Energy Conversion and Management 80 (2014) 46–53

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted least squares regression

Yin Guo [a], Ehsan Nazarian [b], Jeonghan Ko [c,d,*], Kamlakar Rajurkar [e]

[a] Facilities Management and Planning, University of Nebraska-Lincoln, NE 68588, USA
[b] Department of Industrial and Management Systems Engineering, University of Nebraska-Lincoln, NE 68588, USA
[c] Department of Industrial Engineering, Ajou University, Suwon, Gyeonggi-do 443-749, Republic of Korea
[d] Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, USA
[e] Department of Mechanical and Materials Engineering, University of Nebraska-Lincoln, NE 68588, USA

ARTICLE INFO

Article history:
Received 1 October 2013
Accepted 28 December 2013

Keywords:
Cooling load
Forecasting
ARX
Weighted least squares

ABSTRACT

This paper presents a robust hourly cooling-load forecasting method based on time-indexed autoregressive with exogenous inputs (ARX) models, in which the coefficients are estimated through a two-stage weighted least squares regression. The prediction method includes a combination of two separate time-indexed ARX models to improve prediction accuracy of the cooling load over different forecasting periods. The two-stage weighted least-squares regression approach in this study is robust to outliers and suitable for fast and adaptive coefficient estimation. The proposed method is tested on a large-scale central cooling system in an academic institution. The numerical case studies show the proposed prediction method performs better than some ANN and ARX forecasting models for the given test data set.

1. Introduction

Accurate cooling load prediction is essential for the optimal scheduling and planning of cooling system operation. As energy prices increase, optimal control of cooling systems has become more critical for energy cost reduction. Some studies reported that the accuracy in cooling load forecasting and optimal control of air-conditioning systems resulted in significant saving in buildings energy cost [1]. Thermal energy storage (TES) or hybrid cooling systems can be utilized to offset the cooling load during peak periods and lower energy costs [2]. However, without accurate cooling load prediction, it is almost impossible to take full advantage of such systems and their optimization; the efficient utilization of operation schemes in cooling systems necessitates reliable prediction of the cooling load [3].

Cooling load can be predicted based on external and internal factors of cooling systems. In particular, cooling load depends on many external factors of a cooling system, such as outdoor weather parameters, indoor equipment usage, and indoor human activities [4,5]. These external factors are often considered as the main driving forces of cooling load [6–8]. Cooling load also depends on internal factors such as historical load values or current load status. For instance, cooling systems are usually unable to respond immediately to the changes in the external factors driving the required load. This is especially common for large-scale central cooling systems. Thus, considering the external and internal factors, a system can be represented as a feedback system, as shown in Fig. 1.

Although numerous studies have proposed excellent methods for cooling load forecasting, the effect of outliers on cooling load prediction has not been addressed sufficiently in the literature. Moreover, some of the existing tools on cooling load forecasting appear to require substantial computation or are hard to be conveniently customized for a large building complex. Therefore, it is desirable to develop a more robust and adaptive method to mitigate the outlier effect and reduce computation.

This paper presents an hourly cooling-load prediction method based on time-indexed autoregressive with exogenous inputs (ARX) models with a two-stage weighted least squares regression. The prediction method uses a combination of two separate time-indexed ARX models for cooling load prediction: (1) 1 to 6 hours (1-6 h) ahead and (2) 7–24 h ahead. The first model results in improved cooling load forecasting for very short-term forecasting. The second model is more accurate for prediction periods greater than 7 h. The two-stage weighted least squares regression approach in this study allows fast and adaptive coefficient estimation and provides the capability of reducing the influence of outliers. Such a capability of outlier detection could also provide a potential extension for risk management. The proposed approach is applicable to meet a variety of accurate short-term cooling load and electrical demand forecasting needs.

* Corresponding author at: Department of Industrial Engineering, Ajou University, Suwon-si, Gyeonggi-do 443-749, Republic of Korea. Tel.: +82 31 219 2421.
E-mail addresses: jko@ajou.ac.kr, jeonghan@umich.edu (J. Ko).

0196-8904/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.enconman.2013.12.060

2. Literature review

A variety of prediction methods have been studied in cooling load forecasting. Such methods include simulation, grey-box, regression, support vector machines and artificial neural networks. Simulation-based methods usually rely on commercial energy simulation programs, e.g. EnergyPlus, TRNSYS, DOE-2, and ESP-r, to simulate the thermal response of buildings [9–12]. Although these programs are straightforward to use, they often require a large number of inputs and substantial computational effort. Furthermore, measuring for some of the required inputs is not an easy task or even possible. Thus, it could be difficult and costly to implement such simulation programs.

Grey-box modeling builds the cooling load prediction model partially based on expert knowledge by creating thermal networks [13,14]. However, this may be computationally expensive and very difficult to customize for a large-scale building complex for several reasons: (1) the structure of a thermal network has a significant effect on the prediction accuracy, (2) modern building complexes are designed and built with a high degree of complexity, and (3) different areas of a building complex may use different materials and have different indoor temperature set points and outdoor conditions.

Artificial neural networks (ANN) and support vector machines (SVM) are also common models for cooling load prediction [15–24]. These two models have been well studied and discussed in recent decades. They are known to be good at modeling nonlinear problems. In addition, some studies have improved the accuracy of these models. For example, Yao et al. [24] integrated residual error correction with radial based function (RBF) neural networks to increase the accuracy of the model. However, they usually suffer from high computational complexity and local optimums. There is also a risk of overtraining an ANN.

Regression and time-series models are also widely used for cooling load forecasting. Common models reported in the literature include multiple linear regression (MLR), autoregressive integrated moving average (ARIMA), and autoregressive moving average with exogenous inputs (ARMAX) models. An MLR model predicts the cooling load by establishing the correlation between cooling load and external inputs, which are mainly weather conditions. An ARIMA model predicts the cooling load by using only its previous records. The ARMAX model integrates the MLR and ARIMA models. Compared with the models mentioned in the previous paragraph, regression models usually require a much shorter training time, but are less capable of modeling nonlinearity.

