ESL-PA-94/07-02 (5)

OR-94-17-5

A PIECEWISE-LINEAR REGRESSION ON THE ASHRAE TIME-SERIES DATA

Makoto Iijima **Kenji Takagi**

Ryo Takeuchi **Takashi Matsumoto, Dr.Eng.**

ABSTRACT

A piecewise-linear regression was performed on the 1993 ASHRAE time-series data. For data set A, several preprocessing steps were done including normalization, outlier removal, and trend removal. Then training was done via a straightforward steep descent of the coefficients of piecewise-linear maps derived from representation theorems of piecewise-linear maps.

For data set B, no preprocessing was done except for a straightforward normalization. Again, a simple steep descent was used to adjust the coefficients of the maps. Despite its simplicity, the algorithm performs quite reasonably. This indicates that the classic linear algorithms have obvious limitations in solving real-world problems, so one should use algorithms incorporating nonlinearities of the problems.

INTRODUCTION

Time-series prediction problems from real-world data are interesting as well as difficult. In most of the difficult time-series prediction problems, typically the 1993 ASHRAE Building Energy Prediction Shootout problems, one naturally expects that the classic linear approaches will have limited success for various reasons. Among the many possible nonlinear methods available, this paper performs a piecewise-linear regression based on representative theorems for continuous piecewise-linear maps developed for a discipline that is entirely different from time-series prediction (Matsumoto et al. 1993). Training is done via a straightforward steep descent in the parameter space of piecewise-linear maps. Other training is possible, such as that described in Breiman (1991). Our intention in this paper is *not* to claim that this is the method for difficult time-series prediction problems. Rather, our purpose is to test the potential capabilities of piecewise-linear regression for the class of problems given in the ASHRAE competition. Since this paper does not perform any deep statistical analysis, the term "regression" should be understood in its most basic sense.

DESCRIPTION OF THE PROBLEMS

Our description of the problem will be brief since details will be given elsewhere by the organizers (Kreider and Haberl 1994).

Problem A

Training Set The hourly measurements of dependent variables included electric energy consumption (WBE), hot water consumption (WBHW), and chilled-water consumption (WBCW) together with the independent variables of temperature, humidity, solar flux, and wind velocity. The measurements were made from September 1 to December 31, 1989.

Test Set Weather data were gathered and measurements were made from January 1 to February 23, 1990.

The nature and the location of the building from which the data were taken were not disclosed to the participants. The contestants were asked to predict hourly values of whole-building electric (WBE), whole-building hot water (WBHW), and whole-building chilled water (WBCW) from January 1 to February 23, 1990.

Problem B

Training Set The dependent variable was true beam insolation and the independent variables were horizontal solar flux, southeast solar flux, south solar flux, and southwest solar flux. Measurements were made from August 1988 to May 1989.

Test Set The test set consisted of independent variables at random times. Given four independent variables from the fixed devices at random times, the contestants were asked to predict the true beam insolation (dependent variable).

THE ALGORITHM

Representation Theorems for Piecewise-Linear Maps

Several definitions are necessary to state the representation theorems for piecewise-linear maps. For $\alpha_i \in \mathbf{R}^n$, $\beta_i \in \mathbf{R}$, $i = 1, \ldots, k$, let

Makoto Iijima, **Kenji Takagi**, and **Ryo Takeuchi** are students, and **Takashi Matsumoto** is a professor in the Department of Electrical Engineering at Waseda University, Tokyo, Japan.

THIS PREPRINT IS FOR DISCUSSION PURPOSES ONLY, FOR INCLUSION IN ASHRAE TRANSACTIONS 1994, V. 100, Pt. 2. Not to be reprinted in whole or in part without written permission of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 1791 Tullie Circle, NE, Atlanta, GA 30329. *Opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of ASHRAE.* Written questions and comments regarding this paper should be received at ASHRAE no later than July 6, 1994.

$$U(\alpha_i, \beta_i) := \{x \in \mathbf{R}^n \mid \langle \alpha_i, x \rangle = \beta_i\},$$

$$B := \bigcup_{i=1}^{k} U(\alpha_i, \beta_i)$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbf{R}^n$. $U(\alpha_i, \beta_i)$ is called a *boundary*, whereas B is called the *boundary set*. Let ω: $\mathbf{R}^n \to \{0,1\}^k$ be defined by

$$\omega(x) := (\mathrm{sgn}(\langle \alpha_1, x \rangle - \beta_1), \ldots, \mathrm{sgn}(\langle \alpha_k, x \rangle - \beta_k))$$

where

$$\mathrm{sgn}(t) = \begin{cases} 0 \ (t \le 0) \\ 1 \ (t > 0) \end{cases},$$

and let

$$\Omega = \bigcup_{x \in \mathbf{R}^n} \omega(x) \subset \{0, 1\}^k.$$

Then,

$$R_\omega := \{x \in \mathbf{R}^n \mid \omega(x) = \omega\}, \ \omega \in \Omega$$

gives rise to a nonoverlapping partition of $\mathbf{R}^n$:

$$\mathbf{R}^n = \bigcup_{\omega \in \Omega} R_\omega, \ R_\omega \cap R_{\omega'} = \varnothing \text{ if } \omega \ne \omega'.$$

