Energy efficiency and management could become a key strategy to address sustainable and advanced energy systems. It is a challenge to use mathematical models based on the analysis of parameters of modern energy conversion systems in today's buildings.

To conclude, this study has allowed to foretell the thermal power efficiency (dependent variable), with significant accuracy and highly satisfactory outcome, on the basis of the real data observed employing the GPR–relied technique.

Fig. 5 shows the observed and predicted thermal power efficiency values and a 95% confidence interval for the predicted values. Fig. 6 shows the observed vs. the predicted values.

**Fig. 5** Observed and predicted thermal power efficiency values for the testing set employing DE/GPR–relied approach ($R^2 = 0.9687$) with the 95% confidence interval

**Fig. 6** Observed vs. predicted thermal power efficiency values for the testing set employing DE/GPR–relied approach ($R^2 = 0.9687$)

The GPR methodology with the DE optimiser (i.e., DE/GPR–relied model) proved to be a robust and effective approach to tackle this nonlinear regression problem.

## 4 Conclusions

The Energy Performance of Buildings Directive requires all new buildings to be nearly zero-energy by the end of 2020. In this context, the European challenge for energy in advanced thermal systems consists in the implementation of machine learning on advanced energy conversion systems in buildings in order to improve their energy performance. Relied on the former results, several core discoveries of this study can be drawn and indicated as follows:

- First of all, it is important to note that analytical models currently used to foretell the TPE in the energy systems from the observed values are not accurate enough because they greatly simplify a highly nonlinear complex problem. Consequently, the use of machine learning methods as the novel hybrid DE/GPR–relied approach employed in this study has revealed itself as the best choice to estimate the thermal power efficiency accurately from energy performance certificates at residential dwellings for energy management.
- In the second place, the hypothesis that the identification of thermal power efficiency can be determined accurately by means of a hybrid DE/GPR–relied approach in residential buildings has also been validated.
- Thirdly, the application of this GPR–relied methodology to the complete experimental dataset belonging to the thermal power efficiency resulted in a satisfactory coefficient of determination whose value was 0.9687.

- In fourth place, the ranking (or order of importance) of the predictive variables entailed in the estimation of the thermal power efficiency from energy performance certificates at residential dwellings was established. This is one of the principal core conclusions in this research about energy management. The $CO_2$ emissions (ECO2) in particular must be taken into account as the most important issue in the forecasting of TPE in the energy systems. On this matter, it is also noteworthy to emphasize the principal task of the opaque enclosures and holes and skylights in the obtained thermal power efficiency outcome.
- Conclusively, the principal role of the accurate hyperparameters determination in the GPR–relied methodology about the regression performance carried out for thermal power efficiency is established. The calculation of these hyperparameters was successfully carried out here using the heuristic optimizer known as differential evolution (DE).

To sum up, this procedure could be successfully applied to other energy performance certificates on either the same or different types of residential buildings. However, it is usually mandatory to consider individual characteristics for each dwelling and experiment. Hence, it is possible to conclude that the DE/GPR–relied method is a robust useful answer to the nonlinear problem of the estimation of the thermal power efficiency from energy performance certificates at residential houses.

**Acknowledgements** The authors gratefully acknowledge the computational help supplied by the Department of Mathematics at University of Oviedo as well as financial help of the Research Projects PGC2018-098459-B-I00 and FC-GRUPIN-IDI/2018/000221, both partial funding from European Regional Development Fund (ERDF). In this sense, the authors acknowledge the collaboration based on the Research Project FUO-118-19. Likewise, it is mandatory to thank Anthony Ashworth their revision of English grammar and spelling of this article.

