DEVELOPMENT OF LOAD PROFILE PREDICTION USING TCBM AND ARIMA HYBRID-MODELING

Tadahiko MATSUBA*, Hiroaki TSUTSUI**
and Kazuyuki KAMIMURA**
*Yamatake Building Systems Co., Ltd., **Yamatake Corp.
*Tokyo 1086030, **Fujisawa 2518522

ABSTRACT

For effective energy use and level off electricity demand throughout the day, thermal energy storage system is getting popular in HVAC systems in Japan. Load profile prediction can provide information to plan optimal operation, and now, various prediction approaches have been developed, for example, Autoregressive Integrated Moving Average (ARIMA) model, Topological Case Based Modeling (TCBM), Artificial Neural Network (ANN), and other approaches. To be considered characteristics of these approaches, hybrid prediction has have been applied.

In this paper, load profile prediction approach using TCBM and ARIMA hybrid-modeling that has been developed for use in the optimal operation of thermal energy storage system is proposed. Performance of this approach was assessed by benchmark test held by the technical committee for optimization of thermal energy storage systems (TC-OTES) of heating, air-conditioning and sanitary engineers in Japan (SHASE). Results show good agreement between actual and predicted load.

INTRODUCTION

In electrically operated HVAC systems, the introduction of thermal energy storage systems can help level off electricity demand throughout the day and thereby increase the overall operation coefficient of power plants run by utility companies. As a result, construction of additional plants can be avoided, and a contribution can be made to the stability and safety of power plants. Thermal energy storage systems can thus help protect the global environment, particularly by reducing CO2 emissions, which is one of the most important environmental issues we face today.

In reality, however, thermal energy storage systems often do not operate as efficiently as expected at the design stage. Typical reasons for this are:
Excessive storage of thermal energy leads to significant heat loss through tank surroundings; and peak operation of energy plants becomes necessary because stored energy is completely discharged early in the day. In order to avoid these problems, a strategy for optimum operation is needed. Predicting air-conditioning loads is one essential component of such a strategy.

For this purpose, thermal load prediction should have some performances and functions as followed:
1) Prediction of hourly load profiles within several hours or a few days future.
2) Prediction accuracy for planing the operation of the plants.
3) Adaptability to the changing of characteristics of the thermal load.
4) Automatic predicting procedures.
Especially in the case of using thermal energy storage systems, the daily load is necessary to plan storage of thermal energy. For these requests, ARIMA model can predict the load pattern precisely. On the other hand, TCBM can treat the nonlinear factors of load level fluctuations as the input variables. TCBM and ARIMA hybrid-modeling approach has the advantages of both technologies, and satisfies above requested performances and functions for load prediction.

In this paper, TCBM and ARIMA hybrid-modeling approach that has been developed for use in the optimal operation of thermal energy storage system is proposed. Performance of this approach was assessed by the open benchmark tests in Japan. The results show good agreement between actual and predicted load.

TCBM OVERVIEW

TCBM is a modeling technology established by applying the framework of case-based reasoning to modeling. It is based on the concept of topology and is applicable to a general target for which continuity can be established in the input/output relationship.
Conventional modeling technology identified model parameters including the degree of a model and the network structure. TCBM identifies the input space topology by specifying the output error limit. TCBM features data accumulated as a case in the identified input space and the reliability to the estimated output value indicated by the topological distance (similarity degree) between inputs and accumulated input cases during output estimation (Figure 1).

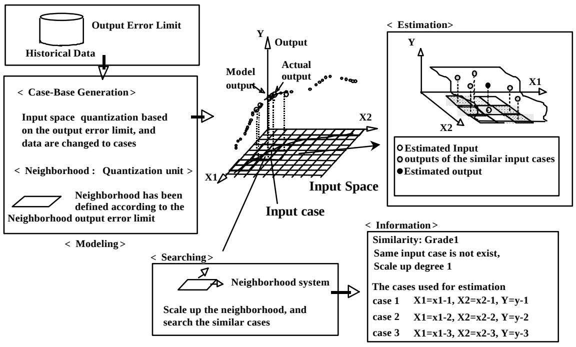

Figure 1 TCBM overview

Conventional modeling has certain problems shown Table 1.

These problems can be resolved by using the framework of Case-Based Reasoning (CBR). CBR finds the solution by directly using past cases similar to a given problem and has the basic framework (Table 2).

