ESL-PA-94/06-01
ASHRAE Transactions, Vol. 100, Pt. 2

PREDICTING HOURLY BUILDING ENERGY USE: The Great Energy Predictor Shootout—Overview and Discussion of Results

Jan F. Kreider, Ph.D., P.E.
Member ASHRAE

Jeff S. Haberl, Ph.D., P.E.
Member ASHRAE

ABSTRACT

Analysis of measured data from buildings has become increasingly important during the past half-decade for reasons ranging from the needs of diagnostic expert systems to predicting the efficacy of energy conservation measures. In order to evaluate many of the analytical methods in use today and to assess new methods not widely applied to building data studies in the past, an open competition was held in the summer of 1993. The objective was to identify the most accurate method for making hourly energy use predictions based on limited amounts of measured data. Two data sets that typified two different needs of building data analysts—future building energy use prediction and insolation data prediction—were prepared. Responding to electronic and conventional publicity, more than 150 entrants requested data sets for which they were to make specific challenging analyses and predictions using their empirical tool of choice.

The results summarized in this paper showed that connectionist methods excelled in the analytical tasks when used either by experts or novices. The six identified winners of the competition used different methods, all within the broad definition of connectionist approaches. Quantitative results are presented in this overview paper, which is supported by six separate papers giving details of the winners' methods.

INTRODUCTION

The prediction of energy used by heating, ventilating, and air-conditioning (HVAC) systems is important for purposes of HVAC diagnostics, system control, parameter and system identification, optimization, and energy management. A wide range of new techniques is now being applied to the analytical problems involved with predicting the future behavior of HVAC systems and deducing properties of these systems. Similar problems arise in most observational disciplines, including physics, biology, and economics. New tools, such as genetic algorithms, simulated annealing, the use of connectionist models for forecasting, and tree-based classifiers or the extraction of parameters of nonlinear systems with time-delay embedding, promise to provide results that cannot be obtained with more traditional techniques. Unfortunately, the realization and evaluation of this promise has been hampered by the difficulty of making rigorous comparisons between competing techniques, particularly ones that come from different disciplines.

To facilitate such comparisons and to foster contact among the relevant disciplines, ASHRAE's TC 4.7 and TC 1.5 organized a building data analysis and prediction competition in the form of an ASHRAE seminar held in Denver in June 1993. The goal of the competition was forecasting or prediction using empirical models (system simulation code validation was not the subject of this seminar). Two carefully chosen sets of energy and environmental data from real buildings were made available to more than 150 contestants. Each contestant was required to prepare quantitative analyses of these data (including predictions) and submit them in a specific format to the seminar co-chairs prior to the Denver ASHRAE seminar. The first author evaluated all submittals using the same software. The six authors with the most accurate results each made a presentation at this seminar.

This competition was organized to help clarify the conflicting claims among many researchers who use and analyze building energy data and to foster contact among these persons and others not previously involved in building energy data analysis. The intent was not only to declare winners but also to establish a format in which rigorous and impartial evaluations of techniques could be made. Because there are natural measures of performance accuracy, a rank ordering was given. In all cases, the goal was to collect and analyze quantitative results in order to understand similarities and differences among the approaches.

THE DATA SETS

Two distinct data sets were provided to all contestants—two sets of independent variables along with the

Jan F. Kreider is a professor and director of the Joint Center for Energy Management in the Department of Civil, Environmental, and Architectural Engineering at the University of Colorado, Boulder. **Jeff S. Haberl** is an assistant professor at the Energy Systems Laboratory in the Mechanical Engineering Department at Texas A&M University, College Station.

corresponding values of dependent variables (e.g., energy use or solar radiation). The accuracy of the predictions of the dependent variables from values of the independent variables from this data set was presented as part of each competitor's submittal. This is the classic method of testing regressions or other curve fits to data.

However, a more rigorous test was also required. Some of the values of the dependent variables were withheld from each of the two data sets (this is explained in detail later) and were known only to the seminar co-chairs. The data sets from which the independent variables were withheld are hereinafter called the "testing sets," whereas the data that include a full set of both the independent and the dependent variable values are called the "training sets."[1]

The values of the independent variables in the testing set were used by each participant to make the best predictions of the corresponding dependent variables using models that were developed from the training data set. This was the essence of the competition. The organizers compared the predictions made by each contestant with the true (data) values of the dependent variables, which were known only to the organizers.

Two data sets were chosen to address two different sorts of building-related data analysis problems. In this section, the general features of the data sets are described. The sets were identified as "A" and "B," with no further information given to the contestants other than that quoted in the next two paragraphs:

A Data Set (approximately 3,000 points)

> "This is a time record of hourly chilled water, hot water, and whole-building electricity use for a four-month period in an institutional building. Weather data and a time stamp are also included. The hourly values of use of these three energy forms are to be predicted for the two following months. The testing set consists of the two months following the four-month period."

B Data Set (approximately 2,400 points)

> "These data consist of solar radiation measurements made by four fixed devices to be used to predict the time-varying hourly beam radiation during a six-month period. This four-pyranometer device is used in an adaptive controller to predict building cooling loads. A random sample of data from the full data set has been reserved as the training set of 1,500 points. The value of beam radiation is to be predicted from data from four fixed sensors for the testing set of 900 additional points."

[1]Although this nomenclature is common in some numerical approaches and not in others, it provides an understandable nomenclature for the competition.

Data recorded at a university engineering center (EC) were provided to the contestants, but physical details, the type of building, and its location were not divulged. The EC is a large, multipurpose building that contains classrooms, laboratories, faculty-staff offices, and a large central computer facility. It is located on a university campus in central Texas (30° 40'N, 96° 2'W) about two hours northwest of Houston. The EC (and more than 250 other buildings on the central campus) receives steam, hot water, chilled water, electricity, and communication services from a centralized utility distribution system.

The four-story, 324,400-ft^2 (30,100-m^2) facility was built in the early 1970s. The building measures 339 ft (the long axis) by 221 ft and is 60 ft high (111 m by 73 m by 20 m). The long axis is oriented in a northeast to southwest direction; the building is not shaded by surrounding structures. Parking for 82 cars is provided under the facility. Figure 1 is a photograph of the EC.

One of the distinguishing features of the building is a large, centralized, three-story atrium in the southwestern half of the building that provides access to the surrounding classrooms and offices. About 2,500 ft^2 (232 m^2) of northeast-facing clerestory lighting helps to illuminate the central staircase and computing facility.

The building can be characterized as an internal-load-dominated, high-mass structure with six-inch concrete floors and insulated concrete walls. Only about 9% of the exterior envelope is glazed. This consists of about 2,500 ft^2 (232 m^2) of single-pane clerestory lighting and about 9,000 ft^2 (836 m^2) of single-pane windows that are set back 3 ft (1 m) between exterior concrete utility chaseways. The EC has a maximum occupancy of 2,300 during peak periods each semester. The occupancy profiles are characterized by a weekday schedule of 8 a.m. to 7 p.m. Significant evening use of the building occurs on weekdays between 7 p.m. and midnight. Weekend use is moderate.