Definition 1 An f: $\mathbf{R}^n \to \mathbf{R}^m$ is piecewise-linear if there are $\{(\alpha_i, \beta_i)\}_{i=1}^{k}$ such that

1. f is differentiable except on B and
2. in each interior of R_ω, the Jacobian $Df(x) := A_\omega$ is constant.

Figure 1 shows typical piecewise-linear maps.

Definition 2 Given $\{(\alpha_i, \beta_i)\}_{i=1}^{k}$, the boundary set B is called *transverse* if for any subset $\{U_{i_1}, \ldots, U_{i_s}\}$ of $\{U_1, \ldots, U_k\}$,

$$U_{i_1} \cap \ldots \cap U_{i_s} = \varnothing,$$

or

$$U_{i_1} \cap \ldots \cap U_{i_s} \ne \varnothing \text{ and}$$

$\{\alpha_{i_1}, \ldots, \alpha_{i_s}\}$ are linearly independent.

The following characterizes *continuous* piecewise-linear maps with transverse boundaries. Proof is found in Matsumoto et al. (1993). See Chua and Deng (1988) also.

Theorem 1 A piecewise-linear f: $\mathbf{R}^n \to \mathbf{R}^m$ defined by

$$f(x) = A_\omega x + q_\omega, \ x \in R_\omega, \ \omega \in \Omega$$

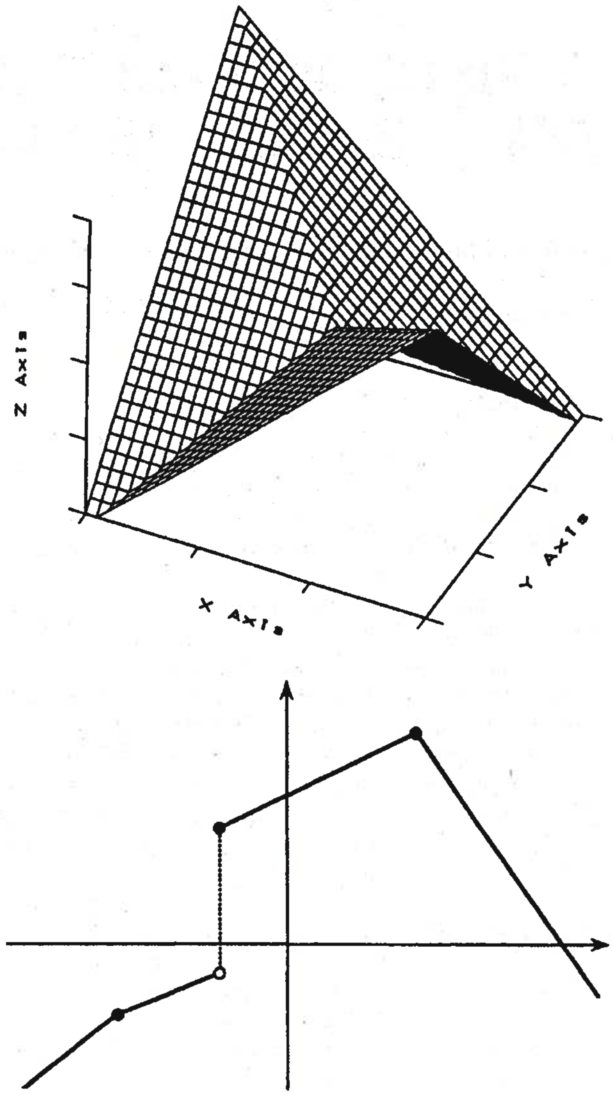

Figure 1 *Typical piecewise-linear maps. (a) A continuous piecewise-linear map: $\mathbf{R}^2 \to \mathbf{R}$. (b) A discontinuous piecewise-linear map: $\mathbf{R} \to \mathbf{R}$*

where transverse B is continuous *if and only if* f admits the following representation:

$$f(x) = A_0 x + q_0 + \sum_{i=1}^{k} p_i \sigma(\langle \alpha_i, x \rangle - \beta_i) \tag{1}$$