**Compliance with ethical standards**

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

## References

1. Bourdeau M, Zhai XQ, Nefzaoui E, Guo X, Chatellier P (2019) Modelling and forecasting building energy consumption: a review of data-driven techniques. Sustain Cities Soc 48:101533

2. Harish VSKV, Kumar A (2016) A review on modeling and simulation of building energy systems. Renew Sust Energ Rev 56:1272–1292

3. European Commission (2002) Directive 2002/91/EC of the European parliament and of the council of 16 December 2002 on the energy performance of buildings, Official Journal of the European Communities

4. Paredes-Sánchez BM, Paredes-Sánchez JP, García Nieto PJ (2020) Energy multiphase model for biocoal conversion systems by means of a nodal network. Energies 13:2728–2740

5. Paredes-Sánchez JP, Conde M, Gómez MA, Alves D (2018) Modelling hybrid thermal systems for district heating: A pilot project in wood transformation industry. J Clean Prod 194:726–734

6. Li Y, Kubicki S, Guerriero A, Rezgui Y (2019) Review of building energy performance certification schemes towards future improvement. Renew Sust Energ Rev 113:109244

7. Khayatian F, Sarto L, Dall'O' G (2016) Application of neural networks for evaluating energy performance certificates of residential buildings. Energy Build 125:45–54

8. Son H, Kim C (2015) Early prediction of the performance of green building projects using pre-project planning variables: data mining approaches. J Clean Prod 109:144–151

9. Melo AP, Versage RS, Sawaya G, Lamberts R (2016) A novel surrogate model to support building energy labelling system: A new approach to assess cooling energy demands in commercial buildings. Energy Build 131:233–247

10. Hensen JLM, Lamberts R (2019) Building performance simulation for design and operation. Routledge, Taylor & Francis Group, New York

11. de Wilde P (2018) Building performance analysis. Wiley-Blackwell, New York

12. Rasmussen CE (2003) Gaussian processes in machine learning: summer school on machine learning. Springer, Berlin

13. Ebden M (2015) Gaussian processes: a quick introduction. https://arxiv.org/pdf/1505.02965.pdf. Accessed 27 May 2020

14. Dym H, McKean HP (2008) Gaussian processes, function theory, and the inverse spectral problem. Dover, New York

15. Storn R, Price K (1997) Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 11:341–359

16. Price K, Storn RM, Lampinen JA (2005) Differential evolution: a practical approach to global optimization. Springer, Berlin

17. Feoktistov V (2006) Differential evolution: in search of solutions. Springer, New York

18. Rocca P, Oliveri G, Massa A (2011) Differential evolution as applied to electromagnetics. IEEE Antennas Propag 53(1):38–49

19. Man K–F, Tang K–S, Kwong S (1999) Genetic algorithms: concepts and designs. Springer, New York

20. Haupt RL, Haupt SE (2004) Practical genetic algorithms. Wiley-Interscience, New York

21. Goldberg DE (2008) Genetic algorithms in search, optimization and machine learning. Dorling Kindersley Pvt Ltd, London, UK

22. Eiben AE, Smith JE (2015) Introduction to evolutionary computing. Springer, New York

23. Kramer O (2017) Genetic algorithm essentials. Springer, Berlin

24. Matthies H, Strang G (1979) The solution of nonlinear finite element equations. Int J Numer Meth Eng 14(11):1613–1626

25. Nocedal J (1980) Updating quasi-Newton matrices with limited storage. Math Comput 35(151):773–782

26. Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. Math Program 45:503–528

27. Byrd RH, Lu P, Nocedal J, Zhu C (1994) A limited-memory algorithm for bound constrained optimization. SIAM J Sci Comp 16:1190–1208

28. Zhu C, Byrd RH, Lu P, Nocedal J (1997) Algorithm 778: L–BFGS–B: Fortran subroutines for large-scale bound-constrained optimization. ACM T Math Software 23(4):550–560

29. Rao SS (2009) Engineering optimization: theory and practice. Wiley, New York

30. Nesterov Y (2018) Lectures on convex optimization. Springer, Berlin

31. Rasmussen CE, Williams CKI (2005) Gaussian processes for machine learning. The MIT Press, Cambridge, MA, USA

32. Duan Y, Cooling C, Ahn JS, Jackson C, Flint A, Eaton MD, Bluck MJ (2019) Using a Gaussian process regression inspired method to measure agreement between the experiment and CFD simulations. Int J Heat Fluid Fl 80:108497