Table 1 Problems in Modeling

Problem 1	A special model structure is used to regulate inputs and outputs. Large amounts of time and labor are needed to find an optimum structure.
Problem 2	When learning with an enormous amount of data, convergent calculation must be performed to identify the parameters belonging to the model structure, greatly increasing the total calculation time
Problem 3	It is difficult to grasp whether the output values of a model are reliable to the input values for estimation.
Problem 4	Parameter identification must be executed each time new information is incorporated, causing adaptive learning to be difficult.

Table 2 Case-Based Estimation

Theory 1	Case (problems/solutions) experienced in the past are accumulated in the case base.
Theory 2	Existing cases with the problems similar to a new problem are searched from the case base.
Theory 3	Solutions to similar searched cases are modified to obtain a solution to the new problem.
Theory 4	After a solution to the new problem is found, that new case is added to the case base.

Table 3 Responses to Modeling Problems

Response to Problem 1	Cases involve the I/O relationship of the system, thus, no special model structure is needed to express the I/O relationship.
Response to Problem 2	TCBM quantizes the input space using the quantization number parameter; this defines the case base and similarity degree. At that time, verification index values (output distribution satisfaction rate, continuity condition satisfaction rate) based on the above concept are merely calculated. With the calculation results, the quantization number is verified. Thus, convergent calculation is not needed but the processing speed can be increased by software methods. Also, the model completion can be proven from the verification index values without using test data.
Response to Problem 3	The similarity degree of searched cases can be determined against the input values for estimation. This similarity degree can be used for reliability verification of output values.
Response to Problem 4	New information can be easily incorporated by partial revision of the case base without performing another parameter identification.

Through application of CBR, the problems in Table 1 can be accommodated as shown Table 3.

However, Problem 2 becomes a problem of definition of the case base structure and similarity degree in CBR. This is a serious engineering problem that the conventional CBR cannot define the structure and similarity degree without sufficient knowledge concerning the target (Problem 1 in the conventional modeling technique will switch to Problem 2 in CBR).

Then, TCBM creates a general definition of the case base and similarity degree in accordance with the output error limit (requested accuracy) by quantizing the input space to be a topological space based on the concepts of continuous mapping in mathematical topology.

ARIMA MODEL OVERVIEW

ARIMA model is a statistical approach based on time series analysis. In the statistical approach, a statistical model is fitted to the observed data. By using an appropriate statistical model, the procedure discussed in this paper provides a model that takes into account the characteristics of both the load profiles and the noise. As the basic model for the statistical approach, the ARIMA model is adopted.

The target time series such as the load profiles of HVAC systems contain trend components that vary daily as well as seasonally. Therefore, the observed time series have nonstationary and periodic components and cannot be considered to be a stationary time series unless these components are eliminated. When applying the ARIMA model to such a nonstationary time series $\{x_t\}$, Box and Jenkins proposed the use of a time series $\{y_t\}$ derived from

$$y_t=(1-z^{-1})^d x_t \tag{1}$$

which is prepared by making the time-lag z^{-1} and applying $(1-z^{-1})$ to $\{x_t\}$ d times. Here, the prediction model can be written as follows:

$$A(z^{-1})(1-z^{-1})^d x_t=B(z^{-1})e_t$$

where (2)

$$A(z^{-1})=1+a_1z^{-1}+...+a_pz^{-p}$$
$$B(z^{-1})=1+b_1z^{-1}+\dots+b_qz^{-q}$$

and where p and q are the orders of the AR and MA processes and $\{e_t\}$ is the time series of the residual error. This is called an autoregressive integrated moving average (ARIMA) model, represented by ARIMA(p,d,q).

When $\{x_t\}$ contains a period component with an elementary period of s, $(1-z^{-s})$ is applied to $\{x_t\}$ d_1 times, and the ARMA(p,q) model is applied to time series $\{y_t\}$, derivered by

$$y_t=(1-z^{-s})^{d1}x_t \tag{3}$$

the following model is obtainded:

$$A(z^{-1})(1-z^{-s})^{d1}x_t=B(z^{-1})c_t. \tag{4}$$

Next, the periodic variation pattern is obtained from the time series $\{c_t\}$. By taking $c_{t1},c_{t1+s},c_{t1+2s},\dots$for any time t_l within the elementary period, the ARMA(p_1,q_1) model is applied to this time series and the following model is obtained:

$$P(z^{-s})c_t=Q(z^{-s})e_t$$
where
$$P(z^{-s})=1+m_1z^{-1}+\dots+m_{p1}z^{-p1s} \tag{5}$$
$$Q(z^{-s})=1+n_1z^{-1}+\dots+n_{q1}z^{-q1s}$$