The building's average air-change energy use is strongly driven by the 10% to 20% fresh air that is supplied to the air-handling units (AHUs). Internal lighting loads (2 W/ft^2) and equipment loads (2.4 W/ft^2) peak during the weekdays in the early afternoon. Considerable electricity is consumed in the evenings by the central computing facility. Prior to the March 1991 variable-air-volume (VAV) system retrofit, the 12 AHUs in the EC operated continuously. Chilled water and hot water are provided by the campus's central plant. A detailed description of the HVAC system is provided by Katipamula and Claridge (1992); additional information about the building can be found in Bronson et al. (1992), Bronson (1992), Haberl et al. (1993a, 1993b), Hinchey (1991), and Katipamula and Haberl (1992).

Time-Series Data Set for the Predictor Shootout

The data that were provided to contest participants represent environmental conditions and measured energy use for the EC from September 1, 1989, to February 28, 1990,

Figure 1 *The engineering center. The four-story engineering center (EC) is a large, multipurpose building containing classrooms, laboratories, faculty-staff offices, and a large central computer facility. It is located at a university campus in central Texas (30° 40'N, 96° 2'W) about two hours northwest of Houston. It was built in the early 1970s, is oriented in a northeast to southwest direction, and is not shaded by surrounding structures.*

as shown in Figures 2 and 3. The training set consisted of whole-building electricity;[2] chilled-water, hot-water, and environmental data (i.e., ambient temperature, absolute humidity ratio, wind, and horizontal insolation); and hourly and daily time stamps for the four-month period beginning September 1, 1989, and ending December 31, 1989.

The testing data set provided to contestants began January 1, 1990, and ended on February 28, 1990. This data set consisted of only the environmental data (i.e., no consumption data were provided) and the time stamps. The contestants were then asked to predict the three energy end-uses for the test period using models that were developed from the training data set. Copies of these predictions were submitted for comparison with the actual energy use data for the two-month testing period.

Figure 2 shows the diurnal and weekly energy use patterns. Table 1 provides information about the instrumentation for data set A. Since the EC is located on a university campus, it observes the same holidays as the university, which included a Thanksgiving holiday (November 23-26, 1989) and an extended Christmas holiday (December 21, 1989, through January 1, 1990). Both periods can be clearly seen in the data. The week prior to the Christmas holiday and the two weeks following the holiday in January have lower-than-normal consumption because this is the period between the fall and spring semesters when the building is sparsely occupied. Figure 3 contains graphs of hourly heating and cooling energy consumption versus average outdoor dry-bulb temperature; this type of plot is often used by building energy analysts because temperature is one of the most important parameters influencing building HVAC energy use.

The training data set also contained a period of "abnormal" consumption that was not pointed out to the contestants. Such abnormal consumption periods are typical in large buildings; this represented an additional challenge to the modelers. The decision as to whether or not to include these periods in the model was left to the discretion of each analyst. The period of abnormal consumption is December 22-23, 1989, as indicated by the solid symbols on the scatter plots in Figure 2. The abnormal consumption can be seen as the points that lie below the large cluster of chilled-water data. These data represent conditions that occurred after an extremely cold spell was experienced on the campus; many chilled-water lines froze and ruptured during this period. This caused a large decrease in the amount of chilled water that was available for building cooling, hence the low chilled-water data outliers. The hot water use for this period unexpectedly increased even though one might have expected the opposite because less chilled water was available to the constant-volume air handlers.

The average hourly electricity use in data testing set A is roughly 100 kW less than the electricity in the corresponding training set. This is because several academic departments (i.e., computer science and aerospace engineering) moved out of their offices in the EC and into newly constructed facilities in another building. These vacant offices were then filled by faculty and staff from the other academic departments in the building. This had the effect of reducing both electricity use and chilled-water use while slightly increasing the hot water consumption to make up for the lower internal heating loads, as shown in Figure 3. This fact also was not provided to the contestants.

Solar Radiation Data (Data Set B)

Cooling loads on buildings are affected by the local insolation, most of which is beam radiation on sunny, high-load days. A device for indirectly measuring beam solar radiation has been developed by Curtiss (1992) and Curtiss et al. (1994). Essentially, four fixed pyranometers—one horizontal, one tilted and facing south, and two tilted but facing east and west of south—are used to find the beam and diffuse insolation. The advantage of this fixed device is that lower cost and less maintenance are involved compared to the more common tracking pyrheliometer or shadow band pyranometer. Data set B included measured beam insolation (from a pyrheliometer located at nearby laboratories) along

[2] The whole-building electricity data is a derived data channel that represents the electricity used by the lights and receptacles. In the case of the EC, it is the whole-building electricity use minus the sum of all motor control centers and electricity consumed by a central computing facility. For additional information, see Bronson et al. (1992).