where

$0 = (0, 0, \ldots, 0) \in \Omega$,
A_0 = an $m \times n$ matrix,
$q_0, p_i \in \mathbf{R}^m$, and

$$\sigma(t) := \begin{cases} t \ (t \le 0) \\ 0 \ (t > 0) \end{cases}. \tag{2}$$

Remarks

1. Strictly speaking, the map f defined by Equation 1 is piecewise-affine instead of piecewise-linear. We are simply following a custom.
2. Observe the *if-and-only-if* nature of the statement. Namely, *any* continuous piecewise-linear map: $\mathbf{R}^n \to \mathbf{R}^m$ with transverse B *must be* of the form in Equation 1.
3. Equation 1 says that a continuous piecewise-linear map can be represented as a linear combination of the "basis functions" $\sigma(\langle\alpha_i,x\rangle - \beta_i)$ together with the affine term $A_0x + q_0$.
4. The functional form of $\sigma(\cdot)$ defined by Equation 2 (see Figure 2) sometimes appears in the neural network literature; however, theorem 1 has not been motivated by neural networks.

This theorem presupposes transversality of B. If transversality is violated (see Figure 3), more involved arguments are necessary.

The following is a complete representation theorem for an arbitrary piecewise-linear map: $\mathbf{R}^n \to \mathbf{R}$ for a general B (not necessarily transverse). Proof is given in Matsumoto et al. (1993).

Theorem 2 An arbitrary continuous piecewise-linear map f: $\mathbf{R}^n \to \mathbf{R}$ can be represented as, at most, n *compositions* of the map of the form of Equation 1.

The number of compositions depends on the degeneracy of B, i.e., the degree of linear dependency of the hyperplanes $U(\alpha_i, \beta_i)$ at their intersection points. For a vector-valued piecewise-linear map: $\mathbf{R}^n \to \mathbf{R}^m$, one can apply theorem 2 to each component.

These two results give exact representations for continuous piecewise-linear maps, while the next result (due to Breiman [1991]) guarantees the approximation capability of *smooth* maps by piecewise-linear maps. This is a result from statistics, not from the discipline from which theorems 1 and 2 evolve.

Theorem 3 Let F: $\mathbf{R}^n \to \mathbf{R}$ be smooth such that

1. the support of F is bounded and
2. $\int \|\omega\|^2|\tilde{F}(\omega)|d\omega < \infty$, where $\tilde{F}(\omega)$ is the Fourier transform of F.

Then, F can be approximated by a continuous piecewise-linear map of the form in Equation 1 with arbitrary accuracy where k is *independent* of n.

These theorems provide us with a theoretical justification for utilizing the piecewise-linear regression to be described next.

Piecewise-Linear Regression

Let $x(t) \in \mathbf{R}^m$ and $u(t) \in \mathbf{R}^n$ be the dependent and independent variables at time t, where $x(t)$ is to be predicted given $u(t)$. Generally, one expects that $(x(t),u(t))$ gives rise to a dynamic system

$$x(t+1) = f(x(t), x(t-1), \ldots, x(t-\tau_x), u(t), u(t-1), \ldots, u(t-\tau_u)), \tag{3}$$

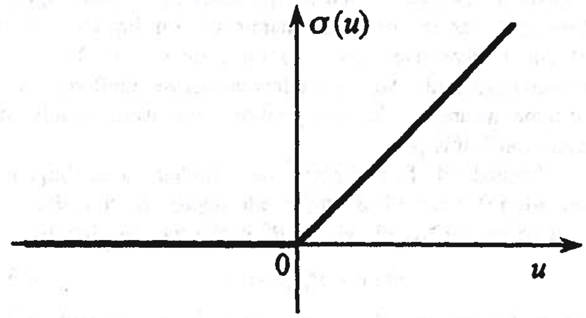

Figure 2 Function $\sigma(\cdot)$ in the representation theorem of piecewise-linear maps.

