ARTICLE IN PRESS

Renewable and Sustainable Energy Reviews xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM

ARTICLE INFO

Keywords:
Heating load
Cooling load
Energy efficient
Buildings
Multivariate
Regression spline
Extreme learning machine
Linear regression
Gaussian process regression
Neural network
Feed-forward
Radial basis function

ABSTRACT

Heating load and cooling forecasting are essential for estimating energy consumption, and consequently, helping engineers in improving the energy performance right from the design phase of buildings. The capacity of heating ventilation and air-conditioning system of the building contribute to the operation cost. Moreover, building being one of the sectors with heavy energy use, it is required to develop an accurate model for energy forecasting of building and constructing energy-efficient buildings. This paper explores different machine learning techniques for predicting the heating load and cooling load of residential buildings. Among these methods, we focus on advanced techniques like Multivariate Adaptive Regression Splines (MARS), Extreme Learning Machine (ELM) and a hybrid model of MARS and ELM along with a comparison of the results with those of more conventional methods like linear regression, neural network, Gaussian processes and Radial Basis Function Network. The MARS model is a non-parametric regression model that splits the data and fits each interval into a basis function and ELM is similar to a Single Layer Feed-forward Neural Network except that in ELM randomly assigned input weights are not updated. As an improvement, we have tried a hybrid model that uses MARS to evaluate the importance of every parameter in the prediction and these important parameters have been fed to the ELM to build hybrid model and it can be seen that this boosts the ELM performance to match up to the accuracy of MARS with lesser computation time. Finally, a comparative study examines the performances of the different techniques by measuring different performance metrics.

1. Introduction

Modern industries in the world are based on a foundation of hydrocarbons, metals, fuels and electricity. These three are intricately connected; each is accessible only if there is sufficient energy in the world to produce the other two. [1] Global consumption of energy for the year 2005 was about 500 exajoules (EJ) and most of which was supplied by fossil fuels. What is more important in terms of the consequences on daily human life, though, is not consumed in an absolute sense, but consumption per capita. Energy loads of constructions are how much energy your building needs. These demands can be provided by electricity, fuel, or by passive means. The concept of building loads involves navigation of a number of interrelated terms. Thermal loads are the quantity of heating and cooling energy that must be added or removed from the building to keep people comfortable. Thermal loads or heating load come from heat transfer from within the building during its operation and between the building and the external environment. These thermal loads can be translated to heating loads (when the building is too cold) and cooling loads (when the building is too hot). These heating and cooling loads aren't just about temperature (sensible heat), they also include moisture control (latent heat). Heating and cooling loads are met by the building's HVAC system 'Heating Ventilation and Air Conditioning'(HVAC) [2], which uses energy to add or remove heat and condition the space. Lower thermal loads indicate that, relatively, the dwelling will require less heating and cooling to maintain comfortable conditions. Lower thermal loads do not necessarily correspond to lower electricity usage. In practice, the heating and cooling loads may be handled by heating or air-conditioning equipment.

The depletion of energy resources is a complex phenomenon [3,4]. Depletion is not a process of simply "running out" of a resource, but instead is a process of technological change and adaptation. This adaptation can lead to the development of new resources, that are often lower quality, costlier to process, or more difficult to access (e.g. deep-water oil resources, shale oil, or bitumen) [4,5]. Keeping these facts in mind, the notion of energy efficiency becomes more important in the current world scenario. According to the International Energy Agency (IEA), energy efficiency is a way of managing and restraining the growth in energy consumption. Something is more energy efficient if it delivers more services for the same energy input, or the same services for less energy input. This paper aims to cater to the imperative need for energy efficiency by proposing a model that predicts the energy performance of buildings and thermal comfort of its occupants. A lot of effort is being invested into this through building simulation tools like DesignBuilder, Energy Plus, ESP-r, DOE-2 [6] to name a few. This ensures that an informed decision can be made while designing a building, consequently leading to more energy efficient residential buildings. Statistics from International Energy Agency (IEA) reports from 2016 show that consume almost 33% of all final energy, over 50% of global energy and consequently contribute to almost 30% of global carbon

Abbreviations: HVAC, Heating Ventilation and Air-Conditioning; MARS, Multivariate Adaptive Regression Spline; ELM, Extreme Learning Machine; IEA, International Energy Agency; GCV, Generalized Cross Validation; RMSE, Root Means Square Error; MAPE, Mean Absolute Percentage Error; WMAPE, Weighted Mean Absolute Percent Error; CL, Cooling Load; SVM, Support Vector Machine; RBF, Radial Basis Function; LR, Logistic Regression; GPR, Gaussian Process Regression

http://dx.doi.org/10.1016/j.rser.2017.05.249
Received 17 April 2016Received in revised form 26 April 2017Accepted 26 May 2017
1364-0321/

Nomenclature

β_0, β_m	Constant coefficients of Basis Functions
$\|\|\omega\|\|$	output weight

emissions [7]. With the trend of increase in population and living standards, this could result in a steep increase in energy use in buildings and makes it imperative to find ways to make building sector energy efficient and sustainable [8]. If the estimation of the operational energy happens at the design stage of the building, the efficiency of the energy consumption can be improved considerably.

This paper mainly aims at exploring advanced prediction models like multivariate adaptive regression splines [9] and extreme learning machines to predict the heating load and cooling load value for building based on certain parameters. Neural networks have for long been an effective prediction tool, used in situations where dependent and independent variables may exhibit non-linear relationship. However, they are criticized for the high computation time needed for their training process. Thus, we use extreme learning machine which has lesser computational complexity and runs faster than traditional algorithms like SVM [10] and LS-SVM [11] as can be seen from many recent studies. Multivariate regression splines have also been a recent favourite of researchers as a powerful tool to model data with non-linear relationship between dependent and independent variables. Additionally, they also give the relative importance of the independent variables in the prediction of the dependent variable in an easily interpretable manner. This makes it a desirable choice to provide support to the prediction models proposed in this paper wherein MARS is used in the first stage to determine the degree of participation of each input variable in the building of the model and providing only the variables with high relative importance as an input in the second stage to the ELM. This optimizes the topology of the ELM [12] model by removing variables of lesser relevance and reducing the time needed to build the ELM model [12] making its predictions at par with that of MARS but with faster computation time which makes a huge difference while handling large number of data-points.

In the last part of the paper, the effectiveness and feasibility of the mentioned models have been examined. The performances of conventional models like neural network, Gaussian processes, linear regression as well as MARS and ELM have been compared by measuring different performance metrics. The experiments use Energy Efficiency dataset. The energy dataset has been split into the training and the test data with 70% and 30% respectively. MARS and Hybrid models show better result than the others. The proposed models in this paper will help engineers for estimating thermal energy consumption of buildings and build more energy efficient ones.

2. Theoretical background

2.1. Multivariate Adaptive Regression Splines (MARS)

Friedman [9] proposed MARS as a non-linear and non-parametric method for predicting the values of a continuous dependent variable using a set of predictors. It is a method of flexible regression modelling for high dimensional data which is motivated by a recursive partitioning approach to regression. Unlike the recursive partitioning, it produces continuous models with continuous derivatives. The relationship is derived from a set of coefficients and basis functions that are driven from the dataset without making any assumptions about any underlying functional relationship between the variables. In MARS, the process of model building happens in a forward-backward way. In the forward pass, basis functions are added in pairs to the model until the change in residual error is too small to continue or the maximum number of terms has been reached. This usually builds an overfit model and hence the backward pass is necessary to improve the generalization ability by pruning the least effective term in the model until the best sub-model is found.

The basis function of MARS is dependent on a segment function which is known as spline function. It divides the dataset into separate piecewise linear segments of differing slope value called splines. Spline functions are formed by joining polynomials of degree n together at fixed points called *knots*. That is, we divide the interval extending from lower limit t_L to upper limit t_U over which we wish to approximate a curve into L+1 sub-intervals separated by L interior boundaries ξ_l called knots, or sometimes *breakpoints*. The resulting model with handles both linear and non-linear behaviour consists of splines that are connected smoothly with the knots while each piecewise curve is called the basis function. The spline function is defined as [13]:

$$h1(x) = (t - x)_+ = \begin{cases} t - x \ if \ t > x \\ 0, \ Otherwise \end{cases} \tag{1}$$

$$h2(x) = (x - t)_+ = \begin{cases} x - t \ if \ x > t \\ 0, \ Otherwise \end{cases} \tag{2}$$

While developing MARS we use functions from a collection I, which includes Eq. (2) as a basis functions if there is no repeated data in the set, and its interaction effects.

$$I = \{(X_i - t)_+, (t - X_j)_+\}, t \in \{x_{1j}, x_{2j}, ..., x_{nj}\}, j = 1, ..., p. \tag{3}$$

The MARS model will have the following structure:

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X) \tag{4}$$

in which, $h_m(X)$ is one or more function from set I, M in the number of functions in the model which will be identified after forward stage. β_m is determined through the minimum sum of squared errors from the h_m functions.

