Applied Soft Computing 22 (2014) 178–188

Contents lists available at ScienceDirect

Applied Soft Computing

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

# Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines

Min-Yuan Cheng, Minh-Tu Cao*

*Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan*

ARTICLE INFO

*Article history:*
Received 24 July 2013
Received in revised form 9 May 2014
Accepted 16 May 2014
Available online 27 May 2014

*Keywords:*
Multivariate adaptive regression splines
Artificial intelligence
Artificial bee colony
Energy performance of buildings
Heating load
Cooling load

ABSTRACT

This paper proposes using evolutionary multivariate adaptive regression splines (EMARS), an artificial intelligence (AI) model, to efficiently predict the energy performance of buildings (EPB). EMARS is a hybrid of multivariate adaptive regression splines (MARS) and artificial bee colony (ABC). In EMARS, MARS addresses learning and curve fitting and ABC carries out optimization to determine the fittest parameter settings with minimal prediction error. The proposed model was constructed using 768 experimental datasets from the literature, with eight input parameters and two output parameters (cooling load (CL) and heating load (HL)). EMARS performance was compared against five other AI models, including MARS, back-propagation neural network (BPNN), radial basis function neural network (RBFNN), classification and regression tree (CART), and support vector machine (SVM). A 10-fold cross-validation approach found EMARS to be the best model for predicting CL and HL with 65% and 45% deduction in terms of *RMSE*, respectively, compared to other methods. Furthermore, EMARS is able to operate autonomously without human intervention or domain knowledge; represent derived relationship between response (HL and CL) with predictor variables associated with their relative importance.



## 1. Introduction

The high energy consumption and long-term, adverse impact on the environment of the building sector [1] has encouraged growing research interest in the energy performance of buildings (EPB). Building energy consumption has increased rapidly over past decades, in line with rising living standards. In the European Community, for example, the building sector accounts for 40% of total energy use and 36% of total $CO_2$ emissions [2]. Designing more energy-efficient, enhanced energy performance buildings is thus important to mitigating further growth in energy demand and $CO_2$ emissions.

Accurate cooling load (CL) and heating load (HL) estimations and correct identification of parameters that significantly affect building energy demand are necessary to set appropriate equipment specifications, install systems properly, and optimize building designs. However, interrelationships between parameters such as relative compactness [3], climate [4], surface area, wall area, roof area [5], and orientation [6] complicate their respective relationships with EPB and make accurate HL and CL estimation a challenging task for architects.

Although building simulation tools are currently used widely to predict and analyze building energy consumption, these tools have several important drawbacks: (1) they are suboptimal for predicting energy use in occupied buildings; (2) different simulation tools can generate significantly different prediction values [7]; (3) their multidisciplinary nature makes these tools very complicated and time-consuming to use; and (4) identifying and comparing the impact of variables on observed quantities of interest is very difficult [8].

Artificial intelligence (AI) inference models are increasingly viewed as a viable alternative approach to predicting EPB. AI used in models that simulate the human inference processes can infer new facts from previously acquired information and change adaptively in response to changes in the historical data. Tsanas [8] stated that AI is not only extremely fast at obtaining answers, but also assists architects to analyze the relative impact of significant parameters of interest.

Many studies have explored the ability of various AI models to predict various quantities of interest in the context of EPB. Most have focused on models based on slightly modified artificial neural network (ANN) [9–12] and support vector machine (SVM) [13–16] integrated with traditional techniques such as classification and

* Corresponding author. Tel.: +886 921634420.
*E-mail addresses:* caominhtu2201@gmail.com, luis_cao2201@yahoo.com (M.-T. Cao).

http://dx.doi.org/10.1016/j.asoc.2014.05.015
1568-4946/

regression tree (CART) [7]. This relatively narrow application of AI to EPB leads the authors to question whether these models adequately address the problem of EPB prediction.

ANNs are the most widely used AI models in the application of building energy prediction. A major disadvantage of ANN is the large number of controlling parameters required to construct the network [17], including the number of hidden layers and neurons in hidden layers, learning rate, and momentum. Moreover, the ANN training process must be obtained via a gradient descent algorithm on the error space, which can be very complex and may contain many local solutions that prevent an ANN model from converging on an optimal solution [18]. It is relatively difficult to attain maximum success using ANN for predicting EPB.

SVM using radial basis function kernel (RBF-SVM) is perceived as highly effective model in a line of improving predictive accuracy of EPB and has been demonstrated in quantitative studies [19]. It is noticing that using SVM is expert-dependent, thus requires expertise to select appropriate penalty parameter $C$ and kernel function parameter $\gamma$ values. Additionally, similar to ANN, the results provided by SVM are not easy to be interpreted because it does not provide explanatory insight into derived relationship between response and input variables. Thus, the two models are merely more appropriate for the prediction problem. As for CART, this method does not generate stable predictions, has difficulty modeling additive structures [20].