It has been reported that some MLR and ARIMA models were not significantly more accurate than other models. For example, Moghram and Rahman [25] compared five models for short-term load forecasting (MLR, ARIMA, general exponential smoothing, state space method, and knowledge-based approach), and showed that traditional MLR and ARIMA models were no better than the other models. Kawashima et al. [26] showed that MLR and ARIMA models were less accurate than ANN models. Li et al. [19] compared ARIMA with ANN and developed an ARIMA-ANN model. Their result also showed that ARIMA was not as good as ANN. Yao et al. [22] applied Analytic Hierarchy Process (AHP) method to combine different point estimations of cooling load and determined their optimal weights. The proposed approach could result in significant improvement in cooling load forecasting by pair-wise judgments between models with periodically updated weights.

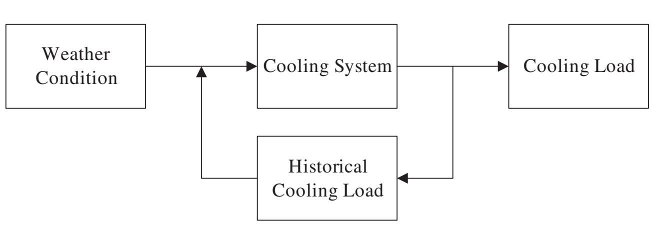

Fig. 1. Schematic diagram of a cooling system.

Several ARMAX models have been studied and showed better prediction. Yang et al. [27] studied an ARMAX model for short-term load forecasting, and compared the results obtained from an evolutionary programming (EP) approach with those from a gradient search based approach implemented in a commercial software package, SAS. The EP approach results were better, because this EP approach was capable of providing a global optimal estimation of the coefficients. Xu et al. [28] proposed time-varying ARMAX models for ice-storage air-conditioning systems. The model contained several sub-models with different temperature intervals to cope with nonlinearity. The results indicated that a model with different temperature intervals could provide a more accurate prediction than a model with a single interval. Yiu and Wang [29] proposed a recursive extended least squares (RELS) approach for parameter estimation in ARMAX models. They verified several orders of ARMAX models to yield the minimum error in forecasting of air-conditioning system performance.

Autoregressive with exogenous inputs (ARX) models, a special case of ARMAX, have also been used for load forecasting. Yoshida and Inooka [30] applied an ARX model to the rational operation of a TES tank. Yun et al. [31] developed a building hourly thermal load prediction scheme based on hourly indexed ARX models. In their study, several ARX models were proposed with different time and temperature intervals and compared with MLR, auto-regressive (AR), and ANN models. The test results showed that these improved ARX models were better than MLR, AR, and ANN models with the same number of inputs.

Least squares methods are commonly used to estimate the coefficients in regression models. However, the sensitivity to outliers is a major drawback of conventional least squares methods. The presence of contaminated data may significantly decrease the accuracy of such models. Some studies have utilized robust regression algorithms, such as iteratively reweighted least squares (IRWLS), to reduce the effect of outliers on the accuracy of regression models. Mbamalu and El-Hawary [32,33] applied IRWLS to predict autoregressive parameters in seasonal and short-term load forecasting of power systems. They used the partial autocorrelation of past predicted load data to determine the sub-optimal models for hourly load prediction. Their result shows that 97% of predictions have absolute errors less than 10% of the actual values. Moreover, it was concluded that by properly adjusting the tuning constant, the results with Andrew and Fair's weight functions were better than those with other weight functions.

The accuracy of IRWLS can be improved if the appropriate weight function is used based on the probability distribution of the observations. The performance of different solution algorithms and weight functions was investigated in IRWLS [34]. The analysis of the numerical examples in that study showed quadratic programming algorithms are efficient for Huber or Talwar weight functions. However, for other types of weight functions some modification of Newton's algorithm was preferred. A Monte Carlo approach was applied to evaluate the efficiency of eight weight functions in IRWLS with Gaussian distributions [35]. The study showed that a common formula for residual scaling might not work well for all weight functions. Wolke and Schwetlick [36] proposed some algorithms on scaling residuals for given coefficient estimates in an IRWLS procedure. Moreover, the study proposed some algorithms for converting a robust regression model to a nonlinear least squares problem. Bissantz et al. [37] showed that convergence of IRWLS methods is significantly faster than modified Newton algorithms in minimization of convex functions.

Kalyani and Giridhar [38] analyzed the convergence rate of IRWLS and M estimator methods.

Many studies in cooling load forecasting have shown improved prediction accuracy. However, none of the above studies has considered the effect of outliers in cooling load forecasting. In addition, the majority of such studies propose models requiring complex calculations and several inputs. ARX models have proven to be computationally efficient and accurate in cooling load forecasting. Such models could be improved by using robust regression methods that decrease the effect of outliers. Thus, this study presents a two-stage robust regression method to improve the accuracy of ARX models.

3. Model

This section explains the hourly cooling-load prediction method based on time-indexed ARX models. This section also presents the two-stage weighted least squares regression algorithm developed to estimate the coefficients of the proposed ARX models. The two stages consist of robust regression to decrease the effect of outliers and use of an exponential forgetting factor to reflect the relative importance of the most recent data.

3.1. Structure of the prediction method

The structure of the proposed prediction method is shown in Fig. 2. The historical data (weather and cooling load) are preprocessed and classified for hourly-indexed ARX models based on the hours of the day and day types. From the historical data, the ARX model coefficients with different values of forgetting factors are estimated through the proposed two-stage reweighted least squares (left side of Fig. 2). Then, the ARX model predicts future cooling loads by using the coefficients with the optimal forgetting-factor value and the necessary weather and cooling load data.

3.2. ARX models

In this study, two different ARX models are used over different forecasting periods to achieve better accuracy. ARX(1,2,24) and ARX(24,168) are used to predict cooling loads 1–6 h ahead and 7–24 h ahead, respectively. The numbers in the parentheses of the two acronyms refer to the time lag of the inputs. For instance, ARX(1,2,24) means the inputs at time $t-1$, $t-2$, and $t-24$ are used in the model. These two different models are used because studies [39,40] showed that, although the prediction accuracy of stochastic AR models decreases as the forecast time increases, the accuracy of a 24 h forecasting model is better than that of a forecast time greater than 6 h. This can be explained by the fact that, as forecast time increases, the effect of recent hours' information diminishes but the daily repeating pattern becomes more dominant.