If we describe MARS using a single input variable x, by considering the observed data as $((x_1, y_1),...,(x_n,y_n))$ where (x_i, y_i) represent the ith observation, at a primary stage this model will have only a y-intercept, that is:

$$\hat{f}_1(x) = \hat{\beta}_0 \tag{5}$$

Taking $\widehat{\beta_0} = \widehat{Y}$ as model 1, model 2 is then created by selecting the model which includes the minimum sum of squares error among the other models which are:

$$\hat{f}_{21}(X) = \hat{\beta}_0 + \hat{\beta}_1(X - x_1)_+ + \hat{\beta}_2(x_1 - X)_+ \tag{6}$$

$$\hat{f}_{22}(X) = \hat{\beta}_0 + \hat{\beta}_3(X - x_2)_+ + \hat{\beta}_4(x_2 - X)_+ \tag{7}$$

From these models, if we select the model 2 as:

$$\hat{f}_2(X) = \hat{\beta}_0 + \hat{\beta}_3(X - x_2)_+ + \hat{\beta}_4(x_2 - X)_+ = \hat{\beta}_0 + \hat{\beta}_3 h_3(X) + \hat{\beta}_4 h_4(X) \tag{9}$$

We move on to generating model 3 by adding more basis functions, which further reduce the sum of squares error, to model 2. Thus model 3 gets selected form the models:

$$\hat{f}_{31}(X) = \hat{\beta}_0 + \hat{\beta}_1 h_3(X) + \hat{\beta}_2{}^* h_4(X) + \hat{\beta}_3{}^* h_1(X) + \hat{\beta}_4{}^* h_2(X) \tag{10}$$

$$\hat{f}_{3(n-1)}(X) = \ldots\ldots\ldots\ldots\ldots\ldots + \hat{\beta}_3{}^* h_{2n-1}(X) + \hat{\beta}_4{}^* h_{2n}(X) \tag{11}$$

Every time a new model is developed a new model coefficient β is estimated as we select a model that has the minimum sum of squares errors $\sum_{i=1}^{n}(y_i - \hat{f}(x_i))$. This goes on until we have added k basis functions. This causes overfitting at the end of the first stage wherein even though the developed model can model the data it does this through overestimation. Thus, the pruning takes place in the next stage to solve this problem by removing basis functions from the model. The basis function whose removal increases the sum of squares error as little as possible are removed and this goes on till all basis functions except y-intercept is removed from the model.

After the completion of this pruning process, 2k-2 models are developed each of which is a contender for the final model. This is determined by calculating the generalized cross-validation (GCV) criteria for each. The GCV is a measure of goodness of fit that takes into account not only the residual error but the model complexity as well. The GCV can be defined as:

$$\mathrm{GCV} = \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{(1 - \frac{C}{N})^2} \tag{12}$$

with C= 1+ cd, where N is the number of cases in the dataset, d is the effective degrees of freedom, which is equal to the number of independent basis functions. The quantity c is the penalty for adding a basis function. Experiments have shown that the best value for C can be found somewhere in the range $2 < d < 3$. The model that has the lowest GCV among 2 K −2 will be selected as the final model and an estimation for MARS. Several applications of MARS in geosciences and agriculture can be found in much research in recent times [14–16].

2.2. Extreme Learning Machine (ELM)

ELM is a modification of a single-layer feed-forward network, which can be obtained by removing backpropagation from a multilayer perceptron [17]. It provides a non-linear model at the speed of a linear model. Unlike SFLN, ELM does not tune the weights using back propagation or any other iterative methods. Instead the weights are initialized analytically and hence are semi-random. Different from BP (Back Propagation) and SVM (Support Vector Machine) which consider multi-layer of networks as a black box, ELM handles both SLFNs and multi-hidden-layer of networks similarly. ELM considers multi-hidden-layer of networks as a white box and trained layer-by-layer. It is, however, different from Deep Learning which requires intensive tuning in hidden layers and hidden neurons, such that all the hidden neurons in hierarchical ELM as a whole are not required to be iteratively tuned. ELM theories show that hidden neurons although important need not be turned (for both SLFNs and multi-hidden-layer of networks), learning can simply be made without iteratively tuning hidden neurons [17]. Several applications of ELM can be found in many fields [18–20] like neurocomputing [21,22], medicine [23], energy estimation [24], cost forecasting [25] and more [26,27].

For generalized SFLN, having one output node the output function for ELM can be defined as:

$$f_L(x) = \sum_{i=1}^{N} \omega_i h_i(x) = h(x)\omega \tag{13}$$

where $\omega=[\omega_1,\ldots,\omega_N]^T$ is the vector of the output weights between the hidden layer of N nodes and the output node and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_L(\mathbf{x})]$ is the output (row) vector of the hidden layer with respect to the input **x.** $\mathbf{h}(\mathbf{x})$ actually maps the data from the d-dimensional input space to the N-dimensional hidden-layer ELM feature space H, and thus, $\mathbf{h}(\mathbf{x})$ is indeed a feature mapping. For the binary classification applications, the decision function of ELM is

$$f_L(\boldsymbol{x}) = sign(\boldsymbol{h}(\boldsymbol{x})\boldsymbol{\omega}) \tag{14}$$

While traditional learning algorithms tend to reach only the smallest training error, ELM tends to reach also the smallest norm of output weights along with it. According to Bartlett's theory, for feedforward neural networks reaching smaller training error, the smaller the norms of weights are, the better generalization performance the networks tend to have. We conjecture that this may be true to the generalized SLFNs where the hidden layer may not be neuron alike. ELM is to minimize the training error as well as the norm of the output weights.

Minimize: $\|\mathbf{H}\boldsymbol{\omega} - \mathbf{T}\|^2$ and $\|\boldsymbol{\omega}\|$

where **H** is the hidden-layer output matrix

$$\mathbf{H} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_m) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdot & \cdot & \cdot & h_N(x_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_1(x_M) & \cdot & \cdot & \cdot & h_N(x_M) \end{bmatrix}$$

Seen from $f_L(x)$, to minimize the norm of the output weights $\|\omega\|$ is actually to maximize the distance of the separating margins of the two different classes in the ELM feature space: $2/\|\omega\|$. The minimal norm least square method instead of the standard optimisation method was used in the original implementation of ELM [12].

$$\boldsymbol{\omega} = \mathbf{H}^{\dagger}\mathbf{T} \tag{15}$$

where $\mathbf{H}^{\dagger}$ is the *Moore–Penrose* generalized inverse of matrix H. Different methods can be used to calculate the Moore–Penrose generalized inverse of a matrix: orthogonal projection method, orthogonalization method, iterative method, and singular value decomposition (SVD). The orthogonal projection method can be used in two cases: when $\mathbf{H}^{\mathbf{T}}\mathbf{H}$ is non singular and $\mathbf{H}^{\dagger} = (\mathbf{H}^{\mathbf{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathbf{T}}$, or when $\mathbf{H}\mathbf{H}^{\mathbf{T}}$ is non singular and $\mathbf{H}^{\dagger} = \mathbf{H}^{\mathbf{T}}(\mathbf{H}\mathbf{H}^{\mathbf{T}})^{-1}$.

2.3. Hybrid model

In the hybrid model, we first build a predictive model using MARS taking all the 8 predictor variables. One of features of MARS is that it evaluates the degree of participation of the independent variables in predicting the target values as can be seen in Table 3. We exploit this feature to segregate the important variables from the unused variables. Then these important variables are taken as the input layer for the ELM in the hybrid model instead of taking all the variables blindly. Using prior information about the participation of variables in the prediction thus makes the process of building the prediction model using ELM more informed. The training of the network is done using different activation functions and by resetting a combination of network parameters like learning rate and momentum to get the best possible result. The predicted values and performance measures are tabulated in Table (5a) and (5b).

2.4. Gaussian process regression

Gaussian process regression has probabilistic characteristics in it. GPR is used to generalize multivariate distribution of input to unbounded-dimensional space using a Bayesian model to find posterior distributions [28,29].

$$y(m) = f[X(m)] + \quad \zeta(m) \tag{16}$$

Here, objective variable is (m) ; $\zeta(m)$~$N(0,\sigma^2)$ is Gaussian noise and σ^2 is variance. X(m) is the space of the regression in data space R^D. All input vector x has attached random variable f(x). In this work, GPR has used RBF kernel.