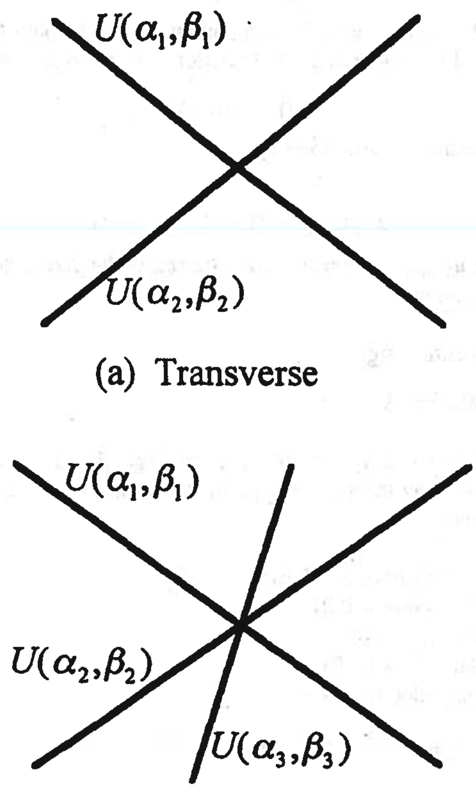

Figure 3 Boundary set for piecewise-linear maps. (a) Transverse. (b) Nontransverse.

so that one expects to find an appropriate $f(\cdot)$. Our original intention was to adjust the parameters of Equation 1 to obtain a piecewise-linear $f(\cdot)$ for Equation 3. Due to a severe time limitation (only a few weeks were left when we became aware of the competition), we were unable to accomplish this plan.

Instead of formulating the problem according to Equation 3, we took a very crude approach. Namely, we first computed "trend" $y(t) \in \mathbf{R}^n$ in the training data by

$$y(t) = F_t(x(\cdot), u(\cdot)) \tag{4}$$

where $F_t(\cdot)$ is an (ad hoc) operation on $x(\cdot) = (x(0), x(1), \ldots)$ and $u(\cdot) = (u(0), u(1), \ldots)$. (The specific form of $F_t(\cdot)$ will be given in the next section.) Then, the "residual,"

$$\hat{x}(t) := x(t) - y(t), \tag{5}$$

would contain the remaining nontrivial information. We adjusted the parameters of Equation 1 so as to realize

$$\hat{x}(t) = f(u(t)). \tag{6}$$

The prediction was done by

$$x(t)_{\text{pred}} = f(u(t)_{\text{test}}) + y(t) \tag{7}$$

where $u(t)_{\text{test}}$ is from the test data set, while $y(t)$ is defined by Equation 4.

Preprocessing

Data Set A

Normalization In order to arrange the data within approximately the same range, the following normalization was done:

hour (military numbering) × 0.001
temperature × 0.01
humidity × 100
solar flux × 0.001
wind velocity × 0.1

Trends

(i) Electric Energy

Apparently, there is a 24-hour periodic component that we will remove. There is a distinct difference between energy consumption on workdays and that on holidays. Let $E_W{}^d(t)$ and $E_H{}^d(t)$ be the WBE at time t of a workday and holiday d, respectively, and let

$$e_W(t) := \sum_{d \in W} E_W^d(t)/n_W$$

where d runs over workdays W between September 1 and December 31 and n_W is the number of workdays in this period. Similarly, let

$$e_H(t) := \sum_{d \in H} E_H^d(t)/n_H$$

where d, this time, runs over holidays H, while n_H stands for the number of holidays during this period. These are the trend functions defined in Equation 4. Instead of subtracting, one can divide the raw data by the trends:

$$\hat{E}_W^d(t) = \frac{E_W^d(t)}{e_W(t)}, \quad \hat{E}_H^d(t) = \frac{E_H^d(t)}{e_H(t)}.$$

Finally, we perform the reparametrization

$$x_{1W}(\tau) \leftarrow \hat{E}_W^d(t), \ d \in W \tag{8}$$

where τ runs linearly over the workdays W. Similarly,

$$x_{1H}(\tau) \leftarrow \hat{E}_H^d(t), \ d \in H \tag{9}$$

where τ runs linearly over the holidays H.

(ii) Chilled and hot water

The major trend is naturally the dependency on temperature. Finding a right $F_t(\cdot)$ for Equation 4 was a rather difficult task. The trend computed was the *two-week moving average:*

$$C^{\text{av}}(\tau) := \frac{\sum_{s=\tau-24\times7}^{\tau+24\times7} C(s)}{24(\text{hrs}) \times 14(\text{days}) + 1} \tag{10}$$

where $C(s)$ is the chilled-water consumption parameterized linearly from September 1 to December 31. We did not distinguish between chilled-water consumption on workdays and that on holidays. The validity of this assumption is, of course, to be discussed. Similarly, let

$$H^{\text{av}}(\tau) := \frac{\sum_{s=\tau-24\times7}^{\tau+24\times7} H(s)}{24(\text{hrs}) \times 14(\text{days}) + 1} \tag{11}$$

where $H(s)$ is the hot water consumption. Finally, let

$$T^{\text{av}}(\tau) := \frac{\sum_{s=\tau-24\times7}^{\tau+24\times7} T(s)}{24(\text{hrs}) \times 14(\text{days}) + 1} \tag{12}$$

be the two-week moving average of temperature.