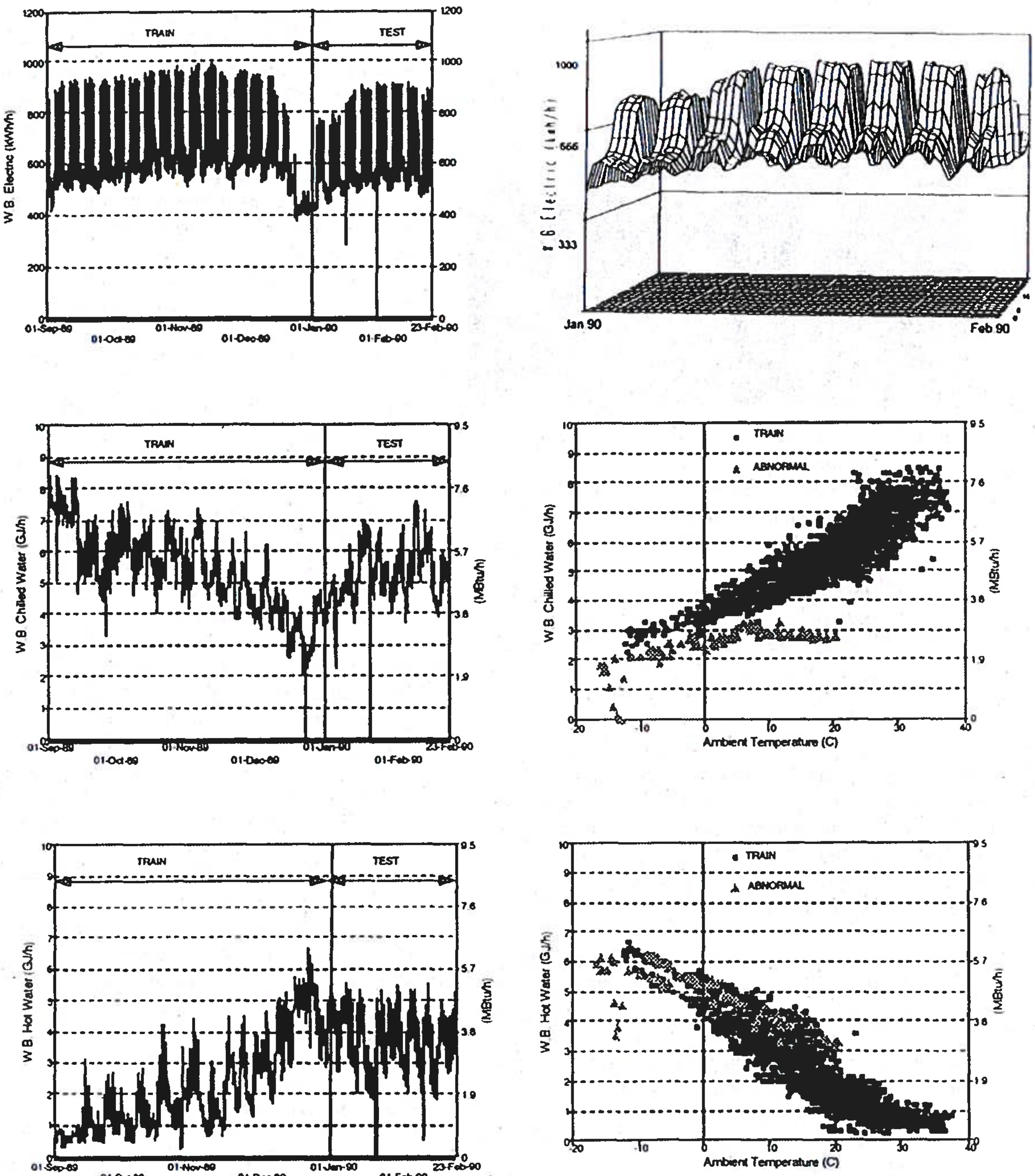

Figure 2 *Training and testing data for the EC. The training period was from September 1, 1989, to December 31, 1989, and the testing data set was from January 1, 1990, to February 28, 1990. The upper-left plot is a time-series trace of the whole-building lights and receptacles load; the upper-right plot is a three-dimensional profile of electricity use for the testing period only. The third figure represents whole-building chilled-water consumption, and the fourth is whole-building hot-water use. The abnormal data represent a period of extremely cold temperatures for this central Texas location during which chilled-water flow was interrupted because of frozen pipes.*

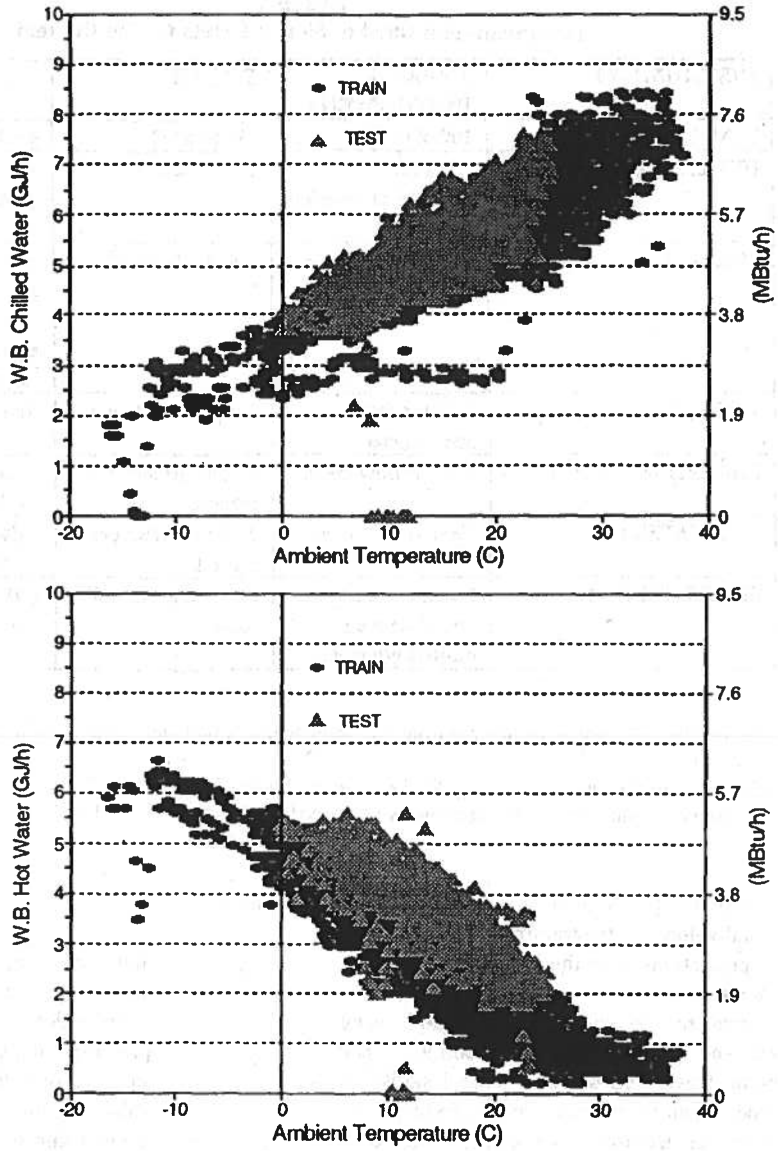

Figure 3 *Plots of the chilled-water and hot-water data. These plots for the training and testing periods show cooling and heating energy vs. outdoor dry-bulb temperature.*

with the data from the four fixed, silicon cell pyranometers. These five insolation values and a time stamp comprised the solar data set. The insolation values were collected between August 1988 and May 1989 in Boulder, Colorado. Hourly averaged insolation levels in this period ranged from 0.0 to more than 1,100 W/m^2.

For the solar testing data set, predictions of hourly radiation were to be made by each entrant for 900 records of only four values of insolation from the multiple-pyranometer array (MPA); the testing set was randomly selected from the full data set. The data set B exercise was essentially one of interpolation in contrast to the predictions required for data set A.

COMPETITION PROCEDURES

The prediction tasks differed between the data sets (the sets were chosen expressly to emphasize different analytical problems). The values of the withheld testing set's dependent variables that were used to evaluate the predictions after the close of the competition were not available to any of the entrants until after all submissions had been received.

For data set A, each contestant submitted hourly predictions (i.e., forecasts) for chilled-water, hot-water, and whole-building electricity use for the two months following the four-month training set. Each contestant also prepared graphs similar to Figures 2 and 3 for their predictions (both

TABLE 1
Instrumentation Used to Measure Data for the Contest

INSTRUMENT:	TYPE OF INSTRUMENT:	RANGE:	ACCURACY:
TEMPERATURE (1)	100 ohm RTD	-50 to 50 C	±0.3 C
HUMIDITY (1)	thin film capacitance type RH sensor	5 to 90% RH	±2%, 0-80%RH ±3%, +80%RH
SOLAR (1)	silicon photodiode solar cell	0 to 3000 W/m^2	±3 % of reading, ±5-10% of reading at large incidence angles, not cosine corrected.
WIND (2)	low threshold anemometer	1 mph to 100 mph	Note (2)
CHILLED WATER (3)	Venturi flow meter.	0.25 to 30 feet per second.	±5% at vel > 3fps, +10% at vel< 3fps.
HOT WATER (3)	Venturi flow meter.	0.25 to 30 feet per second.	±5% at vel > 3fps, +10% at vel< 3fps.
ELECTRICITY (1)	Secondary current transducers on utility kWh meter.	0 to 5 amps at full scale.	± 2% of reading.