2.5. Linear regression

Linear regression is one of the simplest statistical method that models the relationship between to continuous variables by fitting a linear equation to the observed data [30]. In order to fit the model, first it is necessary to check if there is any relationship between the variables of interest which is done using a numerical variable called the correlation coefficient. A linear regression line is represented with the equation of the form Y = a + bX, where X is the independent variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0). In order to find the best fitted line one of the most common measures used is the least square errors, which is done by aiming the minimize the sum of the squares of the vertical deviation of every point from the line or the sum of square of the residuals.

2.6. Neural network

Inspired by the biological nervous system, the neural network tries to mimic the way the brain processes information, where every neuron takes a number of input and gives a particular output based on a firing rule. The input layer is representative of the raw information fed to the neural network. This information moves onto the hidden layers were the input value and the weights associated with each connection determines the output. Finally, we get an output based on an activation function. Neural networks are commonly used non-linear method of predicting data in many domains (Fig.1) [31].

2.7. Radial Basis Function Network

Radial Basis Function Network is an artificial neural network [32–34] used as a supervised algorithm for prediction. Here the basis function is

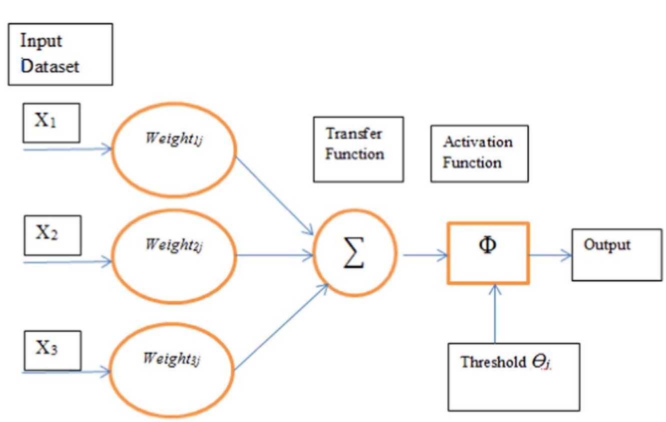

Fig. 1. Artificial Neural Network Architecture [31].

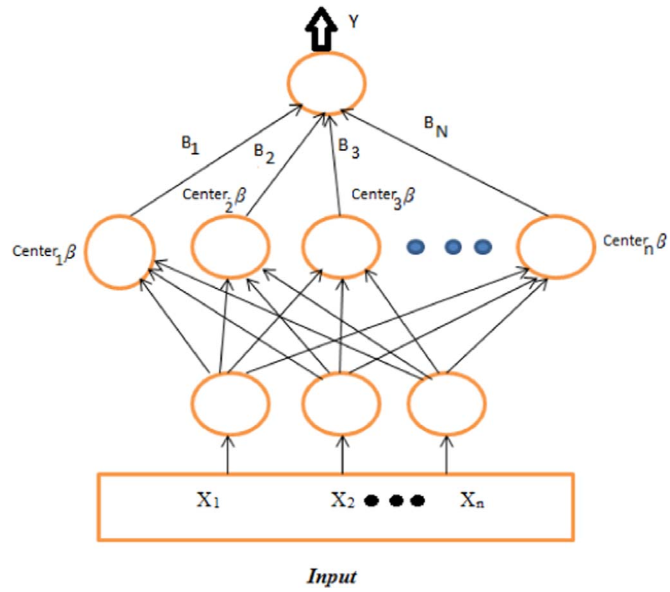

Fig. 2. RBF Network Architecture [32,33].

Table 1
List of dependent and independent variables in dataset.

TYPE	VARIABLE	DEFINITION
DEPENDENT VARIABLES	X1	Relative Compactness
	X2	Surface Area
	X3	Wall Area
	X4	Roof Area
	X5	Overall Height
	X6	Orientation
	X7	Glazing Area
	X8	Glazing Area Distribution
INDEPENDENT VARIABLE	Y1	Heating Load
	Y2	Cooling Load

adopted as K-means clusters. RBFNetwork uses a mapping as below,

$$f_s(x) = \lambda_0 + \sum_{i=1}^{n_s} \lambda_i \varphi(x - c_i) \tag{17}$$

Here, $x \in R^n$, is the input energy data in vector form.$\varnothing(.)$ maps R^n to R is a euclidean nomenclature. λ_i denotes weights and c_i known as centres of RBF (Fig. 2).

3. Materials and methods

3.1. Energy efficiency dataset and statistical analysis

To verify the feasibility and effectiveness of the hybrid model using MARS and ELM we take the energy efficiency dataset created by Angeliki Xifara and processed by Athanasios Tsanas (Oxford Centre for Industrial and Applied Mathematics, University of Oxford, UK). It has been generated by performing energy analysis using 12 different building shapes simulated in Ecotect. The buildings differ with respect to the glazing area, the glazing area distribution, and the orientation, amongst other parameters. Having simulated various settings as functions of the afore-mentioned characteristics, 768 building shapes have been obtained. Xifara and Tsanas, mentioned in their paper about HVAC regulations and that simulated data which they have generated with 12 different building shapes has been obtained by strictly following those regulations [2]. ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) is the leading international body in the HVAC field. The key intention of ASHRAE is to improve the arts and sciences of heating, cooling, ventilation, air conditioning, refrigeration and correlated human issues to serve the evolving needs of the public and ASHRAE members. Heating, ventilation and air conditioning (HVAC) works like a catalyst in controlling the climate inside the building. This HVAC system is responsible for all energy use in a building. Therefore, forecasting of heating and cooling load in building in the design phase itself will directly help to conserve energy.

The dataset comprises 768 samples and each sample in the dataset contains eight attributes (or features, denoted by X1...X8) and two responses (or outcomes, denoted by y1 and y2). The aim is to use the eight features to predict each of the two responses correctly. The independent and dependent variables have been summarized in Table 1:

We can see from the summary of the variables that all variables are continuous in nature and there are no missing values or outliers in any of the variables. Thus the data did not need too much pre-processing. Although we normalise the data before use to remove the amplitude variation and

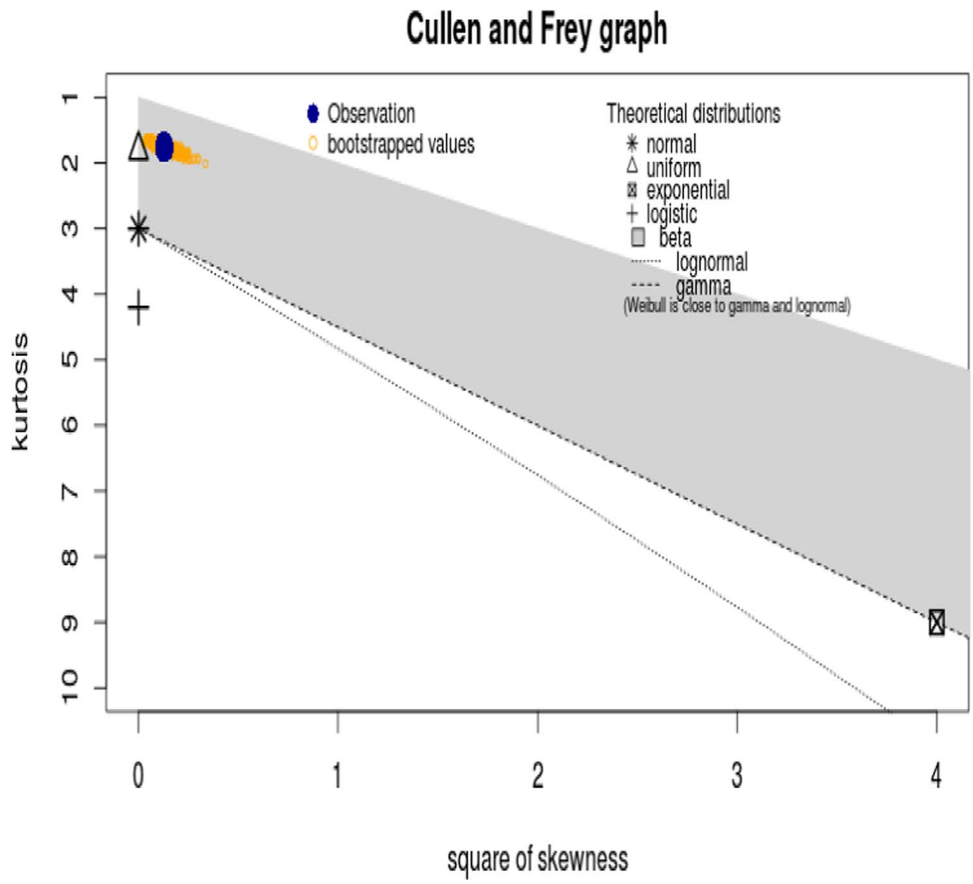

Fig. 3. Data Set analysis: Cullen and Frey graph.

