ESL-PA-96/07-03-1

The Great Energy Predictor Shootout II: Measuring Retrofit Savings—Overview and Discussion of Results

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

Sabaratnan Thamilseran
Student Member ASHRAE

ABSTRACT

A second predictor shootout contest has been developed and conducted to evaluate the most effective empirical or inverse models for modeling hourly whole-building energy baselines for purposes of measuring savings from energy conservation retrofits. This second contest utilized two sets of measured hourly pre-retrofit and post-retrofit data from buildings participating in a revolving loan program in Texas (Claridge et al. 1994). The accuracy of the contestants' models was evaluated by determining their ability to predict data that were carefully removed from the training (or pre-retrofit) period. A comparison of the savings predicted by the models is also presented.

The results from the contest show that neural networks again provide the most accurate model of a building's energy use. However, in contrast to the first contest (Kreider and Haberl 1994a), the second contest's results show that cleverly assembled statistical models also appear to be as accurate or, in some cases, more accurate than some of the neural network entries. When these models were used to forecast the baseline use into the post-retrofit period, large variations in the predicted savings occurred among the models, particularly for the cooling energy savings in one of the case study buildings. These variations appear to be due to the ability of the models (or inability) to capture certain energy performance characteristics and the modeler's assumptions about the post-retrofit energy use.

INTRODUCTION

Overview

Based on the overwhelming response to the first Building Energy Predictor Shootout, a second Shootout was developed to again compare how well different empirical or inverse models predict building energy use from several new data sets and to compare how those models could be used to calculate energy savings from conservation retrofits. ASHRAE's TC 1.5 and TC 4.7 authorized the "Building Energy Predictor Shootout II: Measuring Retrofit Savings," which was held from June to November 1994. This paper describes the second competition and discusses the results from the four top entries.

Purpose/Objective

The purpose of the ASHRAE Predictor Shootouts has been to provide the building analysis community (as well as the scientific community in general) with a clearly defined, scientific test of different methods of predicting hourly building energy use. Predictions of hourly energy use have been shown to be useful for diagnostics and building energy retrofit performance calculations.

In the first contest the competition was a controlled competition that contained two data sets and nothing else (i.e., no description of the building or other specific details about the data). Contestants were asked to build empirical models using the training data sets and then predict energy use and solar measurements for a test data period for which they did not have the answers. Predictions from the contestants were then evaluated by the contest organizers, the top winners were announced, and the specifics about the data sets were made available to the contestants so that papers could be written that detailed their modeling efforts. A summary of the first contest can be found in the June 1994 *ASHRAE Journal* (Kreider and Haberl 1994a) as well as the accompanying *ASHRAE Transactions* papers (Kreider and Haberl 1994b; Feuston and Thurtell 1994; Iijima et al. 1994; Kawashima 1994; MacKay 1994; Ohlsson et al. 1994; Stevenson 1994). ASHRAE has also published a special publication that contains all the papers and the data set from the first contest (ASHRAE 1994).

Jeff S. Haberl is an associate professor in the Department of Architecture and Sabaratnan Thamilseran is a graduate research assistant in the Mechanical Engineering Department, Energy Systems Laboratory, Texas Engineering Experiment Station, Texas A&M University, College Station.

THIS PREPRINT IS FOR DISCUSSION PURPOSES ONLY, FOR INCLUSION IN ASHRAE TRANSACTIONS 1996, V. 102, 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 12, 1996.

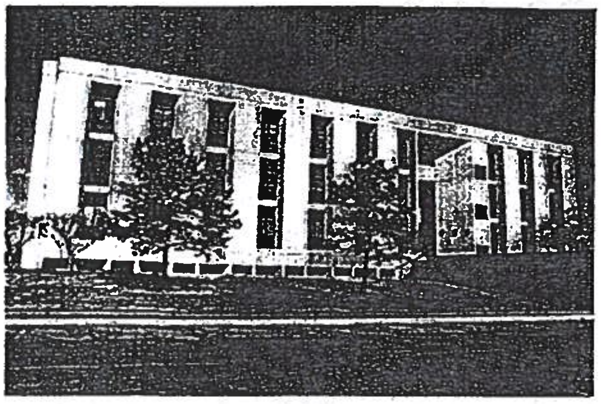

Figure 1 The Engineering Center (EC). The four-story Engineering Center is a large, multipurpose building located on a university campus in central Texas.

DESCRIPTION OF THE CASE STUDY BUILDINGS

To facilitate the second contest, data were collected from two buildings that received retrofits from the Texas LoanSTAR program (Claridge et al. 1994). These buildings were carefully chosen from a group of buildings so that they could provide the contestants with two significantly different data sets. Both of the case study buildings are described in the following section.