The prediction model for a time series containing an elementary period of s is obtained from Equations 4 and 5 as follows:

$$P(z^{-s})A(z^{-1})(1-z^{-s})^{d1}x_t=Q(z^{-s})B(z^{-1})e_t. \tag{6}$$

Furthermore, if $\{x_t\}$ has trend components and periodicity, Equation 6 is rewritten as Equation 7:

$$P(z^{-s})A(z^{-1})(1-z^{-1})^d(1-z^{-s})^{d1}x_t=Q(z^{-s})B(z^{-1})e_t \tag{7}$$

where $\{e_t\}$ is a white-noise sequence. This model is called ARIMA(p,d,q)x(p_1,d_1,q_1).

To determine the order of the model, Akaike's Information Criteria (AIC) is calculated and the order providing the minimum AIC is used.

$$\mathrm{AIC}=N\log\sigma_e^2+2(p+q) \tag{8}$$

Next, the algorithm used to obtain the prediction values from the ARMA model will be described. When the current time is represented by t, the prediction value of x_{t+l} at time $(t+l)$ is represented by $\dot{x}_t(l)$, and the conditional expected value of x_{t+l} when the observed values up to time t are given is represented by

$$E[x_{t+l}|x_t,x_{t-1},\dots]=\underset{t}{E}[x_{t+l}] \tag{9}$$

the following relationship applies [A1]:

$$\underset{t}{E}[x_{t+l}]=\hat{x}_t(l). \tag{10}$$

That is, the prediction values equal the expected conditional values at time t of x_{t+l}.

In the ARMA(p,q) model,

$$\hat{x}_t(l)=\underset{t}{E}[x_{t+l}]=-\sum_{i=1}^{p}a_i\,\underset{t}{E}[x_{t+l-i}] \tag{11}$$
$$+\sum_{i=1}^{q}b_i\,\underset{t}{E}[e_{t+l-i}]+\underset{t}{E}[e_{t+l}]$$

these values can be obtained as follows:

$$\begin{aligned}\underset{t}{E}[x_{t-j}]&=x_{t-j},\ j=0,1,2,\dots\\ \underset{t}{E}[x_{t+j}]&=\hat{x}_t(j),\ j=1,2,\dots\\ \underset{t}{E}[e_{t-j}]&=x_{t-j}-\hat{x}_{t-j-1}(1),\ j=0,1,2,\dots\\ \underset{t}{E}[e_{t+j}]&=0,\ j=1,2,\dots\end{aligned} \tag{12}$$

Although the ARMA models described above have been used to avoid complexity, Equation 9 should be modified based on Equation 2 through 7 since these are actually ARIMA models.

LOAD PROFILE PREDICTION USING TCBM AND ARIMA HYBRID-MODELING

The technique introduced in this paper is to correct the load level that is predicted by the ARIMA model for periodical pattern prediction using the load level that is predicted by TCBM so as to cope with the multiple factors of load level fluctuation.

Concretely, with the aim of improving the efficiency of operation of thermal energy storage systems, a correction method which is intended mainly to improve the accuracy of prediction of cumulative thermal load on a daily basis was adopted. In this method, thermal load is first predicted on an hourly basis by TCBM and ARIMA separately, then the day's cumulative thermal load predicted by ARIMA is corrected in such a way that it coincides with the day's cumulative thermal load predicted by TCBM (Fig. 2).

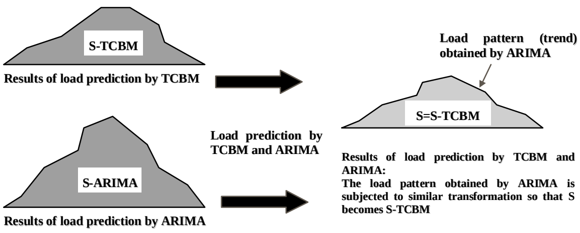

Figure 2 Load prediction using TCBM and ARIMA hybrid-modeling

MODEL VARIABLES

The types of information that can be used in creating load prediction models and in predicting load are load/environment information and unknown future information. It is considered that the more information is available, the better it would be. In practice, however, it is undesirable to use information that can hardly be obtained. Besides, increasing the amount of information used does not always improve the accuracy of prediction. Thus, in load prediction, it is very important to select suitable information. The information (model variables) that can logically be used for load prediction is shown below.