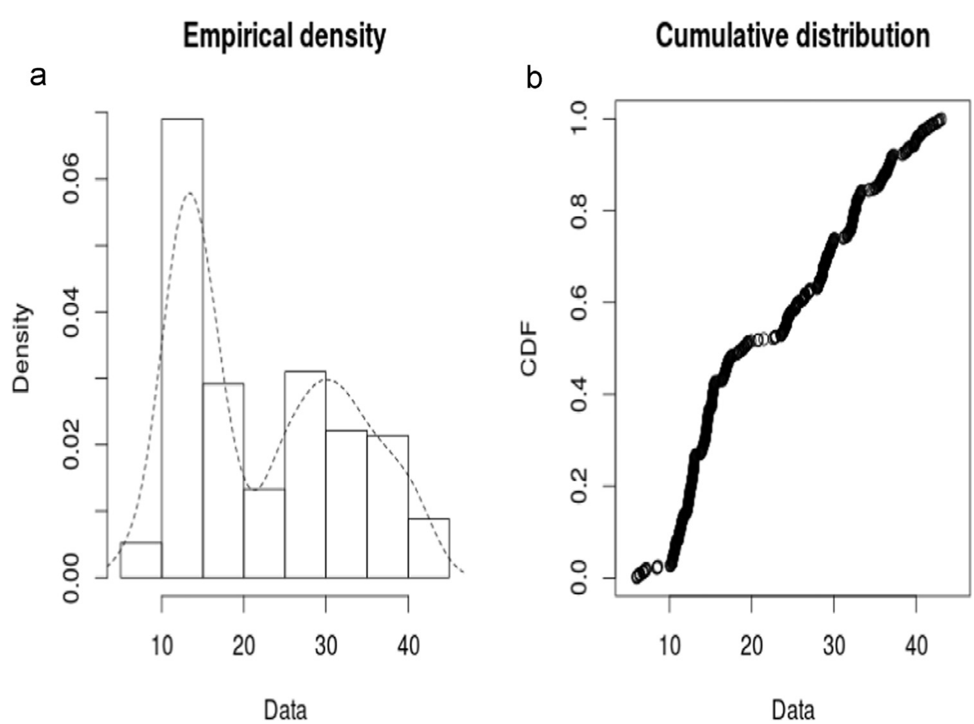

Fig. 4. (a & b) Empirical Density and Cumulative Distribution of the dataset.

focus on the underlying distribution. This will make the model less sensitive to scale of the different features making the data have a zero mean and a unit variance across each feature. We have used the z-score normalisation here which can be expressed as:

$$Xnorm,(i, j) = \frac{X(i, j) - mean(Xj)}{std(Xj)}$$

Table 2
Variable Summary.

VARIABLES	MIN	1ST QUARTILE	MEDIAN	MEAN	3RD QUARTILE	MAX
Relative Compactness	0.6200	0.6825	0.7500	0.7642	0.8300	0.9800
Surface Area	514.5	606.4	673.8	671.7	741.1	808.5
Wall Area	245.0	294.0	318.5	318.5	343.0	416.5
Roof Area	110.2	140.9	183.8	176.6	220.5	220.5
Overall Height	3.50	3.50	5.25	5.25	7.00	7.00
Orientation	2.00	2.75	3.50	3.50	4.25	5.00
Glazing Area	0.0000	0.1000	0.2500	0.2344	0.4000	0.4000
Glazing Area Distribution	0.000	1.750	3.000	2.812	4.000	5.000
Heating Load	6.01	12.99	18.95	22.31	31.67	43.10
Cooling Load	10.90	15.62	22.08	24.59	33.13	48.03

Table 3
Variable selection and Basis Function of MARS Model.

Variable Selection Results		Basis Functions	
Variable name	**Relative Importance**	**Equation Name**	**Equation**
Roof Area (X4)	100	BF1	Max (0, X1-0.527838)
Glazing Area (X7)	45.3	BF2	Max (0, X1-1.28414)
Surface Area (X2)	37.2	BF3	Max (0, X2 to −0.950301)
Relative Compactness (X1)	32.1	BF4	Max (0, −0.394027-X2)
MARS Prediction Function:		BF5	Max (0, X2 to −0.394027)
0.7422228 + 4.234267 * BF1 − 2.00807 * BF2 + 1.228276 * BF3 − 3.45402 * BF4 − 1.109239 * BF5 + 1.2453 * BF6 − 2.190511 * BF7 − 0.6199386 * BF8 − 1.678417 * BF9 − 0.8088267 * BF10 + 0.2216917 * BF11		BF6	Max (0, X2-0.996657)
		BF7	Max (0, X2-1.27479)
		BF8	Max (0, −0.655453-X4)
		BF9	Max (0, X4 to −0.655453)
		BF10	Max (0, −1.00867-X7)
		BF11	Max (0, X7 to −1.00867)

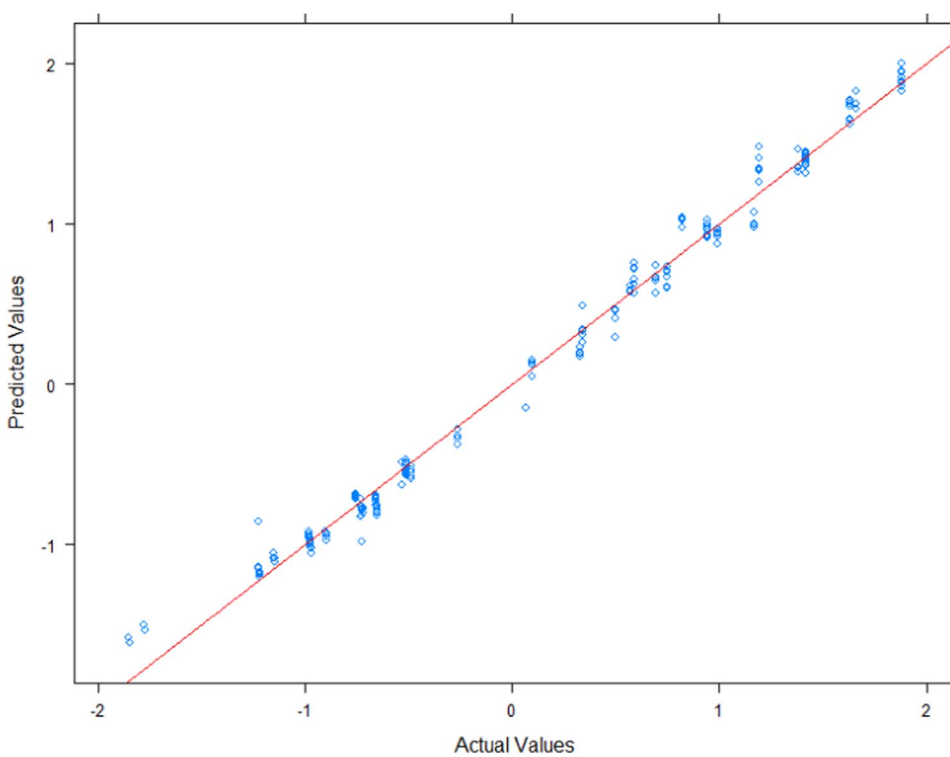

Fig. 5. Actual vs. Predicted graph obtained from test data using MARS.

It is visible from the Fig. 3, that observation falls into the beta distribution. The symmetry of the data or the lack of symmetry is measured by skewness. Kurtosis is a measure of whether the energy data are heavy-tailed or light-tailed relative to a normal distribution. Here, both the parameters' performances of Skewness and Kurtosis are average.

Fig. 4a depicts the density of heating load variable in comparison with other independent variables. It looks like a two peak rightly-skew distribution. Fig. 4b describes cumulative distribution function (CDF) for the selected heating load variable to examine how well the distribution fits data.

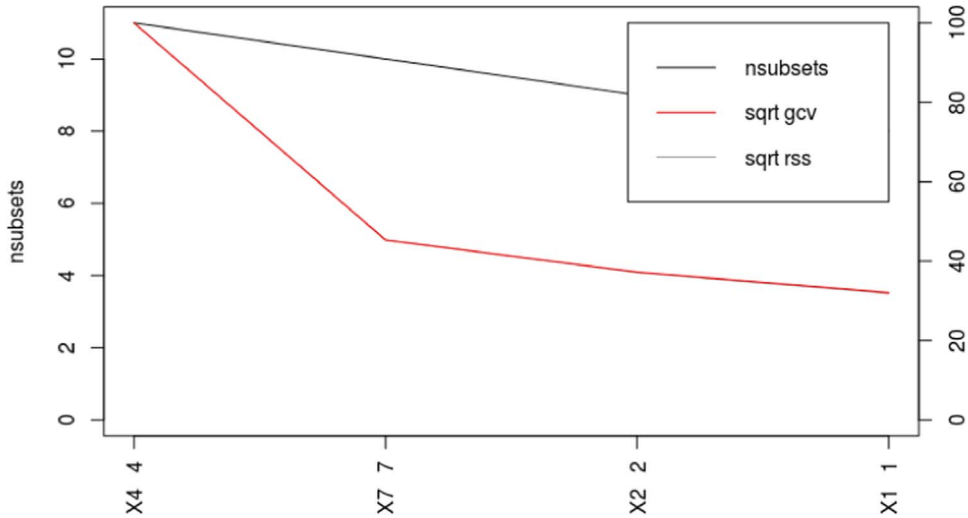

Fig. 6. Variable Importance while predicting heating load.