Notes:
1. The values for the range and accuracy are from the manufacturer's literature. Similar solar cells were used for the B data set.
2. The manufacturer of this device cites only the low threshold wind speed.
3. These are estimated values based on experiments performed by Robinson et al. (1992).

for training and testing sets) as well as a scatter plot of predictions vs. actual values for the training set. For data set B, beam radiation predictions from the four MPA-measured radiation values were required.

The seminar organizers evaluated all submissions using identical software and procedures for all entrants. The performance ranking was based on the standard statistical measures of the coefficient of variation and mean bias error (the mean bias error was accorded a secondary status to be used as a tie-breaker in the event that two contestants produced predictions with the same coefficient of variation, the primary ranking statistic). These two statistical measures are defined by:

coefficient of variation (CV):

$$CV \equiv \frac{\sqrt{\dfrac{\sum_{i=1}^{n} (y_{pred,i} - y_{data,i})^2}{n}}}{\bar{y}_{data}}$$

and mean bias error (MBE):

$$MBE \equiv \frac{\sum_{i=1}^{n} (y_{pred,i} - y_{data,i})}{n\bar{y}_{data}}$$

where

$y_{data,i}$ = data value of the dependent variable corresponding to a particular set of values of the independent variables,

$y_{pred,i}$ = predicted dependent variable value for the same set of independent variables above (these values are the predictions by the entrants),

$\bar{y}_{data}$ = mean value of the dependent-variable testing data set, and

n = number of records of data in the testing set.

RESULTS AND SUMMARY

The results produced by the competitors were the predictions of the dependent variables for the two testing sets of independent variables. This section summarizes the performance of the top six entrants. The final part compares the winning results with several classic statistical techniques for building energy data set A.

Table 2 summarizes the results of the competition. Each contestant is identified by a random identification number (the six winners are identified in the "Acknowledgments" section) shown in the left column. Table 3 shows the methods used by each entrant. Note that for an entry to be accepted, predictions for both data sets were needed.

TABLE 2
Shootout Competition Results
(values in percent)

DATA SET A									SET B		OVERALL	
ID #	WBE CV	WBE MBE	CHW CV	CHW MBE	HW CV	HW MBE	AVG CV	AVG MBE	MPA CV	MPA MBE	GLOB CV	GLOB MBE
9	10.36	8.06	13.02	−6.37	15.24	−5.84	12.87	6.75	3.20	0.32	10.46	5.15
6	11.78	10.50	12.97	−5.95	30.63	−27.33	18.46	14.59	2.75	0.17	14.53	10.99
2	11.89	8.01	13.69	−6.67	31.65	−27.55	19.08	14.08	8.16	−0.15	16.35	10.59
8	16.95	6.20	14.32	−8.25	29.75	−26.19	20.34	13.55	4.91	−0.05	16.48	10.17
11	16.10	12.56	18.06	−9.79	28.08	−21.26	20.75	14.54	3.98	0.20	16.55	10.95
3	12.79	7.33	12.78	−5.31	30.98	−27.10	18.85	13.24	9.78	−0.31	16.58	10.01
21	12.87	8.60	12.96	−5.86	30.57	−26.88	18.80	13.78	12.28	0.23	17.17	10.39
10	22.01	−10.42	22.88	−17.52	17.85	−5.67	20.91	11.20	6.19	0.17	17.23	8.45
16	19.56	6.06	14.80	−7.56	30.25	−25.48	21.54	13.04	4.45	0.12	17.27	9.81
20	15.86	13.85	12.69	−5.16	33.36	−30.42	20.64	16.48	9.26	0.84	17.79	12.57
15	16.44	13.37	14.14	−6.73	40.74	−36.84	23.77	18.98	4.00	0.17	18.83	14.28
4	20.07	17.92	11.65	1.19	38.71	−35.57	23.48	18.23	8.14	−4.75	19.64	14.86
19	20.93	9.14	15.08	−8.22	35.14	−32.38	23.72	16.58	8.70	0.16	19.96	12.47
14	14.59	5.93	17.33	−9.34	29.58	−24.79	20.50	13.35	18.51	0.30	20.00	10.09
12	37.10	1.58	31.21	−12.5	44.92	−17.53	37.74	10.54	6.94	0.17	30.04	7.95
1	20.76	−0.39	24.99	−11.14	32.29	−21.25	26.01	10.92				
17	29.97	18.92	22.09	14.19	30.17	−22.43	27.41	18.51				
5	26.67	5.97	15.54	−7.21	31.75	−26.82	24.66	13.33				
18	30.91	−13.77	33.26	−27.54	66.45	−62.75	43.54	34.69				
7	13.81	11.84	13.63	−6.01	30.57	−27.04	19.34	14.96				

TABLE 3
Methods Used by Entrants

ID	Method
9	Bayesian non-linear modeling (winner #1)
6	Feedforward multilayer perceptron (winner #2)
2	Neural network with pre and post processing (winner #3)
8	Conjugate gradient neural network (winner #4)
11	Piecewise linear regression (winner #5)
3	Neural network (winner #6)
21	Linear regression and kriging
10	Proprietary similarity based algorithm
16	Neural network
20	Regression
15	Piecewise linear perceptron
4	Graphical displays, regression and neural nets
19	Enhanced neural network
14	Machine learning, symbolic, CSNAP
12	Local linear approximations based on K-nearest neighbors
1	Neural network
17	Neural network
5	Cascade correlation feedforward neural net
18	Neural fuzzy
7	Recurrent neural nets

Note: These key words are those used by the entrants to describe their methods.

The fine differences in the predictions of the six top contestants can be seen by examining Figures 2 through 6. In the upper-right plot in Figure 2, a three-dimensional profile of the measured electricity use is presented for testing set A. In this three-dimensional profile, the x-axis contains the 59 days from January 1, 1990, through February 28, 1990. The y-axis (i.e., into the page) is the 24-hour profile of a day's electricity use; the z-axis (or the height above the x-y plane) is the electricity use. Similar three-dimensional profiles are provided for each contestant's predictions in Figure 4.

Several features are worth noting. First, none of the contestants was able to accurately predict the first two weeks of electricity use during the testing period. This is to be expected, since this represents a period between semesters when the building is sparsely occupied as the faculty, staff, and students return from the holidays. Second, there were no holidays in the testing data period—a feature that was not communicated to the contestants. Possible holidays during this period could have included Martin Luther King day, President's day, and Ash Wednesday. Finally, comparisons between the contestants' predictions reveal subtle differences as well. For example, the exceptional performance of contestant 1 (CV = 10.36; see previous section for definition) can be clearly seen at a glance by comparing the testing data and the predictions in this three-dimensional presentation format. Contestants 2 (CV = 11.78) and 6 (CV = 12.79) also had good whole-building electricity profile predictions.

Differences in the chilled-water and hot-water predictions are apparent by comparing Figures 5 and 6 to Figure 3; however, they are slightly more difficult to detect. For example, in Figure 5, all contestants seem to have the appropriate temperature dependence in their chilled-water predictions. However, differences seem to occur in the variation of the chilled-water prediction for a given ambient temperature. This is to be expected in a building that has varying load profiles. The contestants had visibly different predictions, although not necessarily much different CV values.