As load/environment information, measurement measured data that can be obtained by sensors, etc. and recorded by a central monitor system are considered appropriate.

1) Weather condition: open air temperature, open-air humidity, solar radiation, etc.
2) Indoor environment: room temperature, room humidity, set point of temperature, etc. of each system
3) Operating condition: air conditioning system operational information (on-off flag)
4) Thermal load: thermal load of object of prediction, primary energy consumption, secondary thermal load, etc.

The above data are all measured data. In load prediction, therefore, only data obtained up until the time of prediction can be used. It should be noted, however, that with respect to operating condition, the data that has been input in the operation schedule may be used, hence the scheduled operating condition at the time of prediction can be used in the prediction.

As unknown future information, fuzzy information, such as general weather conditions forecast by local meteorological observatories, is considered appropriate.

1) Forecast maximum/minimum temperatures
2) Forecast minimum humidity
3) Forecast general weather conditions

These data supplement the above measured data that are unavailable at the time of prediction. The environment that permits these fuzzy data to be automatically input for prediction is being established.

APPLICATION EXAMPLE

As an example of application of this prediction technique, a benchmark test that was carried out for thermal load prediction is described below.

1) Outline of benchmark test for thermal load prediction

The benchmark test for thermal load prediction was carried out in the following two phases.

- Trial-1. With the aim of evaluating the performance of the prediction engine, not only measured data but also future meteorological data, etc. which were actually unknown at the time of prediction were used as known variables. The object of thermal load prediction was the research center of Tokyo Electric Power Company (seven-story of research rooms having a total floor area of 28,500 m^2).

- Trial-2. From the viewpoint of evaluating the practicality of the prediction engine, only the measured data that were available at the time of prediction were used. The object of thermal load prediction was the TONETS corporation Shinkawa Building (nine-story commercial building having a total floor area of 5,400 m^2; the first floor is a parking space).

The benchmark test reported in this paper concerns Trial-2. First, a model was created and made to learn measured data of the building, such as the cooling load, open-air conditions, solar radiation, and air conditioning system operating condition, obtained during the period from June 1 to October 31 of 1995. Then, using only the measured data, weather forecast, and next day's air conditioner operation schedule at 22:00 of each of the days from June 1 to October 31 of 1996, the time-series load trend till 22:00 of next day was forecast.

2) Conditions of information that can be used in load prediction

Trial-2 was a benchmark test based on data that actually contain a number of problems. Therefore, from the practical viewpoint, the following conditions were set on the information used for load prediction.

i) Unknown future information: Generally speaking, meteorological data for next day are rather fuzzy (e.g., “fine”, “cloudy, later rainy”, etc. for next day's weather and only maximum and minimum temperatures for next day's temperature). Concerning the hours of operation and the use of office rooms too, it is only the present schedule that is certain. In Trial-2, therefore, only future information that is actually available was used.

ii) Limitations on load/environment information: In predicting load, it is desirable that there is as much amount of load/environment information as possible. However, in an actual commercial building, it is almost impossible to install sensors everywhere. Therefore, items of data that could be measured at the building automation system were selected and only data from a small number of sources that were considered usable in the thermal load prediction were used.

The data items that were actually supplied as ones usable in creating a load prediction model and updating it are shown in Table 4.

The data items that can be used for load prediction are those shown in Table 4, FL (operating condition) at the time of prediction, and weather forecast data shown in Table 5.

Table 4. Data items for learning and prediction

	Data item	Unit	State of process	Number of figures
W: Weather data	OT: Outdoor air temperature	deg C	Instantaneous	##.#
	OH: Outdoor air humidity	%	Instantaneous	##.#
	SR: Overall solar radiation (horizontal surface)	W/m^2	Hourly average	###
R: Room environment	RT: Room temperature for each system	deg C	Instantaneous	##.#
	RH: Room humidity for each system	%	Instantaneous	##.#
F: Operating condition	FL: Air conditioner/fan coil unit operation flag for each system		0/1	#
L: Hourly thermal load	TL: Thermal load at secondary side for all systems	kcal/h	Hourly average	#####

Table 5. Weather forecast data items

Data item	Symbol	Content	Unit	Number of figures
F: Weather forecast data	Otmax	Maximum temperature	deg C	##
	OTmin	Minimum temperature	deg C	###
	OHmin	Minimum humidity	%	###
	Wcode	Weather symbol	character	#####*

* See Table 6.