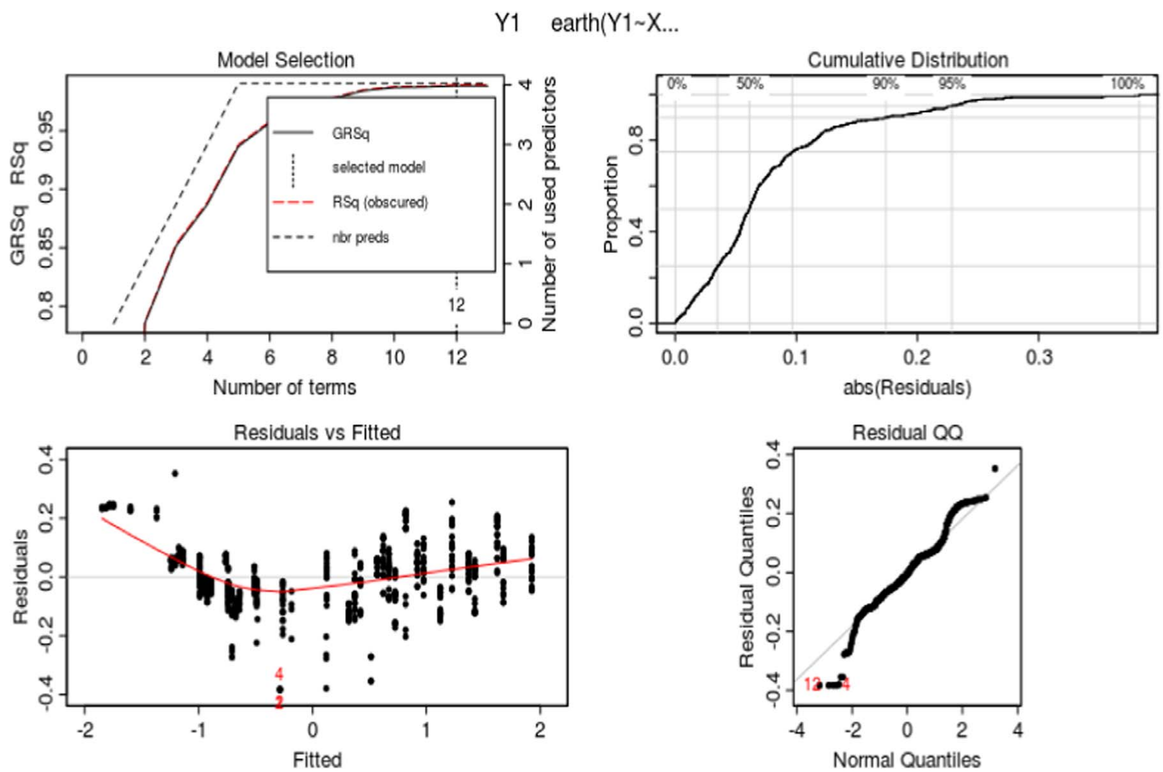

Fig. 7. Summary of the MARS Model for heating load.

The variables in the dataset has been obtained through simulations done on buildings having the same construction material and volume and varying other dimensions and surface areas. The output and input parameters present in the dataset have been obtained from these simulations and it has been seen that these physical properties large influence the energy performance of the buildings. Details of these simulations can been seen in the following research [35]. The dataset is split into training and test data with a probability of 0.7 and 0.3 respectively. The training sample is used to build the model and then the test sample is used to test the capabilities of the model built. Furthermore, for the ELM model, we tune the parameters differently to see which gives the best performance. We tested with number of hidden neurons ranging from 20 to 50 and different activation functions like linear, sigmoid, tan-sigmoid etc.

3.2. Development of the predictive models

The heating load is the amount of heat energy that would need to be added to a space to maintain the temperature in an acceptable range. The cooling load is the amount of heat energy that would need to be removed from a space (cooling) to maintain the temperature in an acceptable range. The heating and cooling loads, or "thermal loads", take into account factors like- the dwelling's construction and insulation; including floors, walls, ceilings and roof; and the dwelling's glazing and skylights; based on size, performance, shading and overshadowing. As mentioned in Table 2, we can see the distribution of the dependent and independent variable. This data has been normalised prior to building the models and used. In the literature, few applications of regression models exists [36–42].

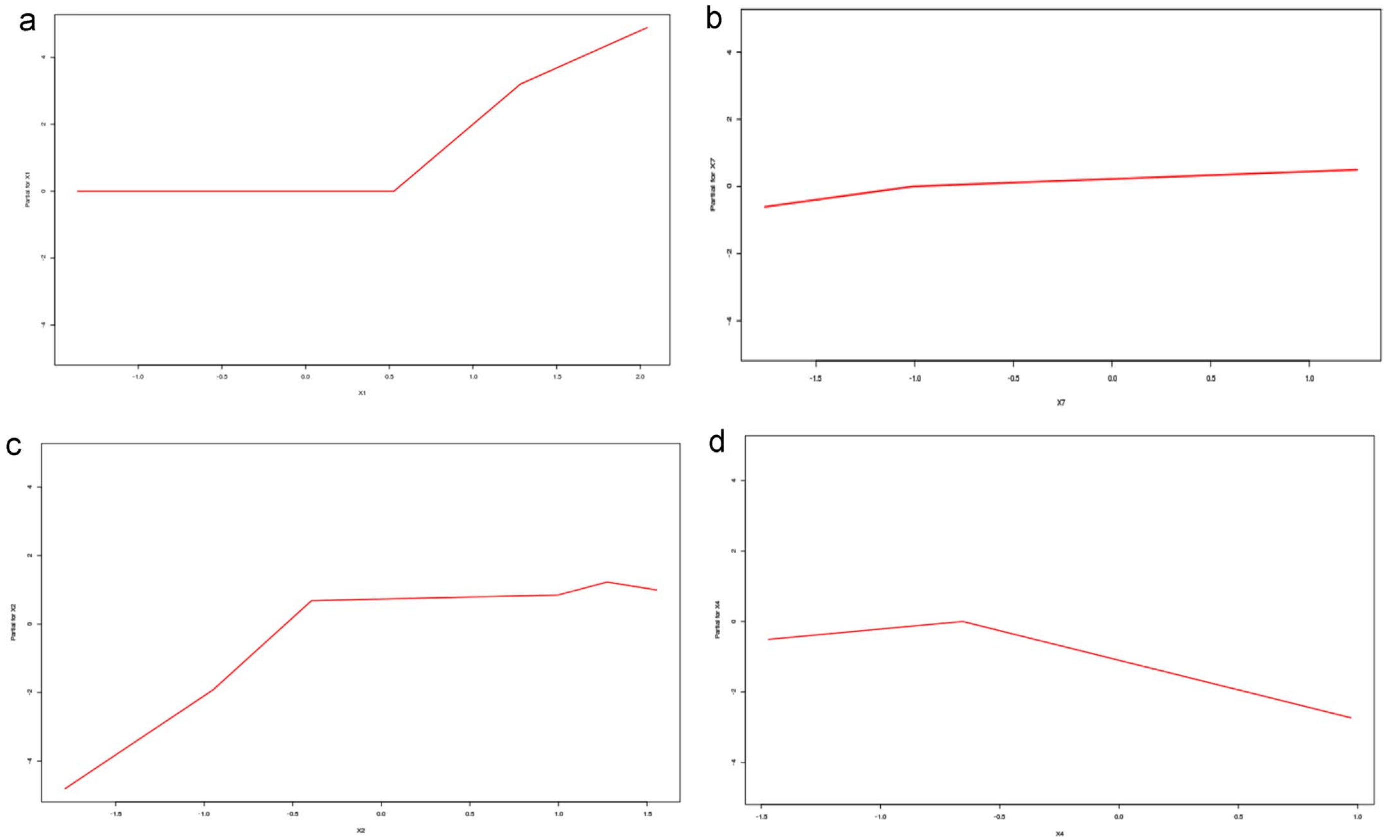

Fig. 8. a. Regression Line for X1. b. Regression Line for X2. c. Regression Line for X4. d. Regression Line for X7.