Each of the two ARX models contains 48 hourly indexed sub-models for each hour in the two types of day (weekdays and weekends). Such a structure can properly reflect the different characteristics of weekdays and weekends as well as those of different hours. It can also reflect similar day-to-day cyclic patterns in the base load, occupancy level, building open hours, and human activity level.

Because separate models are built in this study for weekdays and weekends, a special rule for $t-24$ is defined. The time $t-24$ always refers to the same hour of the most recent day of the same day type. For example, if the target hour t is 6 AM on Tuesday, $t-24$ is 6 AM on Monday in the same week. If it is 6 AM on Monday, then $t-24$ refers to 6 AM on Friday of the previous week. If the target hour t is 6 AM on Sunday, $t-24$ is 6 AM on Saturday. For 6 AM on Saturday, $t-24$ refers to 6 AM on Sunday of the previous week.

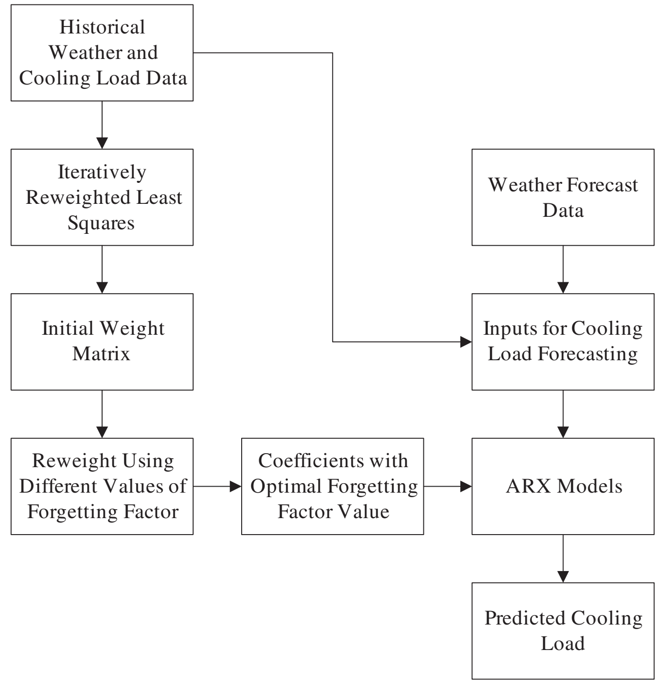

Fig. 2. Flow chart of the prediction scheme.

Without loss of generality and as is often done in practice, this research uses two exogenous input variables: dry bulb ambient temperature and relative humidity. Additional exogenous inputs, such as wind speed, solar radiation and building occupancy, may be added depending on the availability of data and the influence levels.

The ARX(1,2,24) model is formulated as follows.

$$\begin{aligned}CL_t = {} & b_{t,0} + b_{t,1}T_t + b_{t,2}T_t^2 + b_{t,3}RH_t + b_{t,4}T_tRH_t + b_{t,5}CL_{t-1} \\ & + b_{t,6}T_{t-1} + b_{t,7}T_{t-1}^2 + b_{t,8}RH_{t-1} + b_{t,9}T_{t-1}RH_{t-1} \\ & + b_{t,10}CL_{t-2} + b_{t,11}T_{t-2} + b_{t,12}T_{t-2}^2 + b_{t,13}RH_{t-2} \\ & + b_{t,14}T_{t-2}RH_{t-2} + b_{t,15}CL_{t-24} + b_{t,16}T_{t-24} + b_{t,17}T_{t-24}^2 \\ & + b_{t,18}RH_{t-24} + b_{t,19}T_{t-24}RH_{t-24} \end{aligned} \tag{1}$$

where CL is the cooling load at each time point t, and $b_{t,i}$ is the regression coefficient. T and RH are the ambient temperature and relative humidity. Higher order and interaction terms are also included to model nonlinearity.

The ARX(24,168) model is formulated in a similar way.

$$\begin{aligned}CL_t = {} & b_{t,0} + b_{t,1}T_{Max} + b_{t,2}T_{Min} + b_{t,3}T_t + b_{t,4}T_t^2 + b_{t,5}RH_t \\ & + b_{t,6}T_tRH_t + b_{t,7}CL_{t-24} + b_{t,8}T_{t-24} + b_{t,9}T_{t-24}^2 \\ & + b_{t,10}RH_{t-24} + b_{t,11}T_{t-24}RH_{t-24} + b_{t,12}CL_{t-168} \\ & + b_{t,13}T_{t-168} + b_{t,14}T_{t-168}^2 + b_{t,15}RH_{t-168} \\ & + b_{t,16}T_{t-168}RH_{t-168} \end{aligned} \tag{2}$$

where T_{Max} and T_{Min} are the daily maximum and minimum dry bulb temperatures, respectively, for the target hour. Binary variables are used to differentiate holidays and other special event days when the cooling load is expected to be significantly higher or lower than normal days. The data of those days will also be labeled and recorded in the database, so that they will not be used to predict the cooling load for the following normal days after those holidays.

The explanatory variables in the proposed ARX models were chosen based on the past analysis of the cooling load and previous studies in the literature. The accuracy of the models depends on the selection of proper explanatory variables, and simply increasing the number of the variables may result in over-fitting by including insignificant variations in the model. Section 4 discusses the suitability of the chosen variables in detail, and examines the effect of excluding some variables on the calculation complexity and accuracy of the models.

3.3. Two-stage weighted least squares

This section explains the two-stage weighted least-squares regression algorithm. In the first stage, the initial weight matrix is generated and outliers are labeled using IRWLS with the bisquare weighting function. The first stage is necessary for the following reasons: (1) outliers usually exist in data due to many reasons such as incorrect readings, temporary shut-down for maintenance, or other special events, and (2) in regression models, the existence of outliers could result in poor coefficient estimation that can significantly reduce regression accuracy. Therefore, detecting outliers and reducing the outlier effect are important for obtaining the accurate coefficients.

In the second stage, the initial weight matrix will be reweighted by using an exponential forgetting factor with different values. The sum of squared errors for the past 90 days is calculated using those regression coefficients with different values of the exponential forgetting factor. The optimal coefficients with the least sum of squared errors of the past 90 days are then used for prediction. This second stage ensures the adaptiveness of the coefficient estimation. The adaptiveness is important, because the cooling load pattern tends to change over time. For example, the cooling load pattern can change significantly owing to construction of new buildings and implementation of new cooling systems with energy saving features.