Letting[1]

$$(a_1^*, b_1^*) := \arg\max \ \text{Correlation}\,[(a_1 T^{\text{av}}(\cdot) + b_1) C^{\text{av}}(\cdot)],$$

we attempt to remove the temperature dependency of $C(\tau)$:

$$x_2(\tau) := C(\tau) - (a_1^* T^{\text{av}}(\tau) + b_1^*). \tag{13}$$

Similarly, let

$$x_3(\tau) := H(\tau) - (a_2^* T^{\text{av}}(\tau) + b_2^*) \tag{14}$$

[1]Equation $z^* = \arg\max g(z)$ means that z^* gives the maximum of $g(z)$.

with

$$(a_2^*, b_2^*) := \arg\max$$

$$\text{Correlation}[(a_2 T^{av}(\cdot) + b_2) H^{av}(\cdot)].$$

Outliers The following data were removed from the training set since they appear to be outliers:

Electric energy:	1. November 20-26
	2. December 14-31
Chilled and hot water:	3. December 22-24.

Number 1 appears to correspond to the Thanksgiving holidays, while numbers 2 and 3 appear to be attributable to the Christmas holidays.

Data Set B

Normalization All flux data are multiplied by 0.001. Neither trend nor outlier was discernible.

Training

Problem A We used theorem 1 instead of theorem 2 because

1. if B is not transverse, one can perturb (α_i, β_i) slightly in such a way that the perturbed boundary becomes transverse and
2. training will be rather involved if one uses theorem 2.

There are four variables to be predicted (see Equations 8, 9, 13, and 14): $x_{1W}(\tau)$, $x_{1H}(\tau)$, $x_2(\tau)$, and $x_3(\tau)$. Letting

$$u(\tau) := (u_1(\tau), u_2(\tau), u_3(\tau), u_4(\tau), u_5(\tau)),$$

where

$u_1(\tau)$ = temperature at τ,
$u_2(\tau)$ = humidity at τ,
$u_3(\tau)$ = solar flux at τ,
$u_4(\tau)$ = wind velocity at τ, and
$u_5(\tau)$ = hour (military numbering) at τ,

we performed the four minimization problems (recall Equation 1):

$$\min_{\{A_{0j}, q_{0j}, p_{ij}, \alpha_{ij}, \beta_{ij}\}} \frac{1}{2} \sum_{\tau \in T_j} |x(\tau) - \langle A_{0j}, u(\tau)\rangle - q_{0j} - \sum_{i=1}^{k} p_{ij} \sigma(\langle \alpha_{ij}, u(\tau)\rangle - \beta_{ij})|^2 \quad (15)$$

for $j = 1, 2, 3, 4$, where

$$x(\tau) \in \{x_{1W}(\tau), x_{1H}(\tau), x_2(\tau), x_3(\tau)\}$$

and

$$A_{0j}, \alpha_{ij} \in \mathbb{R}^5, q_{0j}, \beta_{ij} \in \mathbb{R}.$$

Note that since outliers have been removed,

$$n_{T_1} + n_{T_2} \neq n_{T_3} = n_{T_4},$$

where n_{T_i} denotes cardinality of T_i, $i = 1, \ldots, 4$. As for k in Equation 15, the number of basis functions, our original plan was to divide each T_j into two subsets $T_j = T_j^{\text{"train"}} \cup T_j^{\text{"test"}}$, perform minimization over $T_j^{\text{"train"}}$, and compare the performance with different ks using appropriate measures over $T_j^{\text{"test"}}$. Again, the time limitation did not permit us to accomplish this scheme. For $j = 1, \ldots, 4$, the number k was set to 100 and the minimization was performed over the entire $T_j, j = 1, \ldots, 4$.

The minimization was by a straightforward steepest descent. Note that the only information necessary to compute gradients is the sign of $(\langle \alpha_{ij}, u(\tau)\rangle - \beta_{ij})$. Although $\sigma(\cdot)$ is not differentiable at 0, $\langle \alpha_{ij}, u(\tau)\rangle - \beta_{ij} = 0$ does not occur generically.

Problem B The problem description indicates that there would be very little, if any, dynamics involved. Of the 2,444 (training) data sets, the first 2,000 were used for training, while the rest were used to test the ability with k = 25, 50, 75, 100. Namely, letting

$x(\tau)$: true beam insolation at τ
$u(\tau) = (u_1(\tau), u_2(\tau), u_3(\tau), u_4(\tau), \tau)$

where

$u_i(\tau)$, $i = 1, \ldots, 4$: four flux data at τ,

minimization similar to that in Equation 15 was done.