The plots of hot-water predictions reveal features similar to the chilled-water predictions with one exception. Contestant 1 had a remarkably good fit for the hot-water prediction (CV = 15.24). The next best result was that of contestant 5 (CV = 28.08). In the graphs, this can be seen as an underprediction of hot water by all contestants other than contestant 1. The remarkable accuracy of the prediction of contestant 1 is borne out by the fact that the measured hot water for the testing period actually did increase when compared to hot water use for the training period, as shown in Figure 3. Again, this is thought to be due to the fact that several academic departments moved out of the building during the holiday break—the most noteworthy feature of which was the removal of the computer science department's computers; this had the effect of reducing the whole-building electricity, thereby increasing hot water use (to make up for the reduced heating produced by computer-generated heat).

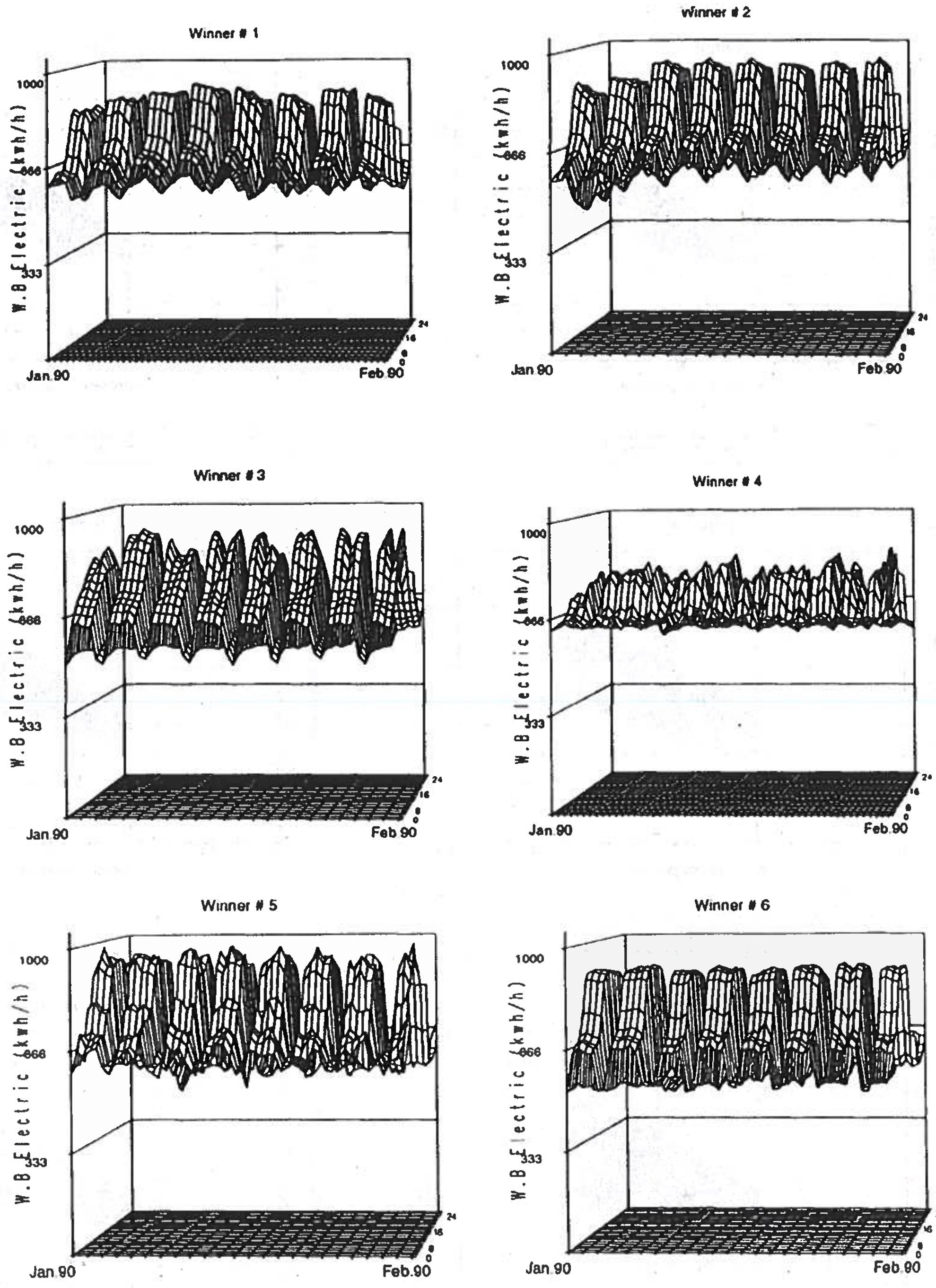

Figure 4 *Three-dimensional profiles of the predicted whole-building electricity use by winners 1 through 6. The results of the prediction efforts by the six winners are displayed. The actual data for this period are shown in the upper-right plot of Figure 2.*