33. Wang S, Zhu L, Fuh JYH, Zhang H, Yan W (2020) Multi-physics modeling and Gaussian process regression analysis of cladding track geometry for direct energy deposition. Opt Laser Eng 127:105950

34. Akhlaghi YG, Zhao X, Shittu S, Badiei A, Cattaneo MEGV, Ma X (2019) Statistical investigation of a dehumidification system performance using Gaussian process regression. Energ Buildings 202:109406

35. Alghamdi AS, Polat K, Alghoson A, Alshdadi AA, Abd El-Latif AA (2020) Gaussian process regression (GPR) based non-invasive continuous blood pressure prediction method from cuff oscillometric signals. Appl Acoust 164:107256

36. Li X, Yuan C, Li X, Wang Z (2020) State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression. Energy 190:116467

37. Zeng A, Ho H, Yu Y (2020) Prediction of building electricity usage using Gaussian Process Regression. J Build Eng 28:101054

38. Ambrogioni L, Maris E (2019) Complex-valued Gaussian process regression for time series analysis. Signal Process 160:215–228

39. Cai H, Jia X, Feng J, Li W, Hsu Y, Lee J (2020) Gaussian Process Regression for numerical wind speed prediction enhancement. Renew Energ 146:2112–2123

40. Gao A, Liao W (2019) Efficient gravity field modeling method for small bodies based on Gaussian process regression. Acta Astronaut 157:73–91

41. Gonçalves IG, Echer E, Frigo E (2020) Sunspot cycle prediction using warped Gaussian process regression. Adv Space Res 65(1):677–683

42. Sarkar D, Contal E, Vayatis N, Dias F (2016) Prediction and optimization of wave energy converter arrays using a machine learning approach. Renew Energ 97:504–517

43. Zhang J, Taflanidis AA, Scruggs JT (2020) Surrogate modeling of hydrodynamic forces between multiple floating bodies through a hierarchical interaction decomposition. J Comput Phys 408:109298

44. Zhao H, Magoulès F (2012) A review on the prediction of building energy consumption. Renew Sustain Energy Rev 16:3586–3592

45. Ahmad T, Chen H (2020) A review on machine learning forecasting growth trends and their real-time applications in different energy systems. Sustain Cities Soc 54:102010

46. Jovanovic RZ, Sretenovic AA, Zivkovic BD (2015) Ensemble of various neural networks for prediction of heating energy consumption, Energy Build 94:189–199

47. Rampazzo M, Lionello M, Beghi A, Sisti E, Cecchinato L (2019) A static moving boundary modelling approach for simulation of indirect evaporative free cooling systems. Appl Energ 250:1719–1728

48. Chou JS, Bui DK (2014) Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energ Build 82:437–446

49. Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, Saidur R (2014) A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 33:102–109

50. Aydinalp-Koksal M, Ugursal VI (2008) Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. Appl Energ 85:271–296

51. Amasyali K, El-Gohary NM (2018) A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 81:1192–1205

52. Yoon YR, Moon HJ (2018) Energy consumption model with energy use factors of tenants in commercial buildings using Gaussian process regression. Energy Build 168:215–224

53. Gray FM, Schmidt M (2016) Thermal building modelling using Gaussian processes. Energy Build 119:119–128

54. Mustapa RF, Dahlan NY, Yassin AIM, Nordin AHM (2020) Quantification of energy savings from an awareness program using NARX-ANN in an educational building. Energy Build 215:109899

55. Asturian Energy Foundation (FAEN) (2020) Technical report. http://www.faen.es/ceee/estadisticas/ceee_estadisticas_municipios.html. Accessed 29 May 2020