A comparison between different k was done by using

$$\frac{\sum_{\tau=2001}^{2444} (x(\tau)_{\text{train}} - \bar{x}_{\text{train}})(x(\tau)_{\text{pred}} - \bar{x}_{\text{pred}})}{\left\{ \begin{array}{l} \sqrt{\sum_{\tau=2001}^{2444} (x(\tau)_{\text{train}} - \bar{x}_{\text{train}})^2} \\ \sqrt{\sum_{\tau=2001}^{2444} (x(\tau)_{\text{pred}} - \bar{x}_{\text{pred}})^2} \end{array} \right\}} \quad (16)$$

and

$$\frac{\sum_{\tau=2001}^{2444} (x(\tau)_{\text{train}} - x(\tau)_{\text{pred}})}{\sum_{\tau=2001}^{2444} (x(\tau)_{\text{train}} - \bar{x}_{\text{train}})} \quad (17)$$

where $\bar{x}_{\text{"train"}}$ and $\bar{x}_{\text{"pred"}}$ stand for the mean of $x(\tau)_{\text{"train"}}$ and $x(\tau)_{\text{"pred"}}$, respectively. The one with $k = 100$ performed best, both in terms of Equation 16 and Equation 17.

RESULTS

Problem A

Let f_{1W}, f_{1H}, f_2, and f_3 be the piecewise-linear maps: $\mathbb{R}^5 \to \mathbb{R}$ obtained by Equation 15. Then the predicted values are computed by

$$\mathrm{WBE(workday)}_{pred} := e_W(\tau) f_{1W}(u(\tau)_{test})$$

if τ is of a workday, while

$$\mathrm{WBE(holiday)}_{pred} := e_H(\tau) f_{1H}(u(\tau)_{test})$$

if τ is of a holiday,

$$\mathrm{WBCW}_{pred} := f_2(u(\tau)_{test}) + (a_1^* T^{av}(\tau) + b_1^*)$$
$$\mathrm{WBHW}_{pred} := f_3(u(\tau)_{test}) + (a_2^* T^{av}(\tau) + b_2^*)$$

where $e_W(\tau)$, $e_H(\tau)$, etc., are the trend functions defined previously, and $u(\tau)_{test}$ is the weather data together with τ itself (see previous section) where τ runs through January and February 1990.

Figures 4 through 6 show the relevant graphs. Significant errors seem to be incurred on Martin Luther King's Day and on Presidents' Day, both of which turned out to be workdays instead of holidays.

Problem B

The predicted value of the true beam insolation was

$$(\text{true beam insolation})_{pred} = 1000 f(u(\tau)_{test})$$

where f is the piecewise-linear map obtained by the training. Figure 7 shows the results. Table 1 summarizes the results for data sets A and B.

DISCUSSION

The number $k = 100$ in Equation 1 for problem A appears to be too large because the regression performed well on the training set, i.e., the data for September 1 to December 31, 1989, while its performance in the prediction phase degraded significantly. Namely, an overfitting appears to be happening. Figure 8, for instance, shows the results of WBE for September to December, where CV = 8.44% and MBE = 2.39%.

Despite the extreme simplicity of the algorithm, our results on problem B appear to be third among the competitors. This brings us to a natural question: "How would a linear regression do?" We did the least-squares fit:

$$\min_{\alpha,\beta} \frac{1}{2} \sum_{\tau} |x(\tau) - \langle \alpha, u(\tau) \rangle - \beta|^2 .$$

The results are CV = 12.74% and MBE = 1.03%, which indicate that the performance of linear regressions will be limited. This linear result, however, appears to be the tenth in the competition in terms of problem B, which suggests that choosing the right model is crucial.

CONCLUSION

A piecewise-linear regression was used to predict the time-series data provided by the ASHRAE competition. After much preprocessing (including normalization and outlier removal), trends were removed from the raw data of