Table 6 Weather code

Code	Content	Code	Content
F	Fine	=	Occasionally
C	Cloudy	/	for a time
R	Rainy	M	In the morning
-	Later	E	In the evening

Examples:
F-C means, “Fine, later cloudy”.

C-C=R means, “Cloudy, later occasionally rainy” but symbolized as “Cloudy, later cloudy and occasionally rainy”.

FMR means, “Fine but rainy in the morning”.

3) Selection of load factors

ARIMA uses time-series load information of the past as the only input variables for the model. By contrast, TCBM can handle not only the load information but also some other information as the input variables. Therefore, suitable data selected from the information described in 2) were used as input variables (load factors) for the model.

To select suitable data, basic statistic calculation, correlation analysis, stepwise analysis (analysis of linear factor combinations), and cluster analysis for selecting nonlinear factors were carried out. With TCBM, it is also possible to determine in the process of model identification the validity of each of the selected input variables from unique evaluation indexes (required continuity rate, required output distribution rate). Using those indexes, the validity of the selected data items was analyzed. In the analysis, a data mining program was used. The analysis results could be obtained in about an hour.

Since the analysis results obtained were simply based on given data, the input variables to be used were ultimately decided taking into consideration the mechanisms by which a thermal load occurs. The selected factors are as follows.

- Input variables used for TCBM (total of 10 input variables)

i) Outdoor air temperature (For learning, data at the time of learning was used; for prediction, data 24 hours before the time of prediction was used.)
ii) Outdoor air humidity (For learning, data at the time of learning was used; for prediction, data 24 hours before the time of prediction was used.)
iii) AHU on each floor on-off (total of 7 variables because there are seven floors)
iv) Calendar information (weekdays, Saturday, Sunday)

- Input variables used for ARIMA (total of 2 input variables)

i) Measured load (load measured before the time of prediction)
ii) Logical sum of air conditioner and fan coil unit on and off (for post-processing)

4) Prediction results

Figure 3 compares the predicted values with the measured values on a time-serial basis for the period from August 3 to October 10. The prediction error on an hourly basis for the entire period of prediction is shown in Table 7, and the prediction error on a daily basis is shown in Table 8.