Generally, a regression model with p inputs and n instances ($n > p$) can be written in a matrix form as follows.

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{3}$$

where $\mathbf{Y}$ is an n-by-1 vector of output (or dependent) variables and $\mathbf{Y} = [y_1, y_2, \ldots, y_n]^T$. $\mathbf{X}$ is an n-by-(p + 1) matrix of input (or independent) variables. $\mathbf{X}$ can also be written as $[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$ or $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$, where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ are n-by-1 vectors of each input variable and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are 1-by-(p + 1) vectors of each instance. $\mathbf{b}$ is a (p + 1)-by-1 vector of regression coefficients. $\mathbf{e}$ is an n-by-1 vector of errors with a standard deviation of σ.

The conventional least squares method minimizes the sum of squared errors.

$$\min z = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \mathbf{x}_i\mathbf{b})^2 \tag{4}$$

To meet the above requirement, the coefficients are estimated through the following equation.

$$\hat{\mathbf{b}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} \tag{5}$$

If a weight is added to each squared error, the conventional least squares regression becomes a weighted least squares regression, and Eqs. (4) and (5) change to the following equations.

$$\min z = \sum_{i=1}^{n} W_{ii} e_i^2 = \sum_{i=1}^{n} W_{ii}(y_i - \mathbf{x}_i\mathbf{b})^2 \tag{6}$$

$$\hat{\mathbf{b}} = (\mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{Y} \tag{7}$$

where $\mathbf{W} = \mathrm{diag}(W_{11}, W_{22}, \ldots, W_{nn})$.

If we denote $\mathbf{X}' = \mathbf{w}\mathbf{X}$ and $\mathbf{Y}' = \mathbf{w}\mathbf{Y}$ with $\mathbf{w} = (\sqrt{W_{11}}, \sqrt{W_{22}}, \ldots, \sqrt{W_{nn}})^T$, we will have a new least squares form with weighted input $\mathbf{X}'$ and weighted output $\mathbf{Y}'$.

$$\mathbf{Y}' = \mathbf{X}'\mathbf{b}' + \mathbf{e}' \tag{8}$$

$$\min z' = \sum_{i=1}^{n} e_i'^2 = \sum_{i=1}^{n} (y_i' - \mathbf{x}_i'\mathbf{b}')^2 \tag{9}$$

$$\hat{\mathbf{b}}' = (\mathbf{X}'^{\mathbf{T}}\mathbf{X}')^{-1}\mathbf{X}'^{\mathbf{T}}\mathbf{Y}' \tag{10}$$

In the first stage of the two-stage weighted least squares procedure, the initial weight matrix is calculated through IRWLS. The IRWLS approach for robust regression focuses on minimizing the sum of squared errors of estimates by using a weight function that reduces the effect of outlying data points.

The evaluation and discussion of different weight functions is beyond the scope of this paper. In this paper, a bisquare function, often used in practice, is used as the weight function. The bisquare weight function proposed by Tukey (1974) as appeared in [41] is as follows.

$$W_{ii}(r_i) = \begin{cases} (1 - u_i^2)^2 & |u_i| \leqslant 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $u_i = \frac{(y_i - \hat{y}_i)}{cS\sqrt{1-h_i}}$. h_i is the leverage, $c = 4.685$ is a tuning constant chosen based on 95% efficiency of the robust regression to least squares method, and S = MAD/0.6745 where MAD is the median of the absolute deviation of the residuals. The solution procedure starts with initial estimates for $\mathbf{b}$ and σ. In each iteration the weighted least squares procedure is applied and estimates of $\mathbf{b}$, σ and r_i are updated. The procedure is continued until convergence. The convergence criterion based on [33] is:

$$\frac{(\sqrt{\mathbf{W}}\mathbf{b})^{\mathbf{T}}(\sqrt{\mathbf{W}}\sigma\mathbf{r})}{\|\sqrt{\mathbf{W}}\mathbf{b}\|_2\|\sqrt{\mathbf{W}}\mathbf{b}\|_2} \tag{12}$$

where $\|\cdot\|$ is the Euclidean norm.

In the second stage, a new weight matrix $\mathbf{W}' = \mathrm{diag}[W_{ii}'] = \mathrm{diag}[\lambda^{n-i}]$ is used in the weighted least squares again for the new problem described by Eqs. (8)–(10). Therefore, the new least squares problem is defined as follows.

$$\mathbf{Y}'' = \mathbf{X}''\mathbf{b}'' + \mathbf{e}'' \tag{13}$$

$$\min z'' = \sum_{i=1}^{n} e_i''^2 = \sum_{i=1}^{n} (y_i'' - \mathbf{x}_i''\mathbf{b}'')^2 \tag{14}$$

$$\hat{\mathbf{b}}'' = (\mathbf{X}''^{\mathbf{T}}\mathbf{X}'')^{-\mathbf{1}}\mathbf{X}''^{\mathbf{T}}\mathbf{Y}'' \tag{15}$$

where $\mathbf{X}'' = \mathbf{w}'\,\mathbf{X}'$ and $\mathbf{Y}'' = \mathbf{w}'\,\mathbf{Y}'$.

Thus, the final regression coefficients are estimated through the two-stage weighted least squares.

4. Results and discussion

This section presents the test results of the proposed prediction method. The test is performed on a large-scale cooling system for the campus of an academic institution with 7,580,265 square feet of indoor area. The cooling is provided by chilled-water based cooling systems. The test is performed during the cooling period of 2012 (April 1 to October 31). The historical data from the same cooling periods in 2010 and 2011 are used as the initial learning data set. The total capacity of the chilled water production by six electric chillers and one steam turbine was 23,000 tons. The weather characteristics were obtained from the National Oceanic and Atmospheric Administration (NOAA).

The proposed prediction scheme was implemented by a computer code written in Matlab®, and run on a workstation with Intel® Xeon® 3.40 GHz CPU and 32 GB memory. The calculation time for the two ARX models were around 59 and 37 s, respectively.