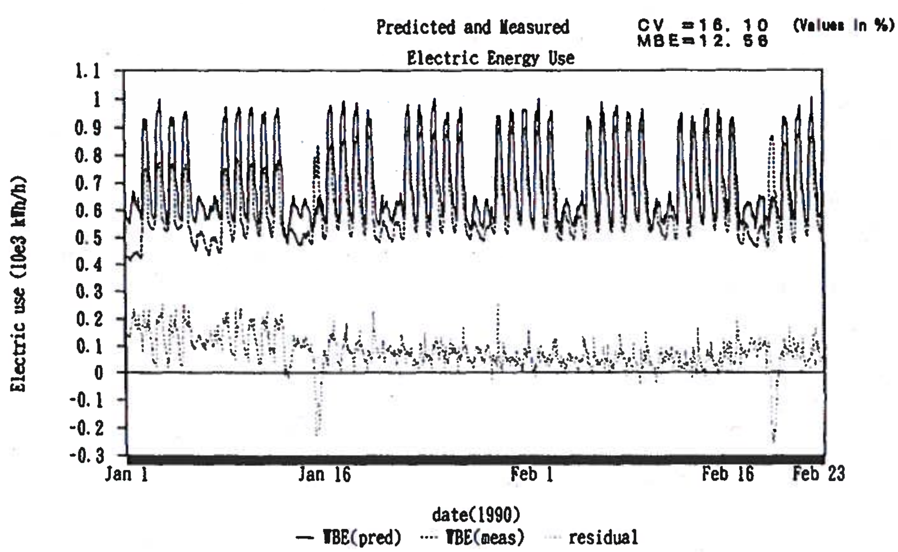

Figure 4 Whole-building electric (WBE).

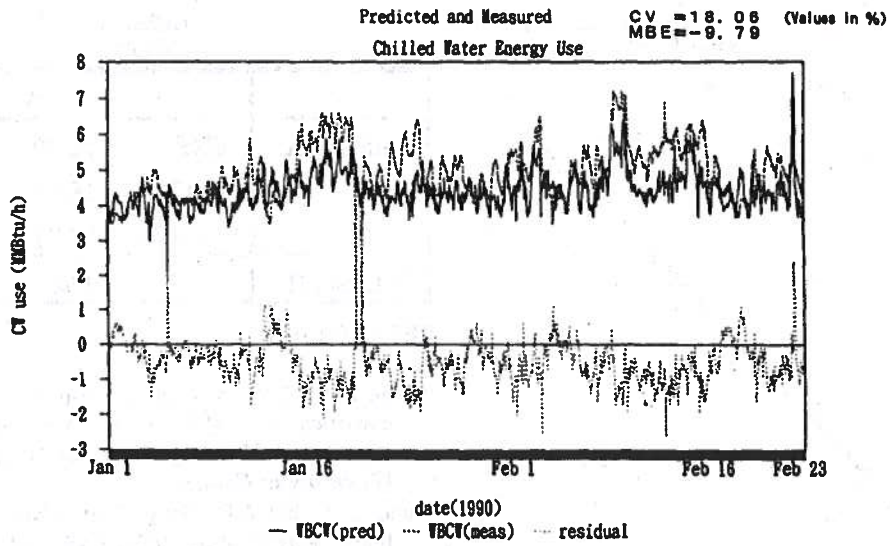

Figure 5 *Whole-building chilled water (WBCW).*

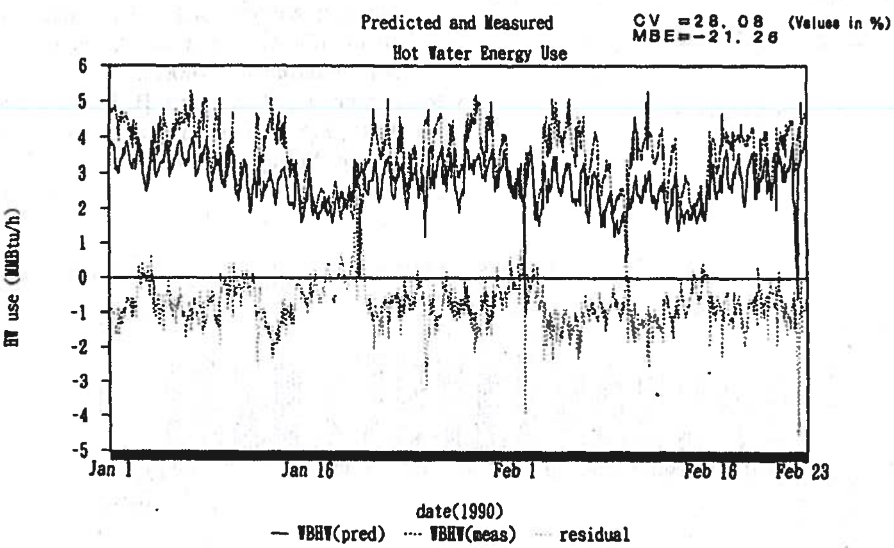

Figure 6 *Whole-building hot water (WBHW).*

data set A so that the residual would contain nontrivial information. Then a training was done to adjust the parameters of the piecewise-linear maps in such a way that the independent variables (weather data) give rise to the residual. The training was via a straightforward steepest descent. The prediction was done by producing residuals corresponding to the weather data from the test set and then by performing an inverse operation of the preprocessing.

