Energy and Buildings 107 (2015) 445–455

Contents lists available at ScienceDirect

Energy and Buildings

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

# An improved office building cooling load prediction model based on multivariable linear regression

Qiang Guo, Zhe Tian*, Yan Ding, Neng Zhu

*School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China*

ARTICLE INFO

*Article history:*
Received 21 May 2015
Accepted 21 August 2015
Available online 24 August 2015

*Keywords:*
Cooling load prediction
Multivariable linear regression
Principal component analysis
Cumulative effect of high temperature
Dynamic two-step correction

ABSTRACT

The cooling load prediction of heating, ventilating and air-conditioning (HVAC) systems in office buildings is fundamental work for optimizing the operation of HVAC systems. In this paper, an improved multivariable linear regression model is proposed to predict the daily mean cooling load of office buildings in which three main measures, including the principal component analysis (PCA) of meteorological factors, cumulative effect of high temperature (CEHT) and dynamic two-step correction, are used to improve prediction accuracy. The site measured cooling load of two office buildings in Tianjin is used to validate the model and evaluate the prediction accuracy. Meanwhile, four contrast models with one or two of the three measures are also built. A comparison among the models proves that a combination of the three measures could effectively improve the prediction accuracy. The predicted load of the proposed model has acceptable agreement with actual load, where the mean absolute relative error is less than 8%.



## 1. Introduction

Public buildings consume considerably more energy than residential buildings [1,2]. As a major portion, energy consumption in office buildings is larger than any other building because of their tremendous amount of floor area [3]. Currently because heating, ventilating and air-conditioning (HVAC) systems are generally integrant in office buildings, making HVAC systems more efficient is an effective way to lower the total energy consumption [4]. In addition to directing the design of HVAC systems, cooling load predictions can guide the operation and optimization of building energy systems to improve efficiency, especially for systems with thermal storage devices [5].

Cooling load prediction models can be built by many methods. At first, traditional methods are applied to calculate and predict cooling load, such as the admittance method and Fourier method [6], transfer function method [7], semi-analytical method [8], and correlation method [9,10]. Along with the development of computer technologies and prediction methods, more methods are applied in cooling (thermal) load predictions, such as the artificial intelligence analysis [11,12], Monte Carlo simulation method [13], simulation analysis [14,15], regression analysis [16] and so on. The most popular methods are simulation analysis, artificial intelligence analysis and regression analysis [17].

Load simulation software, which is developed to simulate heat and mass transfer among indoor environments, outdoor environments and building envelops, is capable of predicting building cooling load as long as boundary conditions are input. To obtain accurate simulation results, detailed building information, including construction materials, structure, windows, shading types, occupancy and so on, is required before running the software. Meanwhile, meteorological data are needed, and its completeness and accuracy affect the simulation results directly. Some simulation software, such as TRNSYS [14], ESP-r [11] and DOE-2 [15], are applied in the cooling (thermal) load prediction of HVAC systems. Simulation analysis is usually used as a contrast method to validate the prediction accuracy of other methods. Obviously, the prediction accuracy of simulation software depends on the deviation of building information from reality. In practice, it is hard to build up a building model with identical information to reality in the software because not all of the input variables that relate to the cooling load of HVAC systems can be determined and measured accurately [18]. At the same time, it costs considerable manpower and time to build and operate the simulation models. Therefore, simulation analysis is not widely used in practice to predict the cooling load of HVAC systems in public buildings [19].

The most popular artificial intelligence methodology applied in the cooling load prediction of HVAC systems is the artificial neural network (ANN). The structure of an ANN model is a network consisting of many processing units arranged in layers [20], which is similar to the neurons in a human brain. A processing unit (neuron) is used to make weighted nonlinear calculations and

* Corresponding author.
*E-mail address:* tianzhe@tju.edu.cn (T. Zhe).

http://dx.doi.org/10.1016/j.enbuild.2015.08.041
0378-7788/

**Nomenclature**

| | |
|---|---|
| $HT_i$ | calculated value of cumulative effect of high temperature (kW) |
| $t$ | temperature (°C) |
| $t_0$ | boundary value of high temperature (30 °C in this paper) (°C) |
| $k_f$ | cumulative coefficient (kW/°C) |
| $l_i$ | actual cooling load in historical data (kW) |
| $\widehat{l_{i \cdot p}}$ | predicted cooling load in the $p$th period (kW) |
| $\Delta l_x$ | increased cooling load caused by the change of internal factors (kW) |
| $\Delta \widehat{l_x}$ | estimated value of $\Delta l_x$ (kW) |
| $l'_i$ | actual cooling load in prediction periods (kW) |
| $l_{r \cdot i}$ | replacement of $l'_i$ after the first correction (kW) |
| $\widehat{l_{r \cdot i}}$ | estimated value of $l_{r \cdot i}$ (kW) |
| $b$ | constant term (kW) |
| $F_i$ | principal component of meteorological parameters (°C) |
| $a$, $h$ | regression coefficient |
| $N'$ | number of the prediction results that meet the accuracy requirement |
| $N$ | total number of prediction results |
| $MO_p$ | prediction model in the $p$th prediction period |
| $MO_0$ | prediction model built based on a dataset completely composed of historical data |
| $\widehat{l_{i \cdot 0}}$ | prediction result based on $MO_0$ |