An example of the coefficients of the two models for $t = 14$ in weekdays is given in Table 1. Because the ARX models contain large number of coefficients (48 sets for each model), the rest of the coefficients are not presented in this paper.

The predicted values of the proposed ARX models is compared with the actual values of cooling load for five consecutive weekdays of summer 2012. As can be seen in Fig. 3, the forecast is close to the actual values observed. The deviations are analyzed in detail in the remainder of this section.

The performance of the two ARX models proposed in this paper is further compared with those of the selected ANN and ARX models in the literature. First, the ARX model's performance is compared with those of hourly-indexed ANN models through back-propagation with the same number of inputs. The ANN models have one hidden layer, which contains 40 and 30 neurons in that hidden layer for one-hour-ahead and 24-h-ahead models, respectively. The network uses a hyperbolic tangent sigmoid activation function and is trained in 3000 epochs using the data of 2010 and 2011. Second, the performance of the proposed one-hour-ahead forecasting ARX model is compared with that of a recent 4-3-5 ARX model [31]. The temperature intervals and building occupancy for the 4-3-5 ARX model are customized to include sufficient data points in each temperature interval. The 24-h-ahead model ARX(24,168) is compared with the 4-8-24 ARX model [42]. The benchmark ARX models are hourly indexed, similarly to the models in this study. Moreover, the 4-8-24 ARX model [42] follows a logic similar to that of this study for distinguishing weekdays and weekends for $t - 24$.

The performance of each model is investigated in terms of four statistical indices introduced in [26] which measure the accuracy of predicted values by comparing them with the actual values: (1) standard error of estimate (*SEE*), (2) expected error percentage (*EEP*), (3) coefficient of variation (*CV*), and (4) mean bias error (*MBS*). These are calculated as:

$$SEE = \sqrt{\sum_{t=1}^{n} (\overline{CL}_t - CL_t)^2 \Big/ n} \quad (16)$$

$$EEP = \frac{SEE}{Max_{t=1,2,\ldots,n}\{CL_t\}} \times 100\% \quad (17)$$

$$CV = \frac{SEE}{\sum_{t=1}^{n} CL_t \Big/ n} \times 100\% \quad (18)$$

$$MBE = \frac{1}{n}\sum_{t=1}^{n} \frac{\overline{CL}_t - CL_t}{CL_t} \times 100\% \quad (19)$$

where $\overline{CL}_t$ is the predicted cooling load, CL_t is the actual recorded cooling load, and n is the total number of data points in the testing data set.

The comparison reveals that the forecast by the proposed model is closer to the actual vales than those by the other benchmark models; the model proposed in this study has improved accuracy in cooling load prediction by almost 10%. Table 2 shows the comparison of predictions with one hour of forecast time in terms of the four evaluation indices described above. The results indicate better accuracy from the proposed ARX(1,2,24) model in this study compared to the benchmark ANN and 4-3-5 ARX models. Figs. 4 and 5 present the SEEs of the prediction with one-hour ahead of forecast time for each hour in weekdays and weekends. Overall, the ARX(1,2,24) model is better than the ANN and 4-3-5 ARX models for most of the hours in a day. The ARX(1,2,24) model is not better than the ANN model only for 10 AM, but it is better than 4-3-5 ARX.

As shown in Table 2, there is improved prediction from the proposed scheme for the data set. The two-stage weighted least squares regression ensures that the regression coefficients are estimated without any influence from outliers. Thus, a significant improvement in accuracy is achieved. In contrast, it is shown that the existence of outliers decreases the accuracy of the 4-3-5 ARX model. Although it was claimed in [31] that the 4-3-5 ARX model was more accurate than the ANN model, the advantage of the 4-3-5 ARX model diminishes due to its lack of robustness as compared to the ANN model.

Figs. 4 and 5 also indicate different patterns for weekdays and weekends: (1) the SEEs of 9 AM to 6 PM in weekdays are significantly higher than those of the same period during weekends, and (2) the SEE of each hour in weekends is more stable than that in weekdays. A possible explanation for these different patterns is the different levels of variation of human activities on the campus. More human activities and higher occupancy on campus will increase the heat gain inside buildings and cooled air loss through building entrances. The increase in internal heat gain and cooling loss results in increased demand for cooling load. However, in contrast with the cooling load prediction for a standalone building, the occupancy and the level of human activities on the whole campus are difficult to measure and have different patterns from day to day. This means that the ARX model introduced in this paper is able to model average human activities and occupancy even without the data of occupancy and human activities. If the dynamic part of human activities is significant, such information becomes important for a more accurate prediction. Therefore, such information should be added to the model.

Table 3 shows the comparison of test results of 24-h-ahead prediction. Overall, the proposed ARX(24,168) model has a better performance than the compared ANN and 4-8-24 ARX models. However, the ARX(24,168) model is not as good as the ANN model in terms of the mean bias error. Similar to the one-hour-ahead prediction case, different patterns for weekdays and weekends can be seen clearly from Figs. 6 and 7.

Table 1
The coefficient values of ARX models for $t = 14$ in weekdays.

ARX(1,2,24)						ARX(24,168)					
$b_{14,0}$	−95.26	$b_{14,8}$	−8.38	$b_{14,16}$	−7.14	$b_{14,0}$	−222.34	$b_{14,8}$	10.06	$b_{14,16}$	−0.22
$b_{14,1}$	−34.83	$b_{14,9}$	0.10	$b_{14,17}$	0.03	$b_{14,1}$	57.53	$b_{14,9}$	−0.44		
$b_{14,2}$	0.30	$b_{14,10}$	−0.03	$b_{14,18}$	−4.19	$b_{14,2}$	12.35	$b_{14,10}$	21.91		
$b_{14,3}$	−3.25	$b_{14,11}$	−21.49	$b_{14,19}$	0.06	$b_{14,3}$	−99.88	$b_{14,11}$	−0.60		
$b_{14,4}$	0.06	$b_{14,12}$	−0.06			$b_{14,4}$	0.96	$b_{14,12}$	0.08		
$b_{14,5}$	1.02	$b_{14,13}$	7.43			$b_{14,5}$	−80.54	$b_{14,13}$	21.43		
$b_{14,6}$	76.49	$b_{14,14}$	−0.11			$b_{14,6}$	1.58	$b_{14,14}$	−0.15		
$b_{14,7}$	−0.39	$b_{14,15}$	0.00			$b_{14,7}$	0.55	$b_{14,15}$	13.76		