Multivariate adaptive regression splines (MARS) [21] is an AI technique of potential help to architects. In addition to its ability to handle prediction problems, MARS is able to determine the input parameters that significantly impact output parameters and even explore the complex nonlinear relationships between a response variable and various predictor variables, which is essential in analyzing and designing energy-efficient buildings. These advantages underline MARS' successful deployment in various problem areas such as credit scoring [22], computer wholesaling [23], paper manufacturing [24], public water supply [25], transportation [26], geological engineering [27] and engineering software [28]. Despite its success elsewhere, MARS has seen surprisingly little application in building-related studies to date.

Building a MARS model requires that users select three tuning parameters, including maximum number of basis functions ($M_{\max}$), penalty (i.e., smooth parameter) ($d$), and maximum interaction between variables ($I_{\max}$). These parameters were cited by Andalib and Atry [29] as important features in MARS model construction due to their control of model complexity and generalization [29]. The quality of the selected MARS parameters thus significantly affects MARS prediction accuracy. Friedman's prior parameter selection suggestions all have large value ranges, with actual selected values dependent on the dataset at hand [21]. Optimal values may remain outside the suggested ranges.

In machine learning, identifying optimal parameter values is a challenging task and considered as an optimization problem. This paper thus uses the ABC [30] algorithm as a search engine to determine optimal MARS parameter values. ABC was introduced by Karaboga in 2005 and is a swarm intelligence-based optimization algorithm inspired by honeybee foraging behavior. Its relatively small number of control parameters makes ABC flexible and easy to execute for novice users [31]. Researchers have demonstrated that ABC is superior to other algorithms in identifying optimal solutions [32,33]. ABC is also a reliable tool when paired with other data mining techniques [34]. ABC is thus a potentially useful search engine for identifying useful MARS parameters such as $M_{\max}$, $I_{\max}$ and $d$.

The objective of this research was to develop and test the evolutionary multivariate adaptive regression splines (EMARS). The authors created EMARS by fusing MARS and ABC to incorporate the strengths and eliminate the weaknesses of each technique. This newly proposed model operates automatically without human intervention and accurately estimates building CL and HL values under various parameter settings. This study then compared EMARS performance against five other AI techniques, including MARS, BPNN, RBFNN, CART, and SVM.

The remainder of this paper is organized as follows: the second section reviews related research works; the third introduces the EMARS model; the fourth describes the data collection process; the fifth validates and analyzes EMARS performance and compares simulation results; and the last presents conclusions.

## 2. Literature review

### 2.1. Energy performance of buildings

The energy performance of both new and existing buildings, as determined by their volumetric and plan solutions, may be assessed using rating system standards such as the U.S.'s Leadership in Energy and Environmental Design (LEED), Germany's Sustainable Building Council (DGNB), and the U.K.'s Building Research Establishment Environmental Assessment Methodology (BREEAM). The European Union adopted the Energy Performance of Buildings Directive (EPBD) on 16 December 2002. This directive set four key energy efficiency requirements for buildings that were intended to ensure the achievement of the EU's 2020 20% headline target on energy efficiency and to pave the way for further energy efficiency improvements.

Energy used in a building may be classified into the three categories of heating load/cooling load (HL/CL), lighting, and hot water provision. HL/CL, which typically accounts for the largest portion of building energy usage, is a measure of energy required to add and remove heat in a space through heating, ventilation, and air conditioning (HVAC) systems in order to maintain desired ambient temperatures indoors. Achieving high energy performance in buildings thus seeks to reduce the CL/HL significantly. Installing a properly sized HVAC system (right-sizing) is critical to achieving this goal.

Right-sizing requires an accurate understanding of the heating and cooling loads of a building, because calculated CL and HL values determine the HVAC equipment specifications and the air distribution system design necessary to maximize heating and cooling efficiencies. The calculated values for CL and HL thus impact initial construction costs in the short term and the operating energy efficiency, occupant comfort, indoor air quality, and durability of the building in the medium and long terms.

### 2.2. Multivariate adaptive regression splines

MARS was first proposed by Friedman (1991) [21] as a flexible procedure to organize relationships that are nearly additive or involve interactions with fewer variables. MARS makes no assumptions about the underlying functional relationships between dependent and independent variables in order to estimate the general functions of high-dimensional arguments given sparse data [21,35]. One further advantage of MARS is its ability to estimate the contributions of basis functions so that the additive and interactive effects of predictors are allowed to determine the response variable.

MARS is established by fitting a basis function (term) to distinct independent variable intervals. In general, splines (also called piecewise polynomials) have pieces that connect smoothly together. The interface points between pieces are called knots, denoted as $t$. MARS uses two-sided truncated power functions as