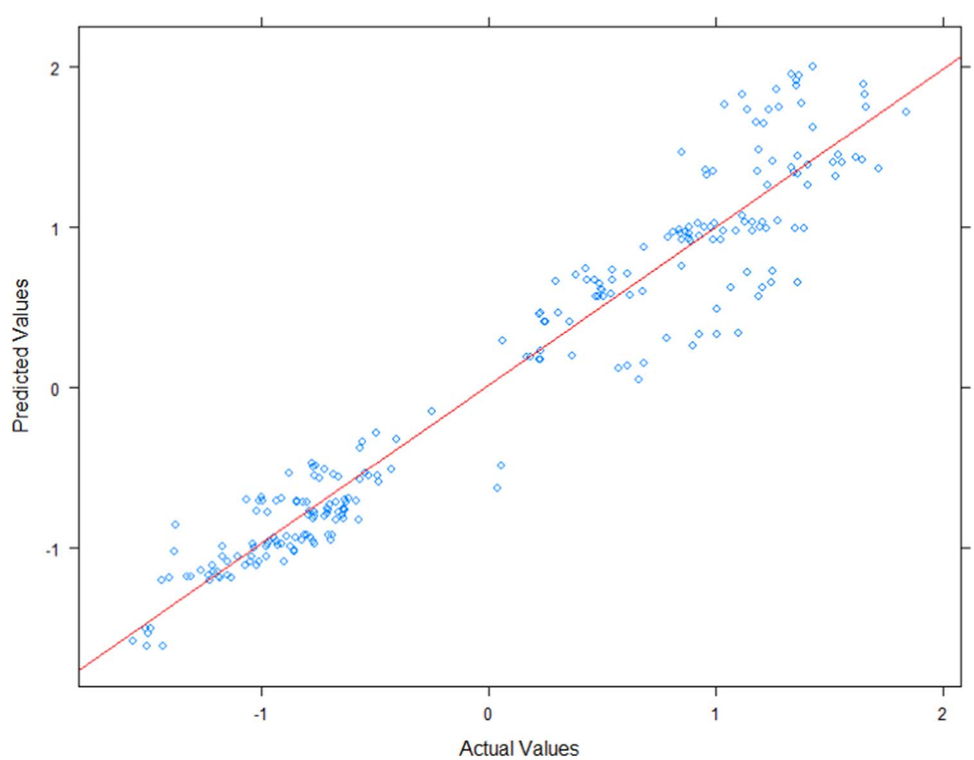

Fig. 9. Actual vs. Predicted output obtained by ELM model.

In order to demonstrate the modelling results of MARS, the first built MARS model will be used as an illustrative example. It is observed that parameters relative compactness (X1), surface area (X2), wall area (X3), roof area (X4) and glazing area (X7) do play important roles in deciding the MARS models. Besides, according to the obtained basis functions and the MARS prediction function, it can be observed that different parameters have different weights and the high and low value for the variables yield different heating load values. The above conclusions from the basis functions and MARS prediction function

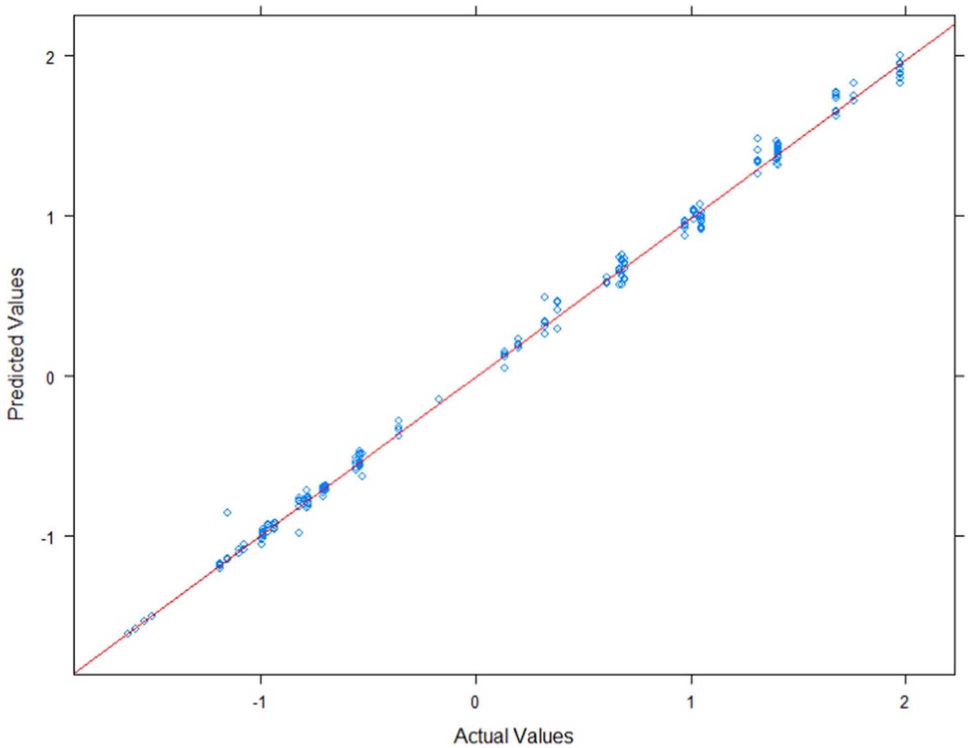

Fig. 10. Actual vs. Predicted output obtained by MARS and ELM.

Table 4
Variable Importance Evaluation by MARS.

Variable	nsubsets	gcv	rss
X4	11	100.0	100.0
X7	10	45.3	45.3
X2	9	37.2	37.1
X1	8	32.1	31.8
X3	unused		
X5	unused		
X6	unused		
X8	unused		

Tablee 5a
Error Measures for heating load.

Method	Measure							
	R^2	RMSE	MAE	C_c	ρ	MAPE	WMAPE	Time
MARS	0.9902	0.1002	0.0769	0.995	0.798	2.2043	0.09768	1.0351
ELM	0.9354	0.2573	0.1887	0.9673	2.079	17.739	1.060408	0.7651
Hybrid Model	0.9973	0.0525	0.0365	0.998	0.417	2.2835	0.21149	0.400907

Table 5b
Error Measures for cooling load.

Method	Measure							
	R^2	RMSE	MAE	C_c	ρ	MAPE	WMAPE	Time
MARS	0.9642	0.1950	0.1458	0.9826	1.2732	19.2319	1.0517	1.0149
ELM	0.9054	0.3170	0.2381	0.9524	2.1020	27.7105	0.8639	0.7622
Hybrid Model	0.9641	0.1952	0.1271	0.9824	1.2751	19.5096	1.1786	0.4462

have important managerial implications since it can help professionals design appropriate buildings that shall be energy efficient. The above-mentioned technical merits of MARS are one of the main reasons for using it in designing this two-stage hybrid model and these concerns are further verified.

From Table 3, it can be observed that the developed MARS has produced eleven basis functions starting from BF1 to BF11. The mathematical equation follows:

Table 6
Results received in conventional methods for heating load prediction.

	Gaussian Processes	Linear Regression	ANN	RBF Network
R^2	0.941951	0.92449	0.9883	0.8043
Correlation coefficient	0.9693	0.9599	0.9945	0.8838
Mean absolute error	0.1753	0.1959	0.0851	0.3535
Root mean squared error	0.2443	0.2788	0.1056	0.4585
Relative absolute error	19.1729%	21.4315%	9.3057%	40.2363%
Root relative squared error	24.4004%	27.851%	10.55%	46.7624%
Time(seconds)	0.98	0.02	0.83	0.04

$$\begin{aligned} &= 0.7422228 + 4.234267 * BF1 - 2.00807 * BF2 + 1.228276 * BF3 \\ &\quad - 3.45402 * BF4 - 1.109239 * BF5 + 1.2453 * BF6 - 2.190511 * BF7 \\ &\quad - 0.6199386 * BF8 - 1.678417 * BF9 - 0.8088267 * BF10 \\ &\quad + 0.2216917 * BF11 \end{aligned} \tag{18}$$

in which, BF1= Max (0, X1-0.527838), BF2 = Max (0, X1-1.28414), BF3 = Max (0, X2 to −0.950301), BF4 = Max (0, −0.394027-X2), BF5 = Max (0, X2 to −0.394027), BF6 = Max (0, X2-0.996657), BF7 = Max (0, X2-1.27479), BF8 = Max (0, −0.655453-X4), BF9 = Max (0, X4 to −0.655453), BF10 = Max (0, −1.00867-X7), BF11= Max (0, X7 to −1.00867).