For data set B, the only operation was normalization. Coefficients of a piecewise-linear map were adjusted to realize the functional relationship between the independent variables (the four solar flux data) and the dependent variable (true beam insolation) via a steepest descent.

While the theoretical representation theorems for piecewise-linear maps are completely rigorous, all the "data-processing" parts are ad hoc. Systematic approaches are necessary.

Formulations in terms of dynamic systems (Equation 3) will be a rather interesting direction to pursue.

ACKNOWLEDGMENT

Discussions with Y. Togawa of Science University of Tokyo were very helpful. Reviewers' comments were also helpful.

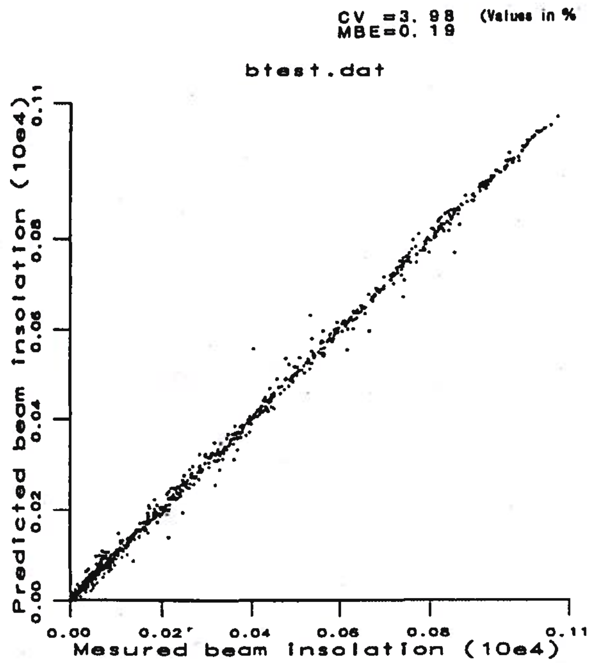

Figure 7 Results for problem B.

TABLE 1
Results

		CV %	MBE %
Data Set A	WBE	16.10	12.56
	WBCW	18.06	-9.79
	WBHW	28.08	-21.26
Data Set B		3.98	0.20

REFERENCES

Breiman, L. 1991. Hinging hyperplanes for regression, classification, and function approximation. Technical Report No. 324, Department of Statistics. Berkeley: University of California.

Chua, L.O., and A.C. Deng. 1988. Canonical piecewise-linear representation. *IEEE Trans CAS*, vol. CAS-35, pp. 101-111.

Kreider, J.F., and J.S. Haberl. 1994. Predicting hourly building energy usage: "The great energy predictor shootout"—Overview and discussion of results. *ASHRAE Transactions* 100(2).

Matsumoto, T., M. Komuro, H. Kokubu, and R. Tokunaga. 1993. *Bifurcations: Sights, sounds, and mathematics.* Tokyo: Springer-Verlag.

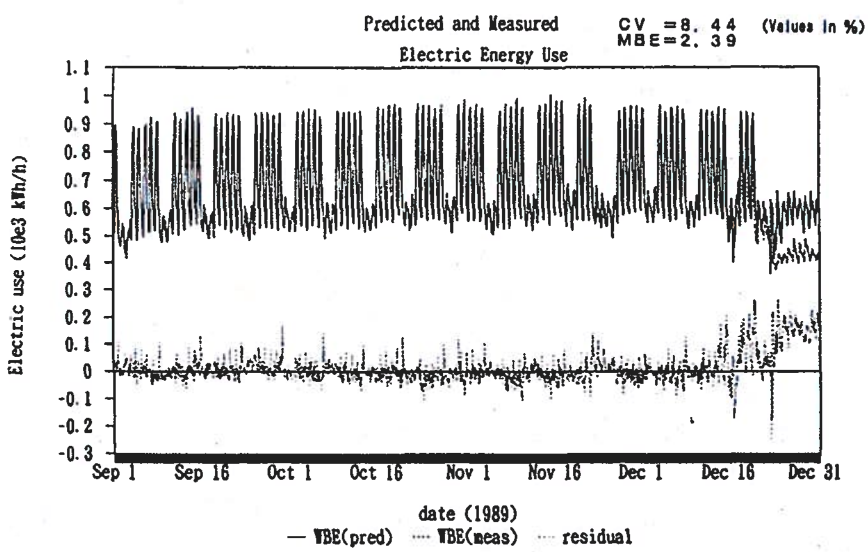

Figure 8 Performance from September 1 to December 31, 1989.