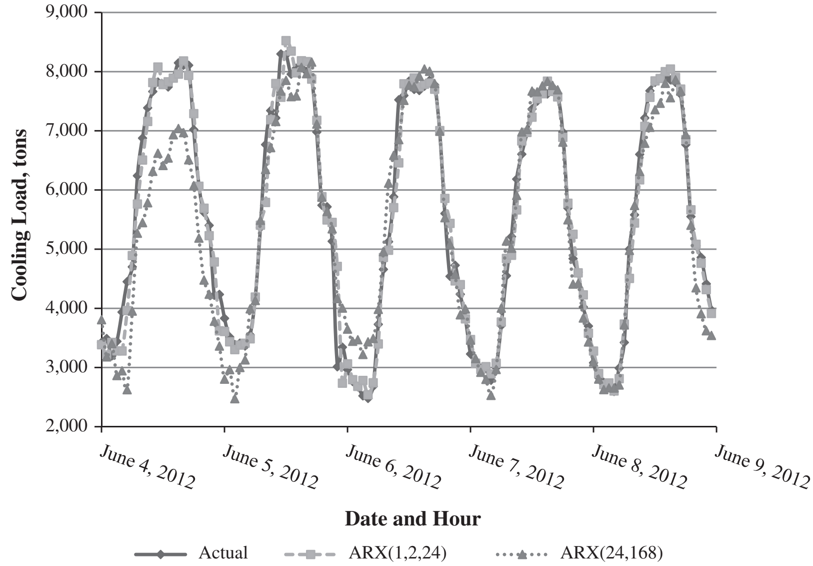

Fig. 3. Comparison of actual and predicted values in June 4th–8th 2012 (weekday).

Table 2
Overall test result of one-hour-ahead prediction using different models throughout the entire testing period.

Model	ANN	4-3-5 ARX	ARX(1,2,24)
SEE (tons)	409.79	395.62	360.25
EEP (%)	3.36	3.24	2.96
CV (%)	8.40	8.11	7.39
MBE (%)	−3.05	−0.39	−0.53

Table 3
Overall test result of 24-hours-ahead prediction using different models throughout the entire testing period.

Model	ANN	4-8-24 ARX	ARX(24,168)
SEE (tons)	746.06	940.67	697.45
EEP (%)	6.11	7.71	5.71
CV (%)	15.29	19.28	14.29
MBE (%)	3.24	4.50	3.34

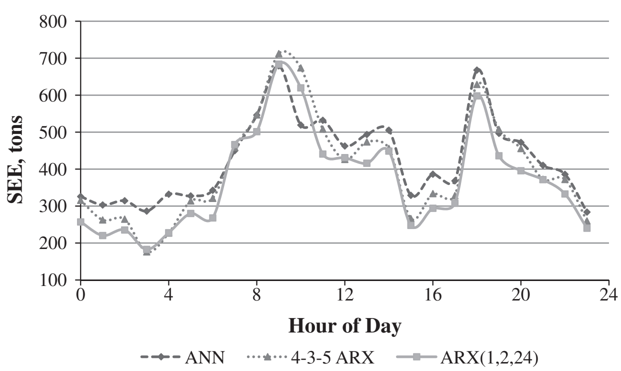

Fig. 4. SEEs of one-hour-ahead prediction for each hour on weekdays.

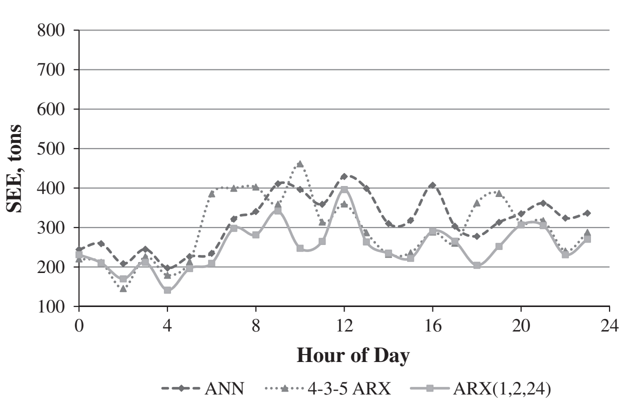

Fig. 5. SEEs of one-hour-ahead prediction for each hour on weekends.

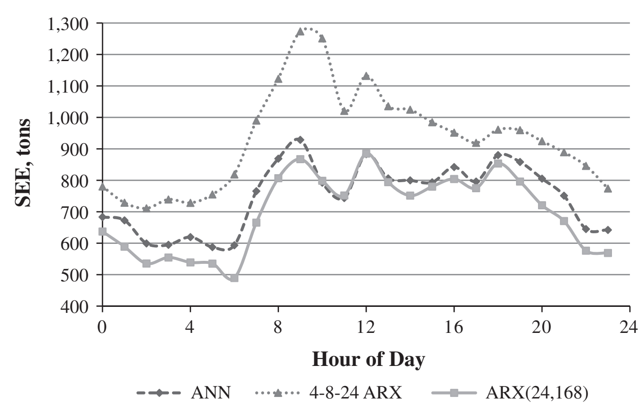

Fig. 6. SEEs of 24-hours-ahead prediction for each hour on weekdays.

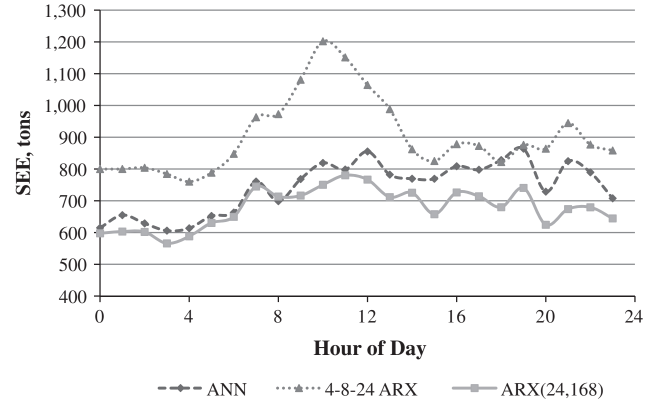

Fig. 7. SEEs of 24-hours-ahead prediction for each hour on weekends.