56. Spanish Institute for Diversification and Energy Saving (IDAE) (2019) Technical software. https://energia.gob.es/desarrollo/EficienciaEnergetica/CertificacionEnergetica/DocumentosReconocidos/Paginas/procedimientos-certificacion-proyecto-terminados.aspx. Accessed 26 May 2020

57. Rychlik I, Johannesson P, Leadbetter MR (1997) Modelling and Statistical Analysis of ocean-wave data using transformed Gaussian processes. Mar Struct 10(1):13–47

58. Bishop CM (2011) Pattern recognition and machine learning. Springer, New York

59. Li M, Sadoughi M, Hu Z, Hu C (2020) A hybrid Gaussian process model for system reliability analysis. Reliab Eng Syst Safe 197:106816

60. Daemi A, Kodamana H, Huang B (2019) Gaussian process modelling with Gaussian mixture likelihood. J Process Contr 81:209–220

61. Ciaburro G (2017) MATLAB for machine learning. Packt Publishing, Birmingham, UK

62. Lantz B (2019) Machine learning with R: expert techniques for predictive modeling. Packt Publishing, Birmingham, UK

63. Simon D (2013) Evolutionary optimization algorithms. Wiley, New York

64. Yang X, Cui Z, Xiao R, Gandomi AH, Karamanoglu M (2013) Swarm intelligence and bio-inspired computation: theory and applications. Elsevier, London

65. Liu J, Lampinen J (2002) On setting the control parameter of the differential evolution method. In: Proceedings of the 8th International Conference on Soft Computing, MENDEL, Brno, Czech Republic, pp 11–18

66. Knafl GJ, Ding K (2016) Adaptive regression for modeling nonlinear relationships. Springer, Berlin

67. McClave JT, Sincich TT (2016) Statistics. Pearson, New York

68. GPy (2014) A Gaussian process framework in python. http://github.com/SheffieldML/GPy. Accessed 25 May 2014

69. Stone JV (2016) Bayes' rule with python: a tutorial introduction to Bayesian analysis. Sebtel Press, London

70. Seeger M (2000) Bayesian model selection for support vector machines, Gaussian processes and other kernel classifiers. In: NIPS'99 Proceedings of the 12th

International Conference on Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, vol. 12, pp 603–609

71. Piironen J, Vehtari A (2016) Projection predictive model selection for Gaussian processes. In: IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE Publisher, Vietri sul Mare, Italy, pp 1–6

72. Paananen T, Piironen J, Andersen MR, Vehtari A (2019) Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution. In: Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), Proceedings of Machine Learning Research (PMLR), Naha, Okinawa, Japan, pp 1743–1752

73. Ye H, Ren Q, Hu X, Lin T, Shi L, Zhang G, Li X (2018) Modeling energy-related $CO_2$ emissions from office buildings using general regression neural network. Resour Conserv Recy 129:168–174

**Table 1** Set of physical predictive variables of the operation used in this study: names, means and standard deviations

| Predictive variables | Name of the variable | Mean | Standard deviation |
|---|---|---|---|
| Useful surface ($m^2$) | US | 1438.15 | 2798.76 |
| Thermal power (kW) | TP | 187.03 | 309.25 |
| $CO_2$ emissions (kg $CO_2$ /$m^2$ year) | ECO2 | 633.71 | 1694.45 |
| Primary energy consumption (kWh/$m^2$ year) | PEC | 10914.08 | 29228.41 |
| Opaque enclosures ($m^2$) | OE | 591.38 | 2062.79 |
| Holes and skylights ($m^2$) | HS | 79.87 | 20.48 |

**Table 2** Initial ranges of the three hyperparameters of the GPR–relied models fitted in this study

| GPS hyperparameters | Lower limit | Upper limit |
|---|---|---|
| RBF kernel variance $\sigma_f^2$ | $10^{-3}$ | $10^{5}$ |
| RBF kernel lengthscale $\ell$ | $10^{-3}$ | $10^{5}$ |
| Gaussian noise variance $\sigma_n^2$ | $10^{-3}$ | $10^{5}$ |