In MARS, the definition of Max is as below,

If (p > q), then max(p,q)=p otherwise it is q. In MARS, the connection points of the polynomials are termed as knots. For example, in the above equations (Eq. (18)) few of the knots are available at x = 0.527838, x = 1.28414 and x = 0.996657, x = 1.27479 and it continues for other basis functions as well. The obtained actual vs predicted graph of MARS has been shown in Fig. 5.

The variable importance graph has been shown in Fig. 6. It has two main components; first one is nsubsets criterion, which counts the number of model subsets that comprise the variable, the second one is sqrt gcv which is the square root of generalized cross validation (GCV). From Fig. 6, it can be seen that for variable X4 nsubsets is 11, and other variables the nsubsets values decreased. The sqrt gcv values decreased more drastically than nsubsets, after X4 variable.

The first sub figure of Fig. 7 shows, GRSq value of the heating load predictions using MARS model. This error says that how better the proposed model predicts while using data not in the energy train data set. The second graph shows the cumulative distribution of the absolute values of residuals. A prediction is when the graph starts at zero and shoots very fast to 1, and which is the case here in our proposed work. The third sub figure shows the graph of fitted and residuals. It describes the residual for each value of the predicted heating response. The redline shows that residual diverges at low and high fitted heating load values. The fourth one is a graph of normal quantiles vs residual quantile of predicted heating load.

Next, we have shown the actual vs. predicted output obtained by ELM when applied on the test dataset. We have built this ELM model using all the dependent variables as the input node and splitting the dataset into training and test data with a probability of 0.7 and 0.3. From the results of MARS model tabulated in Table 3, we can see that only variable X1, X2, X4 and X7 are important in building the prediction model with different measures of their participation while predicting heating load. While the remaining variables X3, X5, X6 and X8 do not have any contribution in developing an accurate model to predict the target variable that is the heating load valuable. When we build the hybrid model, we shall eliminate these unnecessary variables from the dataset and only give the important variables as input to our ELM and to make the performance to the new model better than the initial ELM model. The same has been done to predict the cooling load as well (Figs. 8–10)).

Hence, we have tried with the hybrid model of MARS and ELM to check whether any possibility exists to obtain better performance in comparison to MARS and ELM for heating load prediction. Therefore, we have used MARS along with ELM to improve the model to predict energy efficiency. The input layer of the ELM now becomes the obtained important variables from MARS removing the unnecessary variables. The various error measures have been checked on the predicted value.

4. Performance assessment of the predictive models

To check the performances of MARS, ELM and the Hybrid model of MARS and ELM as well as the linear regression, Gaussian process, neural network and radial basis function network, we have carried on comparative study of all these methods. The importance of variables obtained by MARS are shown in Table 4 and it can be seen that X4 is the most important variable followed by X7. For the advanced models that we are focusing on heating and cooling load prediction respectively and Table 6 represents the conventional ones. To compare each method, we have calculated standard errors obtained by these proposed three methods. The definitions of the errors are given below,

R-squared is a statistical measure of how close the data is to the regression line, given by:

$$R2 = 1 - \frac{\sum_{i=1}^{N}(A_i - P_i)^2}{\sum_{i=1}^{N}(A_i - \overline{A})^2} \tag{19}$$

Root-mean-square error (RMSE) is a measure of the differences between values (actual and predicted values) predicted by a model or an estimator and the values actually observed. It can be defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - P_i)^2} \tag{20}$$

$$\mathrm{Cc} = \frac{\sum_{i=1}^{N}((A_i - \bar{A})(P_i - \bar{P}))}{\sqrt{\sum_{i=1}^{N}(A_i - \bar{A})^2 \sum_{i=1}^{N}(P_i - \bar{P})^2}} \tag{21}$$

$$\rho = \frac{RMSE}{\bar{A}} \times \frac{1}{\mathrm{Cc}+1} \tag{22}$$

The mean absolute percentage error (MAPE), is a measure of predictive accuracy of a forecasting method is statistics. It usually expresses accuracy as a percentage, as follows:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|A_i - P_i|}{A_i} \times 100 \tag{23}$$

Weighted mean absolute percent error (WMAPE) is a weighted modification of MAPE given by:

$$\mathrm{WMAPE} = \frac{\sum_{i=1}^{N}\left|\frac{A_i - P_i}{A_i}\right| \times A_i}{\sum_{i=1}^{N}A_i} \tag{24}$$

As can be seen from Table 5a, in case of heating load the hybrid model outperforms all others by most metrics. The RMSE being more sensitive to errors is a more conclusive metric here and clearly shows good performance from our models with 0.1002 for MARS and an improvement from 0.2573 to 0.0525. For cooling load prediction, the MARS and the hybrid- ELM perform almost similar and are both better than the original ELM (Table 5b). We can also see that by time of execution (for both heating and cooling load) the hybrid ELM has an edge over the remaining which will come in handy while dealing with large number of real-time data (Table 5a and 5b). Also, the comparative studies have been carried out with other very powerful machine learning methods like Gaussian Process Regression, Linear Regression, Artificial Neural Network and Radial Basis Function Network (RBF Network) and the results for heating load prediction have been documented. The comparisons are carried out for several important parameters like R^2, Correlation coefficient, Mean absolute error (MEA), RMSE, Relative absolute error (RAE), Root relative squared error (RRSE), and the execution time (in seconds) of each methods as can be seen from Table 6. The Gaussian Process Regression has used RBF kernel whose mathematical expression is K(x,y) = e^-(1.0* < x-y, x-y > 2), where x and y stands for actual and predicted values of cooling load. This model takes 0.98 s for execution. The R2 value for Gaussian process regression is 0.941, Correlation coefficient is 0.9693, Mean absolute error is 0.1753, Root mean squared error is 0.2443, Relative absolute error is 19.1729% and Root relative squared error is 24.4004%. Next model is Linear Regression Model which has been compared with others. The obtained R2 value is for linear regression is 0.92449, Correlation coefficient is 0.9599, Mean absolute error value is 0.1959, Root mean squared error is 0.2788, Relative absolute error is 21.4315%, and Root relative squared error 27.851% and execution time taken is 0.02 s. We also tried ANN (Artificial Neural Network) to find the predictive values for cooling load of building and it takes 0.83 s to construct the model. ANN has obtained R2 as 0.9883, correlation coefficient is 0.9945, Mean absolute error as 0.0851, Root mean squared error as 0.1056, Relative absolute error as 9.3057%, and Root relative squared error as 10.55%. Finally we compared with our model with RBF Network. The obtained outputs as R2 and Correlation coefficient are 0.8043 and 0.8838 respectively, Mean absolute error is 0.3535, Root mean squared error is 0.4585, Relative absolute error is 40.2363%,Root relative squared error is 46.7624% and it takes 0.04 s to execute. There has been other similar research done and so far, our results are very much comparable to those [43,44]. Research in this field is very pertinent at the moment with the rise in population and living standards and so a lot of effort is being invested in the field of energy conservation and energy efficiency as can be seen from the various techniques and implementations being done at the moment [45–51].

5. Conclusion

This article examines the potential competence of MARS, ELM and the hybrid model of MARS and ELM for heating load and cooling load forecasting in buildings. The formulated MARS has produced excellent performance but has taken more computation time and processing power. The performance of MARS and ELM have been compared with a hybrid model of MARS and ELM together. The RMSE values obtained by the hybrid optimised model have surpassed the MARS and ELM model. Moreover, the comparison to other traditional techniques also has proven that the more advanced models could be used for even better prediction. The MARS model is a non-parametric regression model that splits the data and fits each interval into a basis function and ELM is similar to a Single Layer Feed-forward Neural Network except that the randomly assigned input weights that are not updated. In the proposed hybrid model, the MARS evaluates the importance of every parameter in the prediction and the important parameters have been fed to the ELM to build another model of prediction for load forecasting of energy in buildings.

The limitation these models face at the moment is the lack of more real time data and the models being restricted to only a few parameters. Further research can be done by applying feature engineering techniques or by coupling these models with optimisation algorithms like genetic algorithm, natured-based optimisation and other evolutionary computing techniques that could help improvise the predictions even further taking lesser computation time and power. This kind of prediction is very relevant to energy and building engineers as it helps them foresee the energy requirements and design building more efficiently keeping in mind occupant comfort, lower maintenance cost and greater life of equipments. This will play a vital role in implementing artificial intelligence in energy conservation and thus a lot of research is currently being done in this direction using different techniques.