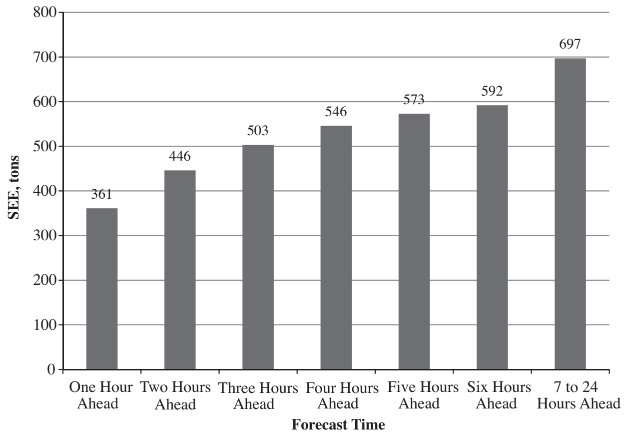

Fig. 8. Overall SEEs with different forecast times in 1–6 h ahead forecast by ARX(1,2,24) and 7–24 h ahead forecast by ARX(24,168).

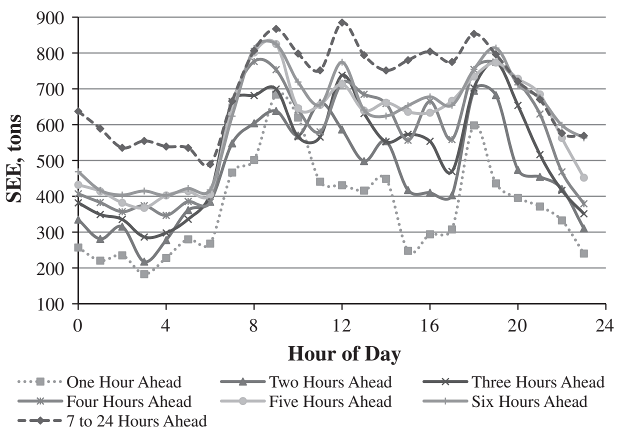

Fig. 9. SEEs of the prediction with different forecast times for each hour on weekdays.

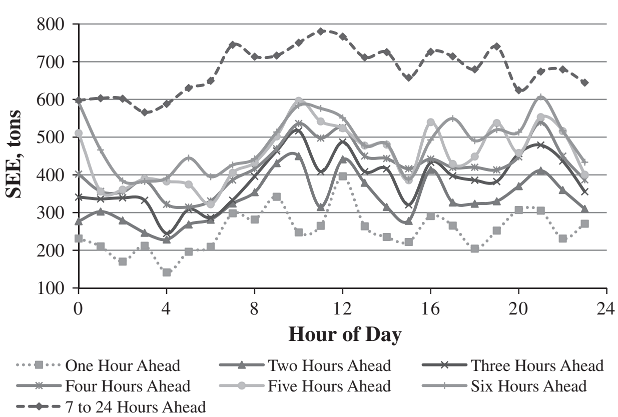

Fig. 10. SEEs of the prediction with different forecast times for each hour on weekends.

Figs. 8–10 show the SEEs with different forecast times by the proposed prediction scheme. As described in the model section, the prediction with 1–6 h of forecast time is estimated using the ARX(1,2,24) model, while the prediction with 7–24 h of forecast time is estimated using the ARX(24,168) model. As shown in the figures, the overall SEE increases, as the forecast time increases from 1 to 6. Although the overall SEE of the ARX(1,2,24) model with 6 h of forecast time is still lower than that of the ARX(24,168) model in Fig. 8, the ARX(1,2,24) model is not much better than the ARX(24,168) model at hours 19–23 in weekdays as shown in Fig. 9. Therefore, there is not much benefit from using the ARX(1,2,24) model to predict the cooling load with 7 h or more of forecast times. However, as shown in Fig. 10, the ARX(1,2,24) is better than ARX(24,168) model in all hours of weekends.

The analysis shows that the number of variables in the ARX models is appropriate in terms of model adequacy and parsimony. Including additional terms in a model, such as more variable interaction terms or higher order terms, did not decrease the SEE for the testing data sets. The model efficiency was also verified. For instance, removing T and RH interaction terms did not change the calculation time while SEE increased to 365 and 771 tons for ARX(1,2,24) and ARX(24,168), respectively. A possible reason for that is T and RH have a large correlation for the cooling load in this study. Also, removing T_{Max} and T_{Min} increased MSE. For example, the SEE increased to 774 in ARX(24,168).

5. Conclusions

This paper introduced an hourly cooling-load prediction method using hourly-indexed ARX models with a two-stage weighted least-squares regression. The proposed method was tested using the cooling load data in an academic institution, and the prediction accuracy was compared with some ARX and ANN models. The results demonstrated that the forecasting method in this study improved the prediction accuracy through the proper combination of the two ARX models for different forecast times as well as the differentiation of day types and outliers in historical load data.

The proposed prediction method includes the following characteristics. First, the two-stage weighted least-squares regression incorporates robustness and adaptiveness. The robust regression by IRWLS is effective in mitigating the outlier influence on coefficient estimation. The exponential forgetting factor ensures that the recent cooling load and weather patterns contribute more than past ones in estimating coefficients. These lead to improved accuracy. Second, the ARX models in this study include higher order terms of temperature and interaction terms among exogenous variables to reflect the nonlinearity and inter-correlation between input parameters, whereas many of the existing cooling-load forecasting studies use piecewise linearization to manage nonlinearity. Although some higher order terms are included in the ARX models of this study, the models are parsimonious. The total number of forecasting variables is less than that of some recent studies in cooling load forecasting. The reason is that the majority of existing studies include additional coefficient sets based on temperature intervals and demand periods, e.g. day, night and transition hours between day and night, to handle the nonlinearity. The reduction in coefficient sets results in decreased computational complexity in comparison with ARX models in the recent literature.

Acknowledgement

This study was partially supported by the Nebraska Center for Energy Sciences Research and the New Faculty Research Fund of Ajou University.

References

[1] Yao Y, Chen J. Global optimization of a central air-conditioning system using decomposition coordination method. Energy Build 2010;42(5):570–83.