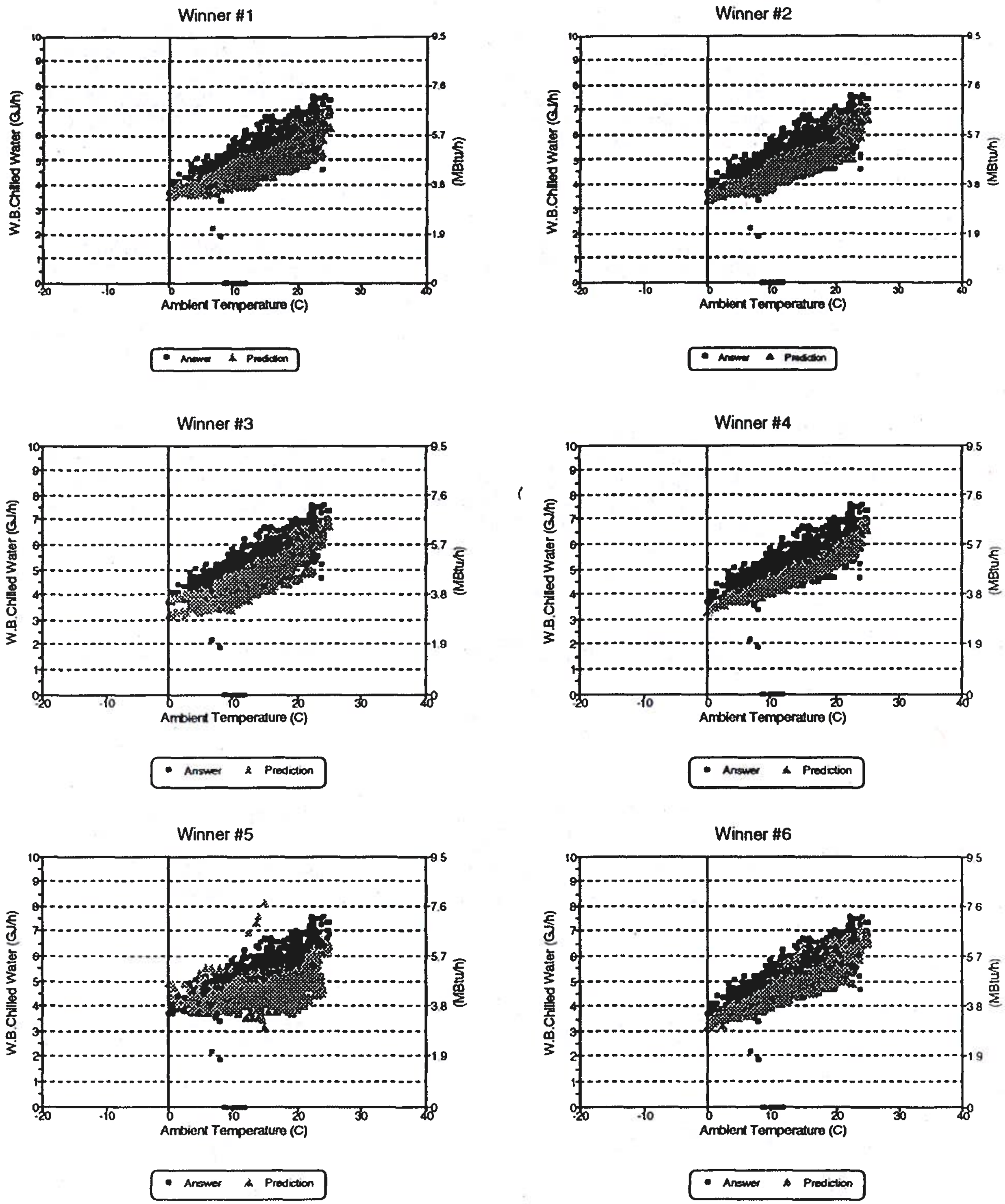

Figure 5 *Chilled-water predictions. The results of the predicted hourly chilled-water use for the six winners are shown. The darkened symbols represent the predictions for the test period. The plus symbols are the actual data for the test period.*

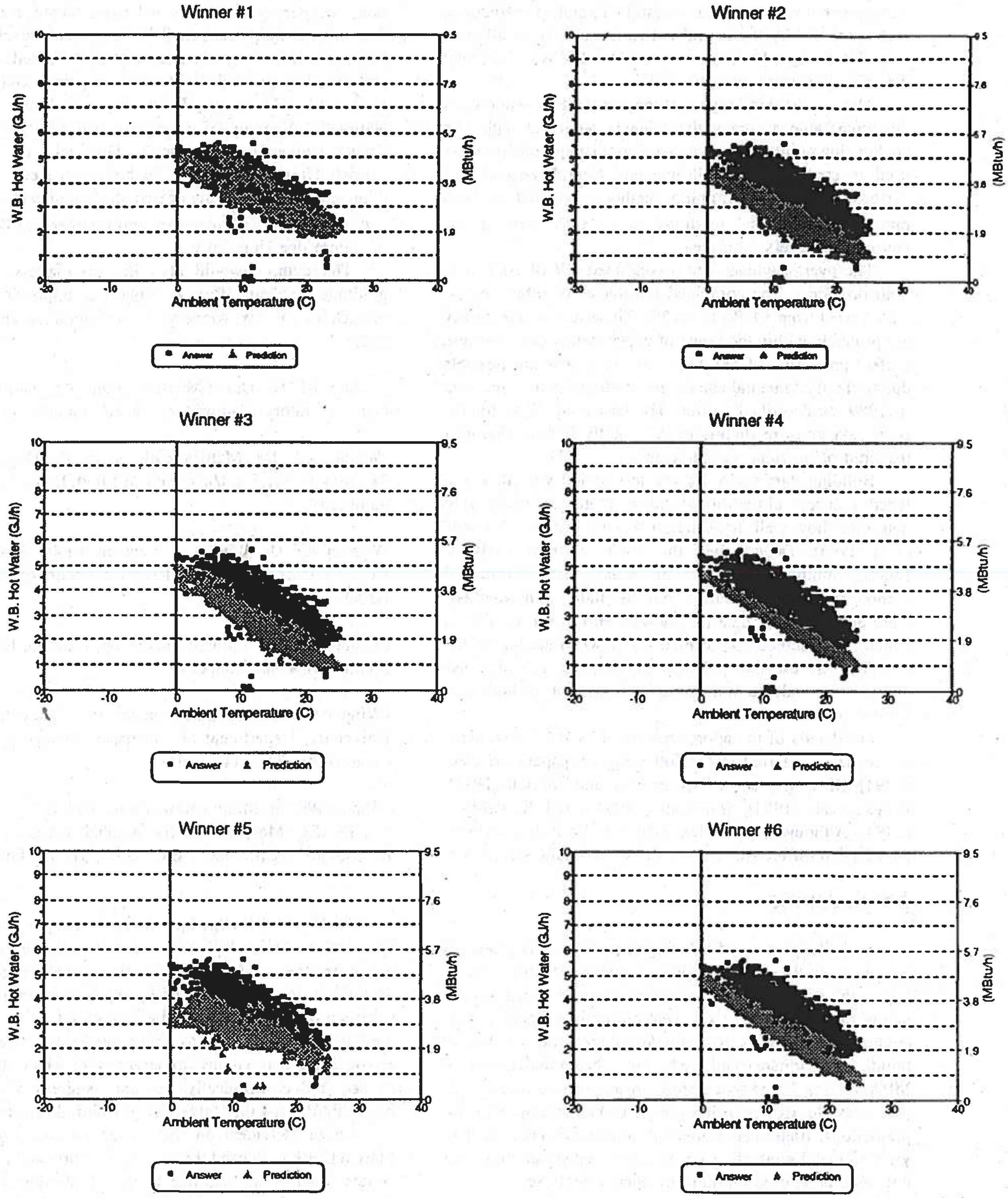

Figure 6 *Hot water predictions. The results of the predicted hourly hot water use for the top six winners are shown. As in Figure 5, the darkened symbols represent the predictions for the test period. The unfilled symbols are the actual data.*

Figure 7 shows the correlation between the 900 actual and predicted values of solar radiation. Excellent predictions were achieved by the neural network methods of all winners. On average, for the six winners the CV was 5.5% and the mean bias error was only 0.2%.

The most striking aspect of the competition, aside from the remarkable accuracy the winners achieved with two challenging problems, is that connectionist approaches were used in some form by all winners. Neural networks of various designs and training methods excelled at both problems. Traditional methods, at least as used by the entrants, were less accurate.

The overall winner had a combined CV of 10.5% for both problems. The combined accuracies of other contestants varied from 14.5% to 16.5%. These levels of accuracy are probably within the range of experimental error because perfect prediction of building data set A was not possible due to the fundamental change in building use that occurred in 1989, as described earlier. The best predictions for the solar data set were slightly more than 3% RMSE, also near the limit of instrumentation accuracy.

Building data set A for the test period was also analyzed with several traditional statistical analysis methods to determine how well these methods could perform (Kissock et al. 1993). Table 4 lists the results of these analyses. Linear, multiple-linear, change-point linear, principal component analysis, Fourier analysis, and binning methods were applied to the data (as shown) with mixed results. In general, all methods performed better when predicting the chilled-water use than when predicting the hot-water use and, as expected, the more complex statistical methods performed better.