Case Study 1: The Engineering Center

The engineering center (EC), located in central Texas on a university campus, contains 30,000 m^2 (324,000 ft^2) of classrooms, offices, a computer center, and laboratory facilities comprising four stories and includes an unconditioned underground parking garage. Figure 1 is a picture of the building. It was constructed in the early 1970s and is a heavy structure with 0.15-m (6-in.) concrete floors and insulated exterior walls made of precast concrete and porcelain-plated steel panels. About 12% of the exterior wall area is covered with single-pane, bronze-tinted glazing. The windows are recessed approximately 0.61 m (2 ft) from the exterior walls, which provides some shading. Approximately 288 m^2 (310 ft^2) of northeast-facing clerestory windows admit daylight into the core of the building. The building is occupied from 7:30 a.m. to 6:30 p.m. on weekdays and has reduced occupancy from 7:30 a.m. to 5:30 p.m. on weekends. The computer facility is operated 24 hours per day.

The building is primarily served by 12 dual-duct air-handling units located in the underground parking garage. Chilled and hot water for the cooling and heating coils are supplied to the building by the campus physical plant.[1] Two multizone units and a dedicated centrifugal chiller serve a supercomputing facility located in a special room within the building. Manual operation of the secondary chilled and hot-water pumps also affects the system's cooling and heating capacity. Prior to the March 1991 variable-air-volume (VAV) retrofit, the outside air dampers were permanently set to supply about 10% to 20% outdoor air and did not operate in an economizer mode (Katipamula and Claridge 1992). The primary retrofit to the building was to replace the existing constant-air-volume (CAV) air distribution systems with variable-speed-drive, VAV air distribution systems. During the retrofit an energy management and controls system was also upgraded. Additional information about the building can be found in Bronson (1992), Bronson et al. (1992), and Haberl et al. (1993, 1995).

In the engineering center about 50 channels of hourly data have been recorded and collected each week since May 1989. The important channels for the retrofit savings measurement are those for air handler electricity consumption and whole-building heating and cooling energy use. Air handler electricity consumption is measured at the building's motor control center (MCC) and represents all of the air-handling units and most of the heating, ventilating, and air-conditioning (HVAC)-related pumps in the building. Cooling and heating energy use are determined by monitoring the fluid flow rate and temperature difference across the supply and return lines of the chilled- and hot-water supply to the building. Figure 2 shows the data that were provided to the contestants for the engineering center for the pre-retrofit training period (c.trn file) and post-retrofit periods (c.tst file).

Case Study 2: The Business Building

The business building (BUS), located at a university in the Dallas-Fort Worth area, contains 13,926 m^2 (149,900 ft^2) of gross conditioned space and consists of six stories of classrooms, offices, and lecture halls. Figure 3 is a photograph of the business building. It was constructed in 1970 and is face brick on block wall construction. About 10% of the exterior wall area is covered with single-pane, tinted glazing. The building is occupied from 8 a.m. to 6 p.m. Monday through Friday for regular classes and occasionally after 6 p.m. for night classes and on weekends for special purposes.

The building is served by three large dual-duct air-handling units that were retrofitted to operate as VAV dual-duct systems in July 1991. Chilled water and hot water (from steam) for cooling and heating coils are supplied to the building by the campus utility plant. During the pre-retrofit period the air-handling units (AHUs) were scheduled to operate 24 hours per day, although there was a considerable variation in the daily electricity consumption. This variation was caused by the following: part-load operation of the AHUs during the building's unoccupied period and irregular on/off combinations of pumps and fans. In the post-retrofit period, a reduced number of the units were operated during occupied periods (i.e., 17 hours per day, seven days per week). The primary retrofits were to replace the constant-air-volume (CAV) air distribution systems with a variable-speed-drive, VAV air distribution system and to improve the control of lighting using occupancy sensors in classrooms.

In the business building 15 channels of hourly data are metered, including whole-building electricity use, MCC electricity use, whole-building heating energy use, whole-building cooling energy use, and lighting energy use. Figure 4 shows

[1] The building also receives steam for heating domestic hot water. These steam consumption data were not included in the current study.