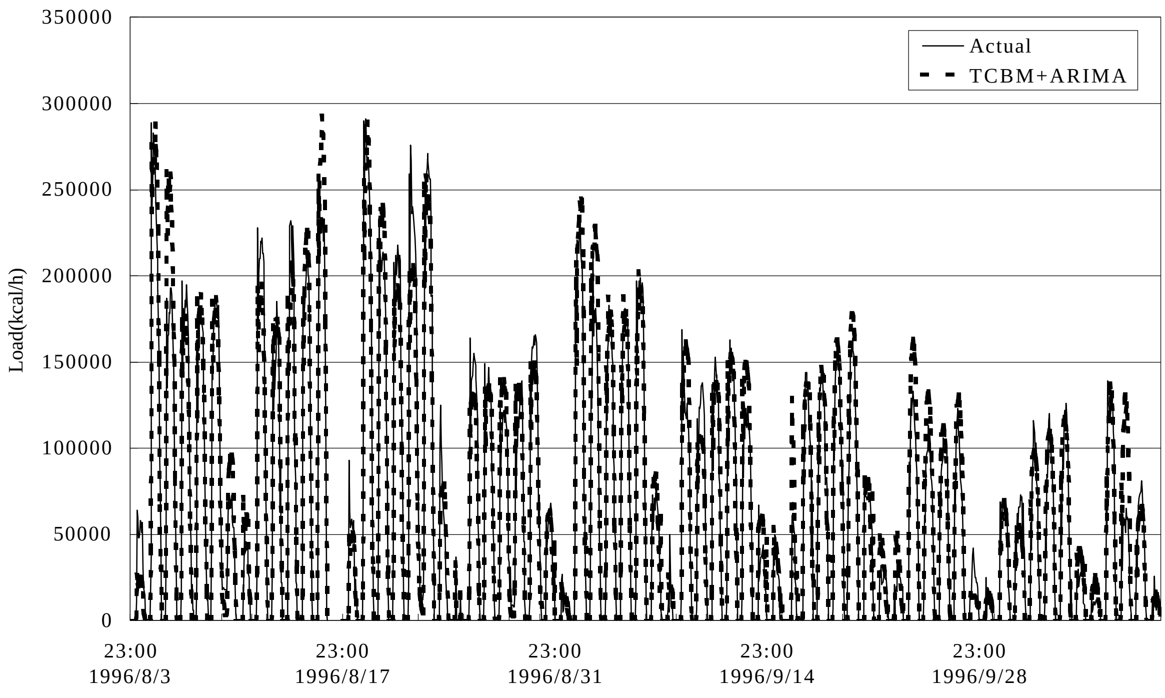

Figure 3 Prediction results

Table 7. Evaluation of prediction error on hourly basis

Prediction technique	Average value of absolute error	Max. error on + side	Max. error on – side	Sum of squares of error
ARIMA	1.577E+04	2.603E+05	-1.587E+05	2.504E+12
TCBM	1.341E+04	1.398E+05	-1.290E+05	1.845E+12
TCBM+ARIMA	1.208E+04	1.633E+05	-1.217E+05	1.454E+12
Hourly average of measured values	6.583E+04			

Table 8. Evaluation of prediction error on daily basis

Prediction technique	Average value of absolute error	Max. error on + side	Max. error on – side	Sum of squares of error
ARIMA	3.129E+05	1.139E+06	-1.379E+06	2.646E+13
TCBM	1.998E+05	8.048E+05	-7.359E+05	1.128E+13
TCBM+ARIMA	1.998E+05	8.048E+05	-7.359E+05	1.128E+13
Daily average of measured values	1.580E+06			

The prediction results shown in Tables 7 and 8 are based on the evaluation standards for the load prediction trial. The reason why the maximum errors, as well as the absolute error, are evaluated is that when the prediction technique is applied to an actual system, if a single predicted value significantly deviates from the measured value, it can have unfavorable effect on the heat source device. From the viewpoint of operational support and automatic operation, in practical application of the prediction technique, it is considered effective to make an accurate analysis of error factors for each input variable in order to improve the efficiency of operation of thermal energy storage systems.

The reason why the above statistical standards of evaluation are employed in the benchmark test is that with conventional statistical models, it is difficult to evaluate error for each input variable. With TCBM, by contrast, it is possible to evaluate not only the statistical error (model accuracy) mentioned above but also the error for each input variable.

DISCUSSIONS

In addition to the evaluation of prediction results described above, the results of another load prediction conducted using the current day's temperature and humidity data under the same conditions as in Trial-1 were evaluated. As a result, it was found that a significant prediction error can occur when the temperature and humidity at the time of prediction differ widely from those measured at the same time of the previous day. The results of still another load prediction conducted using weather forecast data were also evaluated. As a result, it was found that the prediction error was larger than when the temperature and humidity measured on the previous day were used.

From the above facts, it was found that the accuracy of load prediction depends much on how accurately the temperature and humidity at the time of the load prediction can be predicted.

CONCLUSION

In this paper, the concept of thermal load of air conditioning was confirmed first. Then, based on this concept, TCBM and ARIMA hybrid-modeling was taken up as one of load prediction techniques, and the prediction procedure and characteristics of this technique were discussed. Finally, the validity of this prediction technque was proved by applying it in a benchmark test for thermal load prediction.

FUTURE TASKS

At present, the authors et al. are striving to establish an environment which makes it possible to prove that the load prediction technique described in this paper can easily be applied to other objects using various types of measurement measured data.

In the future, they intend to study methods for accurately predicting weather conditions, such as the open-air temperature and humidity, at the time of prediction, and improve the accuracy of load prediction, thereby contributing to the improvement of accuracy and efficiency of operation of heat sources and thermal energy storage systems.

REFERENCES

Akaike,H., “A new look at the statistical model identification”, IEEE Transaction AC-19, 1974.

Box,C.P., and J.M.Jenkins, “Time seies analysis”, New York: Holden-Day, 1965.

Kimbara,A., S.Kurosu, R.Endo, K.Kamimura, T.Matsuba and A.Yamada, "On-line prediction for load profile of an air-conditioning system", ASHRAE Transaction, San Diego meeting, 1995.

Kurosu, S., A. Kimbara, K. Kamimura and N. Suzuki, "Prediction for energy load of air-conditioning system", IV-ICCCBE 1991 Conference, 1991.

Tsutsui, H., "Topological case based modeling", Building Optimization and Fault Diagnosis Source Book, IEA ANNEX25, 1996.

Tsutsui, H. and K. Kamimura, "Chiller Condition Monitoring Using Topological Case-Based Modeling", AT-96-6-4, ASHRAE Atlanta meeting, 1996.