For details of the approaches used by the six winners, the reader is referred to the following six papers: Mackay (1994), Ohlsson et al. (1994), Feuston and Thurtell (1994), Iijima et al. (1994), Stevenson (1994), and Kawashima (1994). Although each author used a connectionist method, the detailed differences among the six are quite significant.

CONCLUSIONS

A challenging set of building energy analysis problems was posed to the data analysis community. The top submittals to the competition were remarkably accurate given the nature of the data and the severe constraints placed on the contestants. It appears that both decent predictions of hourly building performance and prediction of beam radiation from MPA devices, based on carefully measured past history, are now possible. However, in order to make the most accurate predictions, traditional statistical approaches may need to yield to novel methods, such as neural networks, that have not often been used by building energy analysts.

ACKNOWLEDGMENTS

The authors express their recognition to the more than 150 persons who requested the contest data, including the 21 entrants who submitted results from the following institutions: University of Illinois at Urbana-Champaign, London School of Economics, Mobil Research and Development at Princeton, University of Arizona-Tucson, Waseda University (Japan), Universidad Politechnic (Spain), University of Karlsruhe (Germany), University of Lund (Sweden), National University of Singapore, University of Bologna (Italy), University of Colorado, University of Wisconsin, Cornell University, United Technologies Resource Center (Connecticut), University of British Columbia at Vancouver, and the Cavendish Laboratory and Engineering Department of Cambridge University.

The authors would also like to express a debt of gratitude for their efforts in support of improving building science to the six winners of the predictor shootout, in order:

(Winner #1) Dr. David Mackay, Radio Astronomy, Cavendish Laboratory, Cambridge, UK (ID number 9);

(Winner #2) Dr. Mattias Ohlsson et al., Department of Theoretical Physics, University of Lund, Lund, Sweden (ID number 6);

(Winner #3) Dr. Bradley P. Feuston, Mobil Research and Development Corp., Central Research Laboratory, Princeton, NJ (ID number 2);

(Winner #4) Mr. William Stevenson, Pretoria, Republic of South Africa (ID number 8);

(Winner #5) Mr. Makato Iijima and Ryo Takeuchi, Waseda University, Department of Electrical Engineering, Tokyo, Japan (ID number 11); and

(Winner #6) Mr. Minoru Kawashima, University of Wisconsin, TSARC, Madison, WI (ID number 3) (now at Institute of Technology, Shimizu Corporation, Tokyo, Japan).

The Centre d'Energetique provided an opportune setting for one author (Kreider) to initiate and conclude the competition, and he thanks Denis Clodic of the Centre for his hospitality and generosity with computer resources. Peter Curtiss provided the solar radiation data set. Mike Mozer of the University of Colorado Computer Science Department provided the anonymous ftp server from which data could be acquired electronically. He and Andreas Weigend of Xerox PARC and the University of Colorado consulted with one author (Kreider) on the design of the competition. Marian Clark organized the data distribution and logged the entries. Thanks are also due to David Claridge and Kelly Kissock, who assisted with the shootout and this paper; to Dennis O'Neal and John Bryant for providing the building instrumentation; and to Sabaratnam Thamilseran and Mustafa Abbas for the preparation of the graphs in this paper.

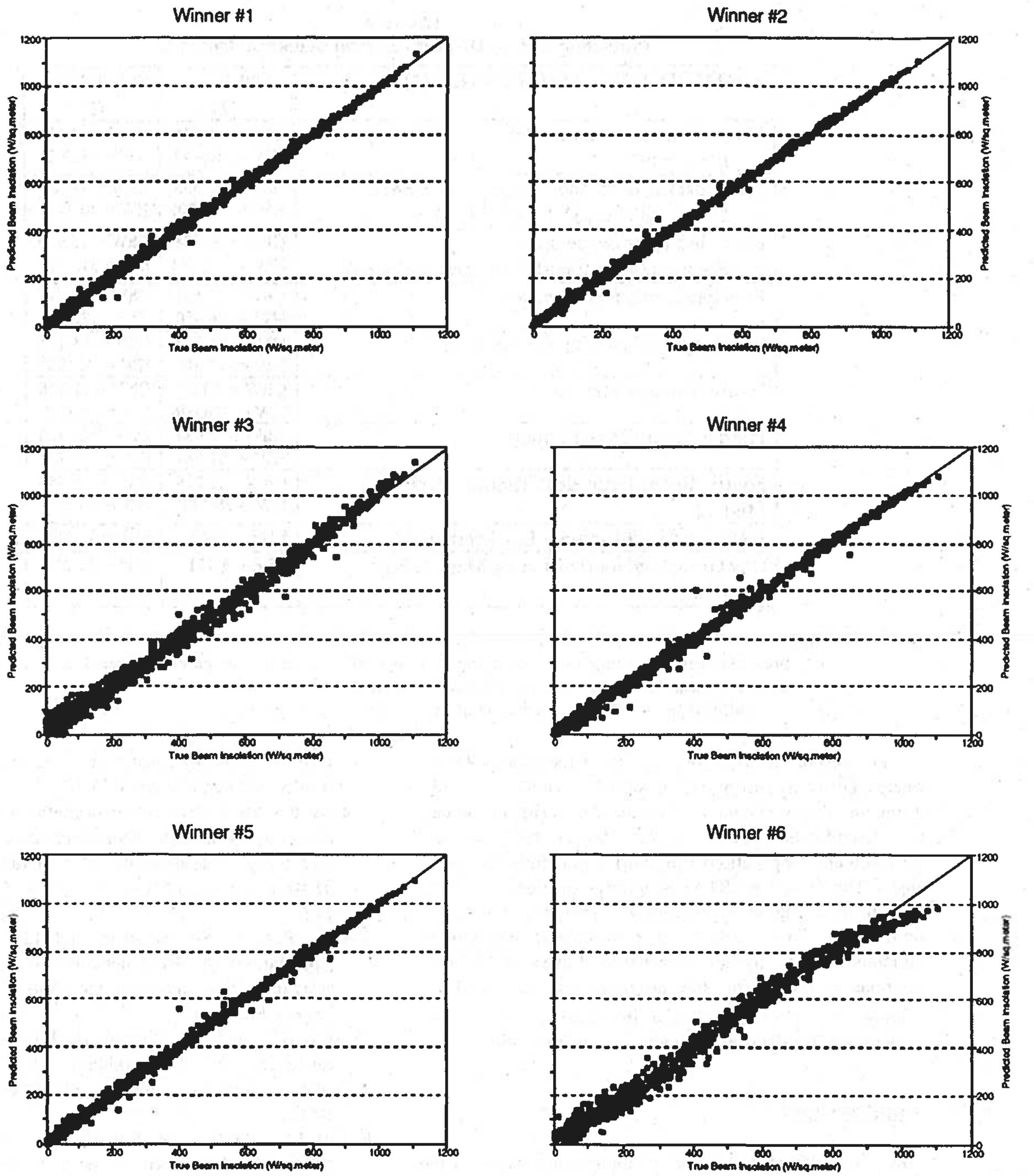

Figure 7 *Predicted vs. actual solar beam radiation for contestants 1 through 6. X-Y plots of the results of the six predictions vs. actual solar beam radiation are shown in these series of plots.*