**Table 3** Minimum, mean, maximum and standard deviation values for the objective function (–log likelihood) for the 50 runs of the optimisation stage for the DE, GA and L-BFGS algorithms

| Optimizer | min. | mean | max. | Standard deviation |
|---|---|---|---|---|
| DE | 101 | 101 | 101 | $5.8\times10^{-9}$ |
| GA | 101 | 108 | 142 | 9.24 |
| L-BFGS | 101 | 103 | 104 | 0.61 |

**Table 4** *p*-values of the Mann-Whitney rank test comparing the sets 50 objective functions values obtained for each of the different hybrid models using DE, GA and L-BFGS as parameter tuning optimisation algorithms

| | *p*-value |
|---|---|
| DE vs. GA | $2.2\times10^{-16}$ |
| DE vs. L-BFGS | $1.94\times10^{-15}$ |
| GA vs. L-BFGS | $6.6\times10^{-7}$ |

**Table 5** Optimal hyperparameters of the best-fitted GPR–relied model encountered with the DE, optimizer: variance $\sigma_f^2$ and lengthscale $\ell$ for the RBF kernel, the Gaussian noise variance $\sigma_n^2$, and the corresponding objective function (-log likelihood) value for the optimized model for the training set

| Parameter | $\sigma_f^2$ | $\ell$ | $\sigma_n^2$ | Objective fun. value |
|---|---|---|---|---|
| Thermal power efficiency (%) | 15.69 | 1.28 | 0.0747 | 101 |

**Table 6** Coefficient of determination ($R^2$), coefficient of correlation *r* and root mean square error *RMSE* for the testing dataset for the DE/GPR regression model

| Model | $R^2$ | *r* | *RMSE* |
|---|---|---|---|
| DE/GPR | 0.9687 | 0.9867 | 7.116 |

**Table 7** Relative importance of the input variables as stated in the DE/GPR–relied approach for the TPE according to VLM method

| Predictive variable | Relative importance |
|---|---|
| $CO_2$ emissions (kg $CO_2$ / $m^2$ year) | 1 |
| Opaque enclosures ($m^2$) | 0.1405 |
| Holes and skylights ($m^2$) | 0.0353 |
| Primary energy consumption (kWh/$m^2$ year) | 0.0167 |
| Useful surface ($m^2$) | 0.0053 |
| Thermal power (kW) | 0.0048 |