[2] Guo Y, Ko J. Cost optimization for a large-scale hybrid central cooling plant with multiple energy sources under a complex electricity cost structure. HVAC&R Res 2013;19(6):754–63.

[3] Yao Y, Wang L. Energy analysis on VAV system with different air-side economizers in China. Energy Build 2010;42(8):1220–30.
[4] Lam JC, Wan KK, Liu D, Tsang C. Multiple regression models for energy use in air-conditioned office buildings in different climates. Energy Convers Manage 2010;51(12):2692–7.
[5] Aranda A, Ferreira G, Mainar-Toledo M, Scarpellini S, Llera-Sastresa E. Multiple regression models to predict the annual energy consumption in the Spanish banking sector. Energy Build 2012;49:380–7.
[6] Duanmu L, Wang Z, Zhai ZJ, Li X. A simplified method to predict hourly building cooling load for urban energy planning. Energy Build 2013;58:281–91.
[7] Sun Y, Wang S, Xiao F. Development and validation of a simplified online cooling load prediction strategy for a super high-rise building in Hong Kong. Energy Convers Manage 2013;68:20–7.
[8] Leung M, Tse NC, Lai L, Chow T. The use of occupancy space electrical power demand in building cooling load prediction. Energy Build 2012;55:151–63.
[9] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001;33(4):319–31.
[10] Crawley DB, Hand JW, Kummert M, Griffith BT. Contrasting the capabilities of building energy performance simulation programs. Build Environ 2008;43(4):661–73.
[11] Carriere M, Schoenau G, Besant R. Investigation of some large building energy conservation opportunities using the DOE-2 model. Energy Convers Manage 1999;40(8):861–72.
[12] Kalogirou SA. Use of TRNSYS for modeling and simulation of a hybrid PV–thermal solar system for Cyprus. Renew Energy 2001;23(2):247–60.
[13] Barrios JA, Torres-Alvarado M, Cavazos A. Neural, fuzzy and Grey-Box modeling for entry temperature prediction in a hot strip mill. Exp Syst Appl 2012;39(3):3374–84.
[14] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renew Sust Energy Rev 2012;16(6):3586–92.
[15] Kalogirou SA. Applications of artificial neural-networks for energy systems. Appl Energy 2000;67(1):17–35.
[16] Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 2008;40(12):2169–76.
[17] Zhou Q, Wang S, Xu X, Xiao F. A grey box model of next day building thermal load prediction for energy efficient control. Int J Energy Res 2008;32(15):1418–31.
[18] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Convers Manage 2009;50(1):90–6.
[19] Li X, Shao M, Ding L, Xu G, Li J. Particle swarm optimization-based LS-SVM for building cooling load prediction. J Computers 2010;5(4):614–21.
[20] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86(10):2249–56.
[21] Pai P, Hong W. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers Manage 2005;46(17):2669–88.
[22] Yao Y, Lian Z, Liu S, Hou Z. Hourly cooling load prediction by a combined forecasting model based on analytic hierarchy process. Int J Therm Sci 2004;43(11):1107–18.
[23] Ben-Nakhi AE, Mahmoud MA. Cooling load prediction for buildings using general regression neural networks. Energy Convers Manage 2004;45(13):2127–41.
[24] Yao Y, Lian Z, Hou Z, Liu W. An innovative air-conditioning load forecasting model based on RBF neural network and combined residual error correction. Int J Refrig 2006;29(4):528–38.
[25] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Trans Power Syst 1989;4(4):1484–91.
[26] Kawashima M, Dorgan CE, Mitchell JW. Hourly thermal load prediction for the next 24 h by ARIMA, EWMA, LR, and an artificial neural network. ASHRAE Trans 1995;101(1):186–200.
[27] Yang HT, Huang CM, Huang CL. Identification of ARMAX model for short term load forecasting: an evolutionary programming approach. In: Power industry computer application conference, IEEE conference proceedings; 1995. p. 325–30.
[28] Xu J, Xiao R, Huang C, Gao R, Feng Z. ARMAX model of ice-storage air conditioning system load based on temperature interval. J Wuhan Univ Technol 2009;10(029):109–12.
[29] Yiu JC, Wang S. Multiple ARMAX modeling scheme for forecasting air conditioning system performance. Energy Convers Manage 2007;48(8):2276–85.
[30] Yoshida H, Inooka T. Rational operation of a thermal storage tank with load prediction scheme by ARX model approach. In: Proceeding of building performance simulations; 1997.
[31] Yun K, Luck R, Mago PJ, Cho H. Building hourly thermal load prediction using an indexed ARX model. Energy Build 2012;54:225–33.
[32] Mbamalu G, El-Hawary M. Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. IEEE Trans Power Syst 1993;8(1):343–8.
[33] El-Hawary M, Mbamalu G. Short-term power system load forecasting using the iteratively reweighted least squares algorithm. Electr Power Syst Res 1990;19(1):11–22.
[34] O'Leary DP. Robust regression computation using iteratively reweighted least squares. SIAM J Matrix Anal Appl 1990;11(3):466–80.
[35] Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. Commun Stat-Theor Methods 1977;6(9):813–27.
[36] Wolke R, Schwetlick H. Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. SIAM J Sci Stat Comput 1988;9(5):907–21.
[37] Bissantz N, Dümbgen L, Munk A, Stratmann B. Convergence analysis of generalized iteratively reweighted least squares algorithms on convex function spaces. SIAM J Optim 2009;19(4):1828–45.
[38] Kalyani S, Giridhar K. MSE analysis of the iteratively reweighted least squares algorithm when applied to M estimators. In: Global telecommunications conference, IEEE; 2007.
[39] Seem J, Braun J. Adaptive methods for real-time forecasting of building electrical demand. ASHRAE Trans 1991;97(1):710–21.
[40] Henze GP, Dodier RH, Krarti M. Development of a predictive optimal controller for thermal energy storage systems. HVAC&R Res 1997;3(3):233–64.
[41] Dumouchel W, O'Brien F. Integrating a robust option into a multiple regression computing environment. Inst Math Appl 1991;36:41–8.
[42] Forrester J, Wepfer W. Formulation of a load prediction algorithm for a large commercial building. ASHRAE Trans 1984;90(2):536–51.