TABLE 4
Predicting Energy Use with Several Statistical Methods

METHOD AND BASIC MODEL:(1)	M.B.E. (2)	CV-RMSE (3)
Two parameter linear model: $E = a + b(T)$	ChW = -6.425 HW = -26.442	ChW = 13.521 HW = 29.963
Three parameter change-point linear model: $E = a + b(T-T_{cp})^{+} + C(T - T_{cp})^{-}$	ChW = -5.553 HW = -27.209	ChW = 13.566 HW = 30.726
Multiple Linear Regression: $E = a + b(T) + c(sph) + d(solar) + e(wind)$	ChW = -6.204 HW = -27.843	ChW = 13.925 HW = 31.342
Principle Component Analysis:	ChW = -7.171 HW = -24.980	ChW = 14.097 HW = 28.680
Weekday/weekend Binning Method:	ChW = -5.176 HW = -26.668	ChW = 13.163 HW = 30.088
Hourly Binning Method:	ChW = 1.196 HW = -26.076	ChW = 12.728 HW = 29.671
Fourier Series/Basic Method:	ChW = -5.634 HW = -37.944	ChW = 13.405 HW = 47.727
Fourier Series/Equivalent Thermal Parameter Method:	ChW = -5.224 HW = 28.123	ChW = 12.883 HW = 31.169
Fourier Series/Electricity Use Method:	ELE = 8.003	ELE = 11.868
Day Grouping/Hourly Binning Method For Electricity Use:	ELE = 3.131	ELE = 11.268

Notes:
1. For additional information on the statistical analysis methods used, see the paper by Kissock et al. (1993).
2. M.B.E. is mean bias error. Notice that the MLR has R^2.
3. CV-RMSE is the coefficient of variation of the root mean square error.

Portions of this work were funded by the State of Texas Energy Office as part of Texas A&M's LoanSTAR Monitoring and Analysis contract, and the Texas Higher Education Coordinating Board (under Project #227 Energy Research and Applications Program) is gratefully acknowledged. The Texas LoanSTAR program is an eight-year, $98 million revolving loan program for energy conservation retrofits in Texas state, local government, and school buildings funded by oil overcharge dollars. Additional information concerning the program can be found in Claridge et al. (1991). For further information and/or copies of the predictor data set, contact either of the authors.

REFERENCES

Bronson, D. 1992. Calibrated computer simulations for the analysis of retrofit energy savings. Master's thesis, Energy Systems Laboratory. Report nos. ESL-TH-92/04-02 and ESL-TH-92/04-01. College Station: Texas A&M University.

Bronson, D., S. Hinchey, J. Haberl, and D. O'Neal. 1992. A procedure for calibrating the DOE-2 simulation program to non-weather-dependent loads. *ASHRAE Transactions* 98(1): 636-652.

Claridge, D., J. Haberl, D. O'Neal, W. Heffington, D. Turner, C. Tombari, M. Roberts, and S. Jaeger. 1991. Improving energy conservation retrofits with measure results. *ASHRAE Journal* 33(10).

Curtiss, P.S. 1992. An analysis of methods for deriving th constituent insolation components from multipyranom eter array measurements. *Proceedings of the 199 ASME International Solar Energy Conference*, pp. 109 117.

Curtiss, P.S., S. Starkweather, and J.F. Kreider. 199 Applications of neural networks to multipyranomete array data. *Transactions of the ASME, Journal of Sola Energy Engineering*.

Feuston, B.P., and J.H. Thurtell. 1994. Generalized nonlin ear regression with ensemble of neural nets: The gre energy predictor shootout. *ASHRAE Transactio* 100(2).

Haberl, J.S., J.D. Bronson, S.B. Hinchey, and D.L. O'Nea 1993a. Graphical tools to help calibrate the DOE- simulation program: Comparative 3-D surface plo improve the ability to view small differences betwee simulated and measured data. *ASHRAE Journal* 35(1 27-28.

Haberl, J., D. Bronson, T. Bou-Saada, and D. O'Nea 1993b. A report on the impact of using measur weather data versus TMY weather data in a DOE simulation of an existing building in central Texa Energy Systems Laboratory Report No. ESL-TR-93/0 02. College Station: Texas A&M University.

Hinchey, S. 1991. Influence of thermal zone assumptions on DOE-2 energy use estimations of a commercial building. Master's thesis, Mechanical Engineering Department, Energy Systems Laboratory Report No. ESL-TH-91/08-06. College Station: Texas A&M University.

Iijima, M., K. Takagi, R. Takeuchi, and T. Matsumoto. 1994. A piecewise-linear regression on the ASHRAE time series data. *ASHRAE Transactions* 100(2).

Katipamula, S., and D. Claridge. 1992. Monitoring air handler performance and comparing it with a simplified system model. *ASHRAE Transactions* 98(2).

Katipamula, S.K., and J. Haberl. 1992. A methodology to identify diurnal load shapes for non-weather-dependent electric end-uses. *Solar Engineering 1991: Proceedings of the ASME-JSES-JSME International Solar Energy Conference*, pp. 457-467, Reno, NV.

Kawashima, M. 1994. Artificial neural network backpropagation model with three-phase annealing developed for the building energy predictor shootout. *ASHRAE Transactions* 100(2).

Kissock, K., T. Bou-Saada, D. Bronson, B. Cox, A. Dhar, S. Katipamula, S. Thamilseran, A. Reddy, J. Haberl, and D. Claridge. 1993. Summary of shootout modeling techniques. Energy Systems Laboratory Report No. ESL-PA-93/04-11 (April).

Mackay, D.J. 1994. Bayesian nonlinear modeling for the energy prediction competition. *ASHRAE Transactions* 100(2).

Ohlsson, M., C. Peterson, H. Pi, T. Rognvaldsson, and B. Soderberg. 1994. Predicting utility loads with artificial neural networks—Methods and results from the great energy predictor shootout. *ASHRAE Transactions* 100(2).

Robinson, J., J. Bryant, J. Haberl, and D. Turner. 1992. Calibration of tangential paddlewheel insertion flowmeters. *Proceedings of the Eighth Symposium on Improving Building Systems in Hot and Humid Climates*, Energy Systems Laboratory, Texas A&M University, pp. 222-228.

Stevenson, W.J. 1994. Predicting building energy parameters using artificial neural nets. *ASHRAE Transactions* 100(2).

DISCUSSION

Will Preska, Senior Development Engineer, Honeywell, Golden Valley, MN: Are there any real products coming out of all this?

Jan F. Kreider: Neural network (NN) products in HVAC are a novelty because of the relatively short time that research first indicated promise for them and because of the conservative nature of the HVAC industry. I am aware of prototype controllers, energy prediction packages, and HVAC system diagnostics software that have at least some NN components.

K. Vincent Wong, Department of Mechanical Engineering, University of Miami, Coral Gables, FL: Many of the nonwinners used neural networks. Can you pinpoint what they did differently to get different answers?

Kreider: The shootout co-chairs did not investigate the methods used by those not in the top six. However, anecdotal evidence suggests that those ranking lower made a less astute choice of inputs, did not use an abundance of physical insight, and did not preprocess the data as extensively. Some of these contestants also had less experience with the practical use of NNs than the winners.