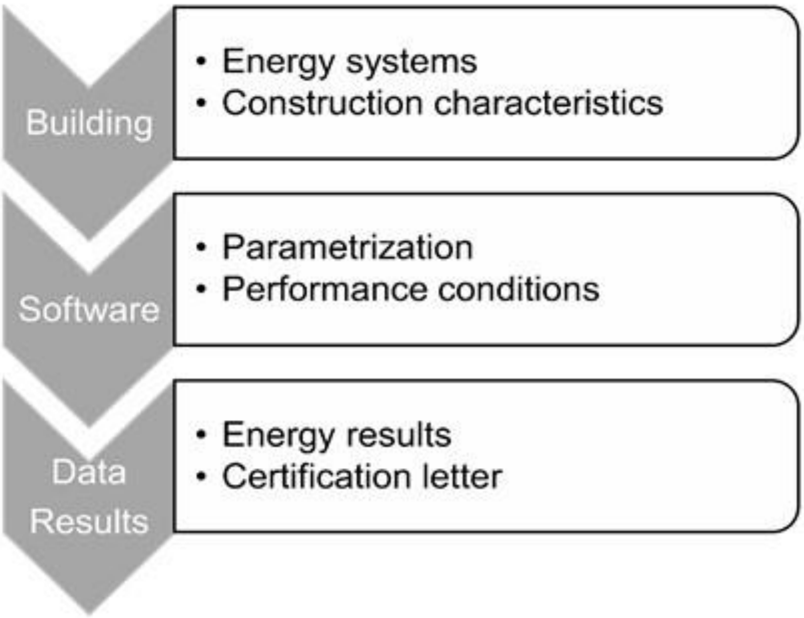


**Fig. 1** Diagram of data collection


**Fig. 2** Example of a building and the parameters analysed in it as an energy system

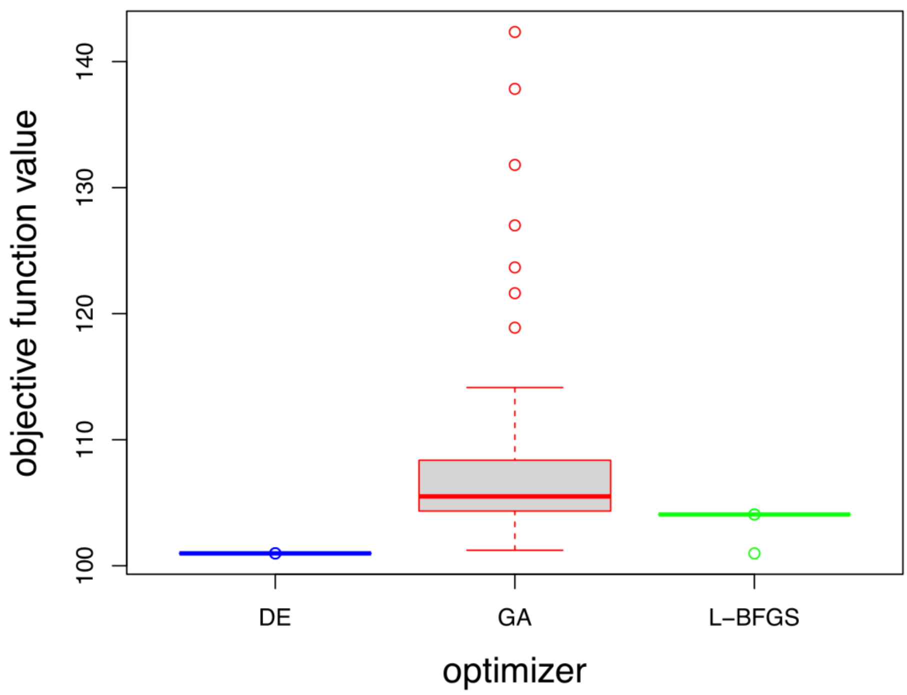


**Fig. 3** Boxplots for the values of the objective function (-log likelihood) obtained from running each of three optimisers fifty times

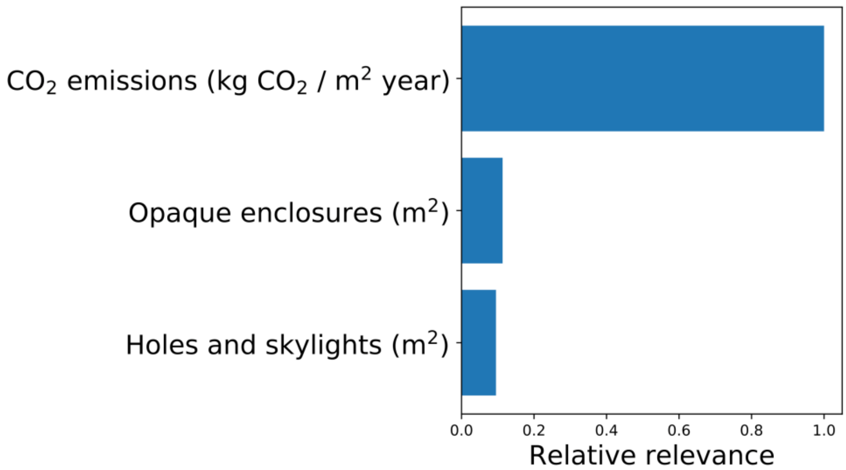


**Fig. 4** Relative importance of the input variables as stated in the DE/GPR–relied approach for the TPE according to VLM method