*Subscripts*

| | |
|---|---|
| $p$ | counter of prediction periods |
| $i$ | counter of weekdays |
| $q$, $f$ | counter of coefficients |
| $w$ | wet bulb |
| $d$ | dry bulb |
| $max$ | maximum value |

transform the activations to others. Many ANN methodologies have been well applied in building load predictions and have become the research focus during the past decade, such as the general regression neural network [11], feedback artificial neural network [21], and adaptive artificial neural network [22]. Compared with traditional methods, ANN has a better ability to solve complex problems with a large quantity of influencing parameters [11,21,22]. A major disadvantage of ANN is that a large number of controlling parameters are required to construct the network, including the number of hidden layers and neurons in hidden layers, the learning rate, and the momentum [23]. It may be a very complex process when that the ANN training process needs to be obtained via a gradient descent algorithm on the error space, which may contain many local solutions that prevent an ANN model from converging on an optimal solution [24]. At the same time, the training process needs a large amount of time. Because of the complexity of building reasonable models, ANN methods are used for theoretical research in most cases and are not widely implemented in practice.

Regression analysis [16] establishes the causal relationship in a functional form among explanatory variables and cooling load. Complete historical data are the basis of regression analysis. The key to regression analysis is to determine all of the possible influential variables and weights. Multivariable linear regression (MLR) models, auto regression (AR) models, and autoregressive exogenous (ARX) models have usually been used for cooling load predictions. The key work of MLR models is establishing a linear relationship between cooling load and input variables. When AR models are applied in prediction, the relationship between the predicted and historical load should be determined. The ARX models are used to predict the cooling load based on the exogenous inputs and historical load. Compared with other methods, it is simpler and more effective to build a prediction model based on regression analysis, which is widely applied. Although regression analysis is deemed to be less accurate in some studies [25], the prediction accuracy can be effectively improved if input variables are selected properly and different models can be combined reasonably. In some cases, the prediction accuracy of the regression method can be better than the artificial intelligence analysis [19]. Lam et al. [26] determined 12 input variables by sensitive analysis to build an MLR model to predict the energy use in air-conditioned office buildings in different climates. The maximum difference between the MLR model and the simulation model built by DOE-2 was within 10%. Yun et al. [16] built an ARX model with time and temperature index for 1 h ahead building thermal load prediction. This proposed model determined the dominant factors that affect the thermal load at a given time. A simulation model was used to determine the prediction accuracy of the proposed model on several different benchmark building types. Guo et al. [27] presented a model for the hourly cooling load prediction based on the time-indexed ARX method. In this model, two separate time-indexed ARX models were combined to improve the prediction accuracy. It was proven robust to outliers and was suitable for fast and adaptive coefficient estimation because the coefficients were estimated through a two-stage weighted least squares regression. Li et al. [19] re-evaluated four popular prediction models: the autoregressive moving average with exogenous (ARMAX) model, MLR model, ANN model and resistor–capacitor network (a simplified physical model). The results showed that the MLR model and the ARMAX model are superior in prediction accuracy and precision.

**Table 1**
The comparison among three prediction methodologies.

| Prediction methodologies | Advantages | Disadvantages |
|---|---|---|
| Simulation analysis | Mature theories<br>Easy comprehension | Time consuming for building models<br>Requiring considerable Manpower for operation<br>Impossibly determining all of the influencing parameters and setting correctly |
| Artificial intelligence analysis | Advanced theories<br>High accuracy<br>Better ability to solve complex problems with a large quantity of influencing parameters | Theories are hard to comprehend<br>Models are complicated to build<br>Time consuming for training process |
| Regression analysis | Mature theories<br>Simple models<br>Widely used<br>Computationally efficient<br>Prediction accuracy can be obviously improved by input selection and method combination | Relatively low accuracy (in some studies) |

The comparison among three prediction methodologies is shown in Table 1.

Throughout the existing studies on the cooling load predictions of HVAC systems, prediction accuracy is usually regarded as a key index while practicability is often ignored. Compared with the artificial intelligence analysis and simulation analysis, the easy operation of the regression analysis is one of the most important advantages in practical projects. Prediction models based on regression analysis can be widely implemented as long as the accuracy can be effectively improved.