References

[1] Duncan RC. The Olduvai Theory: Energy, Population, and Industrial Civilization. The Social Contract, Winter 2005–2006.
[2] Burdick A. Strategy guideline: accurate heating and cooling load calculations. US Department of Energy, Energy Efficiency & Renewable Energy, Building Technologies Program. Jun 1, 2011.

[3] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy Build 2008;40(3):394–8.
[4] Shukuya M, Matsunawa K. A nomograph for estimating annual cooling and heating energy requirements in buildings dominated by internal loads. Energy Build 1988;12(3):207–18.
[5] Duncan RC. The peak of world oil production and the road to the Olduvai Gorge. InPardee Keynote Symposia Geological Society of America, Summit May 13, 2000.
[6] Maile Tobias, Fischer Martin, Bazjanac, Vladimir. Building Energy Performance Simulation Tools - a Life-Cycle and Interoperable Perspective, CIFE Working Paper, December 2007.
[7] International Energy Agency, Transition to Sustainable Buildings, ISBN: 978-92-64-20241-2.
[8] Amoco BP. Statistical review of world energy. BP Amoco 2001.
[9] Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991:1–67.
[10] Cortes C, Vapnik V. Machine learning. Support-vector networks, journal. 1995. 20:273-97.
[11] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293–300.
[12] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1):489–501.
[13] Rounaghi MM, Abbaszadeh MR, Arashi M. Stock price forecasting for companies listed on Tehran stock exchange using multivariate adaptive regression splines model and semi-parametric splines technique. Phys A: Stat Mech Appl 2015;438:625–33.
[14] Zhang W, Goh AT. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci Front 2016;7(1):45–52.
[15] Borodin V, Bourtembourg J, Hnaien F, Labadie N. Predictive modelling with panel data and multivariate adaptive regression splines: case of farmers crop delivery for a harvest season ahead. Stoch Environ Res Risk Assess 2016;30(1):309–25.
[16] Islam T, Srivastava PK, Dai Q, Gupta M, Zhuo L. Rain rate retrieval algorithm for conical-scanning microwave imagers aided by random forest, RReliefF, and multivariate adaptive regression splines (RAMARS). IEEE Sens J 2015;15(4):2186–93.
[17] Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern Part B: Cybern 2012;42(2):513–29.
[18] Kariminia S, Shamshirband S, Motamedi S, Hashim R, Roy C. A systematic extreme learning machine approach to analyze visitors′ thermal comfort at a public urban space. Renew Sustain Energy Rev 2016;58:751–60.
[19] Liu X, Lin S, Fang J, Xu Z. Is extreme learning machine feasible? A theoretical assessment (Part I). IEEE Trans Neural Netw Learn Syst 2015;26(1):7–20.
[20] Wong PK, Wong KI, Vong CM, Cheung CS. Modeling and optimization of biodiesel engine performance using kernel-based extreme learning machine and cuckoo search. Renew Energy 2015;74:640–7.
[21] Nianyin Zeng Hong, Zhang Weibo, Liu Jinling, Liang , Alsaadi Faud E. A switching delayed PSO optimized extreme learning machine for short-term load forecasting. Neurocomputhing 2017, (Elsevier).
[22] Mao Wentao, Wang Jinwan, He Ling, Tian Yangyang. Online sequential prediction of imbalance data with two-stage hybrid strategy by extreme learning machine. Neurocomputing 2016, (Elsevier).
[23] Biredagn Nahato Kindie, Nehemiah Khanna H, Kannan A. Hybrid approach using fuzzy sets and extreme learning machinefor classifying clinical datasets. Inform Med Unlocked 2016, (Elsevier).
[24] Sánchez-Oro J, Duarte A, Salcedo-Sanz S. Robust total energy demand estimation with a hybrid Variable Neighborhood Search – Extreme Learning Machine algorithm. Energy Convers Manag 2016, (Elsevier).
[25] Ou Tsung-Yin, Cheng Chen-Yang, Chen Po-Jung, Perng Chayun. Dynamic cost forecasting model based on extreme learning machine – A case study in steel plant. Comput Ind Eng 2016, (Elsevier).
[26] Guo Peng, Cheng Wenming, Wang Yi. Hybrid evolutionary algorithm with extreme machine learning fitness function evaluation for two-stage capacitated facility location problems. Expert Syst Appl 2016, (Elsevier).
[27] Al-Yaseen Wathiq Laftah, Othman Zulaiha Ali, Nazri Mohd Zakree Ahmad. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst Appl 2016, (Elsevier).
[28] Williams CK, Rasmussen CE. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2006.
[29] Deo RC, Samui P. Forecasting evaporative loss by least-square support-vector regression and evaluation with genetic programming, Gaussian process, and minimax probability machine regression: case study of Brisbane City. J Hydrol Eng 2017:05017003.
[30] Kutner MH, Nachtsheim C, Neter J. Applied linear regression models. McGraw-Hill/Irwin; 2004.
[31] Ascione Fabrizio, Bianco Nicola, Stasio Claudio De, Mauro Gerardo Maria, Vanoli Giuseppe Peter. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: a novel approach. Energy 2016, (Elsevier).
[32] Chen S, Cowan CF, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans Neural Netw 1991;2(2):302–9.
[33] Schwenker F, Kestler HA, Palm G. Three learning phases for radial-basis-function networks. Neural Netw 2001;14(4):439–58.
[34] Billings SA, Zheng GL. Radial basis function network configuration using genetic algorithms. Neural Netw 1995;8(6):877–90.
[35] Tsanas A, Xifara A. Accurate quantitative estimation of energy performanceof residential buildings using statistical machine learning tools. Energy and Build 2012;49(0):560–7.
[36] Shin M, Do SL. Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate. Energy Build 2016;110:57–70.
[37] Naji S, Keivani A, Shamshirband S, Alengaram UJ, Jumaat MZ, Mansor Z, Lee M. Estimating building energy consumption using extreme learning machine method. Energy 2016;97:506–16.
[38] Naji S, Shamshirband S, Basser H, Keivani A, Alengaram UJ, Jumaat MZ, Petković D. Application of adaptive neuro-fuzzy methodology for estimating building energy consumption. Renew Sustain Energy Rev 2016;53:1520–8.
[39] Harish VS, Kumar A. A review on modeling and simulation of building energy systems. Renew Sustain Energy Rev 2016;56:1272–92.
[40] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: lessons from assembling a very large national database of building energy use. Appl Energy 2015;140:85–93.
[41] Yuan X, Wang X, Zuo J. Renewable energy in buildings in China—a review. Renew Sustain Energy Rev 2013;24:1–8.
[42] Cabeza LF, Castell A, Barreneche C, De Gracia A, Fernández AI. Materials used as PCM in thermal energy storage in buildings: a review. Renew Sustain Energy Rev 2011;15(3):1675–95.
[43] Chou Jui-Sheng, Bui Dac-Khuong. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build 2014, (Elsevier).
[44] Moghaddam TB, Soltani M, Karim MR, Shamshirband S, Petković D, Baaj H. Estimation of the rutting performance of Polyethylene Terephthalate modified asphalt mixtures by adaptive neuro-fuzzy methodology. Constr Build Mater 2015;96:550–5.
[45] Stadler P, Ashouri A, Maréchal F. Model-based optimization of distributed and renewable energy systems in buildings. Energy Build 2016;120:103–13.
[46] Tyagi VV, Pandey AK, Buddhi D, Kothari R. Thermal performance assessment of encapsulated PCM based thermal management system to reduce peak energy demand in buildings. Energy Build 2016;117:44–52.
[47] Molina-Solana Miguel, Rosa María, Ruiz M Dolores, Gómez-Romero Juan, Martin-Bautista MJ. Data science for building energy management: a review. Renew Sustain Energy Rev 2017, (Elsevier).
[48] Deb Chirag, Zhang Fan, Yang Junjing, Lee Siew Eang, Shah Kwok Wei. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017, (Elsevier).
[49] Cheng Min-Yuan, Cao Minh-Tu. Accurately predicting building energy performance usingevolutionary multivariate adaptive regression splines. Appl Soft Comput 2017, (Elsevier).
[50] Meloa AP, Versage RS, Sawaya G, Lamberts R. A novel surrogate model to support building energy labelling system: a new approach to assess cooling energy demand in commercial buildings. Energy Build 2016.
[51] Zeyu Wang, Ravi S. Srinivasan. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models, Renewable and Sustainable Energy Reviews. (Elsevier).

Sanjiban Sekhar Roy*, Reetika Roy, Valentina E. Balas
School of Computer Science and Engineering, VIT University, Vellore 632014, India
Automation and Applied Informatics, Aurel Vlaicu University of Arad, Romania
E-mail address: sanjibanroy09@gmail.com

* Correspondence to: School of Computer Science and Engineering VIT University, Room No: 116-A29, SJT, Vellore 632014, India.