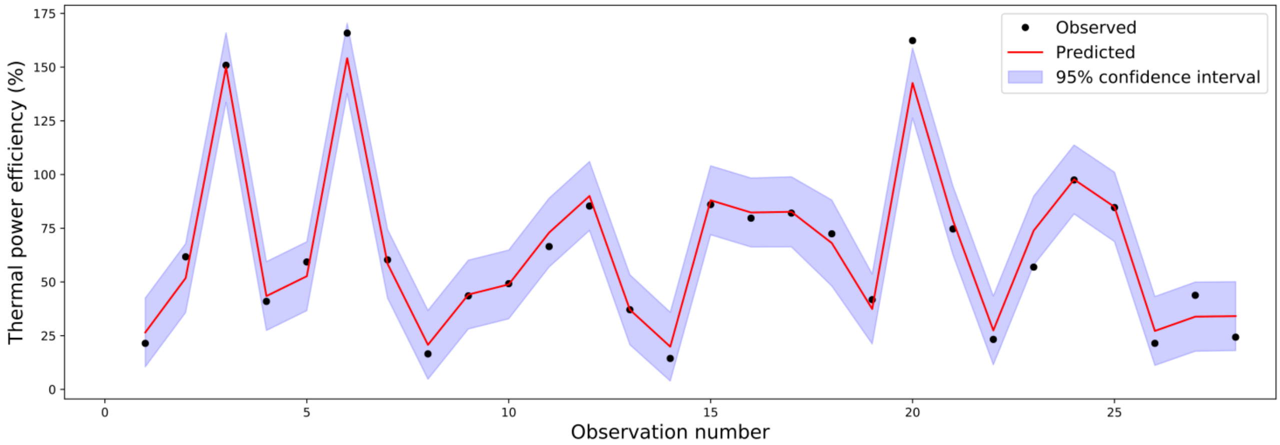


**Fig. 5** Observed and predicted thermal power efficiency values for the testing set employing DE/GPR–relied approach ( $R^2 = 0.9687$ ) with the 95% confidence interval

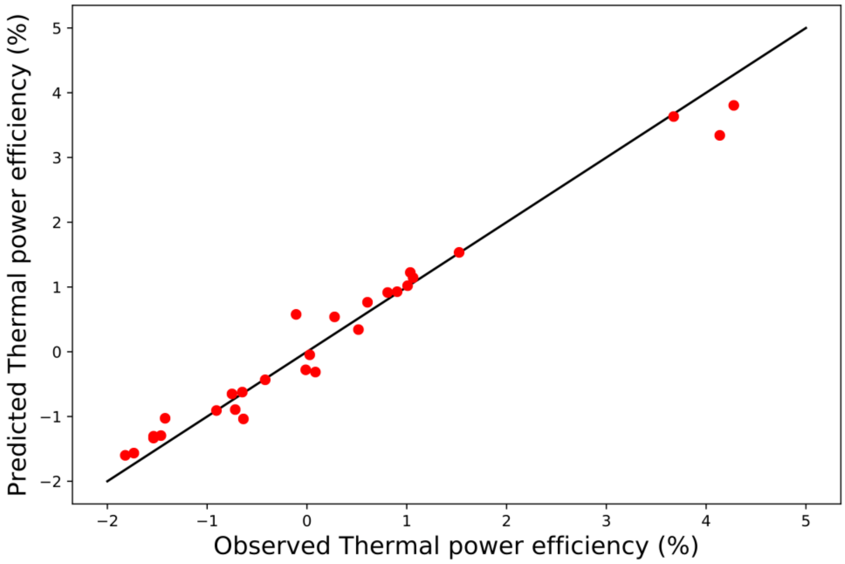


**Fig. 6** Observed vs. predicted thermal power efficiency values for the testing set employing DE/GPR–relied approach ( $R^2 = 0.9687$ )