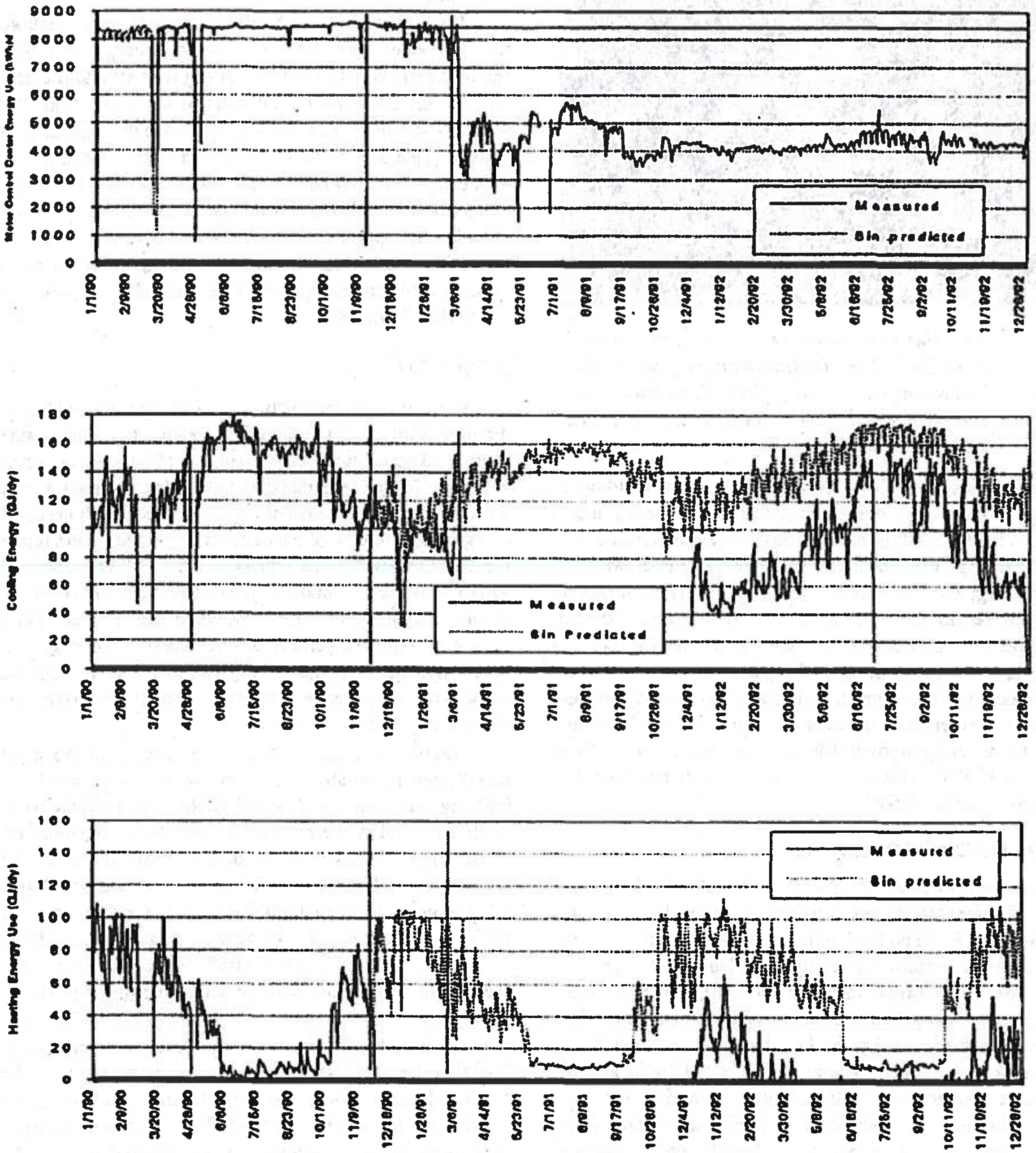

Figure 2 *Daily time series plots for the Engineering Center. In the upper figure daily motor control center electricity use (MCC) is shown, followed by chilled water and hot water use in the middle and lower graphs. Daily summaries of the hourly data are shown for the training, construction, and post-retrofit periods (indicated by the two vertical lines). Data were removed from the pre-retrofit training period for testing purposes. Missing data occur for two months in 1990 for the MCC and lights and equipment channels, from April through November 1991 for chilled water, and during April 1991 and June through December 1991 for hot water. In the post-retrofit period the daily energy use predicted by a bin model is also shown.*

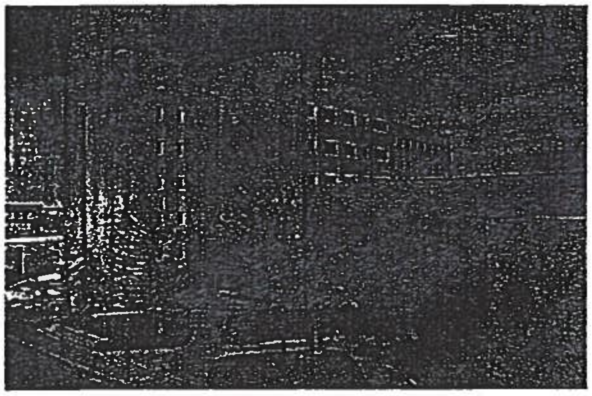

Figure 3 The Business Building (BUS). The six-story business building is also a multipurpose building containing classrooms, offices, and lecture halls. It is located at a university campus in the Dallas-Fort Worth area.

daily time series plots of the data from the business building. In the upper figure, daily lights and receptacle electricity use is shown. In the second figure, MCC electricity use is shown, and in the third graph from the top, chilled-water use is shown, with heating energy use shown in the bottom graph. Data shown are for the pre-retrofit training period and construction and post-retrofit periods. Lighting and receptacle electricity use is a derived channel that is obtained by subtracting motor control center electricity use from the whole-building electricity use. Data were also removed from the Business Building's training period for testing purposes. Missing data occurred for three months in 1991 for chilled water and hot water and for a few weeks in December 1992.

COMPETITION PROCEDURES

The competition began in June 1994 and ended November 1, 1994. To enter the competition contestants needed to visit the anonymous FTP site on the Internet, download the training and testing data sets, perform their analysis, and submit their answers in the agreed-upon format. In the second contest the actual post-retrofit data were supplied to the contestants since their models' accuracy was determined from data that were withheld during the training period (i.e., pre-retrofit period). Graphical daily summaries similar to the data in Figures 2 and 4 were also provided in the contest instructions to give the contestants a general idea of what to expect from the different channels of data.

At the close of the competition the results were analyzed, and the winners were announced and were invited to write papers that describe their modeling methods (Chonan et al. 1996; Jang et al. 1996; Katipamula 1996; Dodier and Henze 1996). Interest in the second contest was significant, although down somewhat from the first contest. Fifty visitors downloaded the general "readme.txt" file that described the contest, and 11 contestants tried their hand at the contest and downloaded the training and testing data sets. Four contestants completed the analysis in the agreed-upon format before the closing date of November 1, 1994.

The results produced by the contestants are predictions of the dependent variables (i.e., building energy use) for the removal periods in the "*.trn" training data set and predictions that use the pre-retrofit parameters in the post-retrofit "*.tst" period. These predictions were submitted to the organizers, who compared the predictions of the removed data against the actual data using an hourly coefficient of variation (CV(RMSE)) and mean bias error (MBE).[2] After the analysis was complete, a statistical and graphical comparison was developed and distributed to the contestants in June 1995 and they were encouraged to begin writing a description of their methods to be published in the *ASHRAE Transactions*.

DATA SETS

As mentioned previously, data sets were provided from two different buildings in four files. The pre-retrofit training data sets (*.trn) and post-retrofit testing data sets (*.tst) contain independent variables (i.e., weather data and a calendar time stamp) and the corresponding dependent variables (e.g., whole-building energy use). Portions of the dependent variables were removed from the training files and replaced with "-99." The independent variables that corresponded to the removed data remained in the training file and were used by the contestants to predict energy use for the removed periods. The predictions of energy use for the removal period in the training data sets were then compared to the actual data (known only by the contest organizers) to test the accuracy of the contestant's model.

The two post-retrofit "*.tst" files contained independent and dependent variables from the post-retrofit period for each building. The contestants were required to use their baseline models to predict energy use for every hour contained in the "*.tst" file and submit their predictions in the required format. Predictions in the post-retrofit period were compared against other contestants' predictions to determine the differences in the savings calculations. Savings were calculated as the difference between the contestants' post-retrofit predictions (which use the pre-retrofit parameters) and the actual measured post-retrofit energy use.

The data for the engineering center were contained in the "C.trn" and "C.tst" data sets and consisted of the following hourly data: whole-building electricity (kWh/h), motor control center electricity (kWh/h), lights and equipment electricity (kWh/h), whole-building chilled water (MBtu/h), whole-building hot water (MBtu/h), ambient temperature (°F), ambient relative humidity (%), global horizontal solar (W/m^2), and wind speed (mph). It had the following format:[3]

---

[2]The equations for the CV(RMSE) and MBE are described in the Appendix.

[3]The -99 has been used to replace the removal data in the pre-retrofit "*.trn" training set. Instrumentation used for measuring the energy use in both buildings is similar and has been previously described in Kreider and Haberl (1994b).

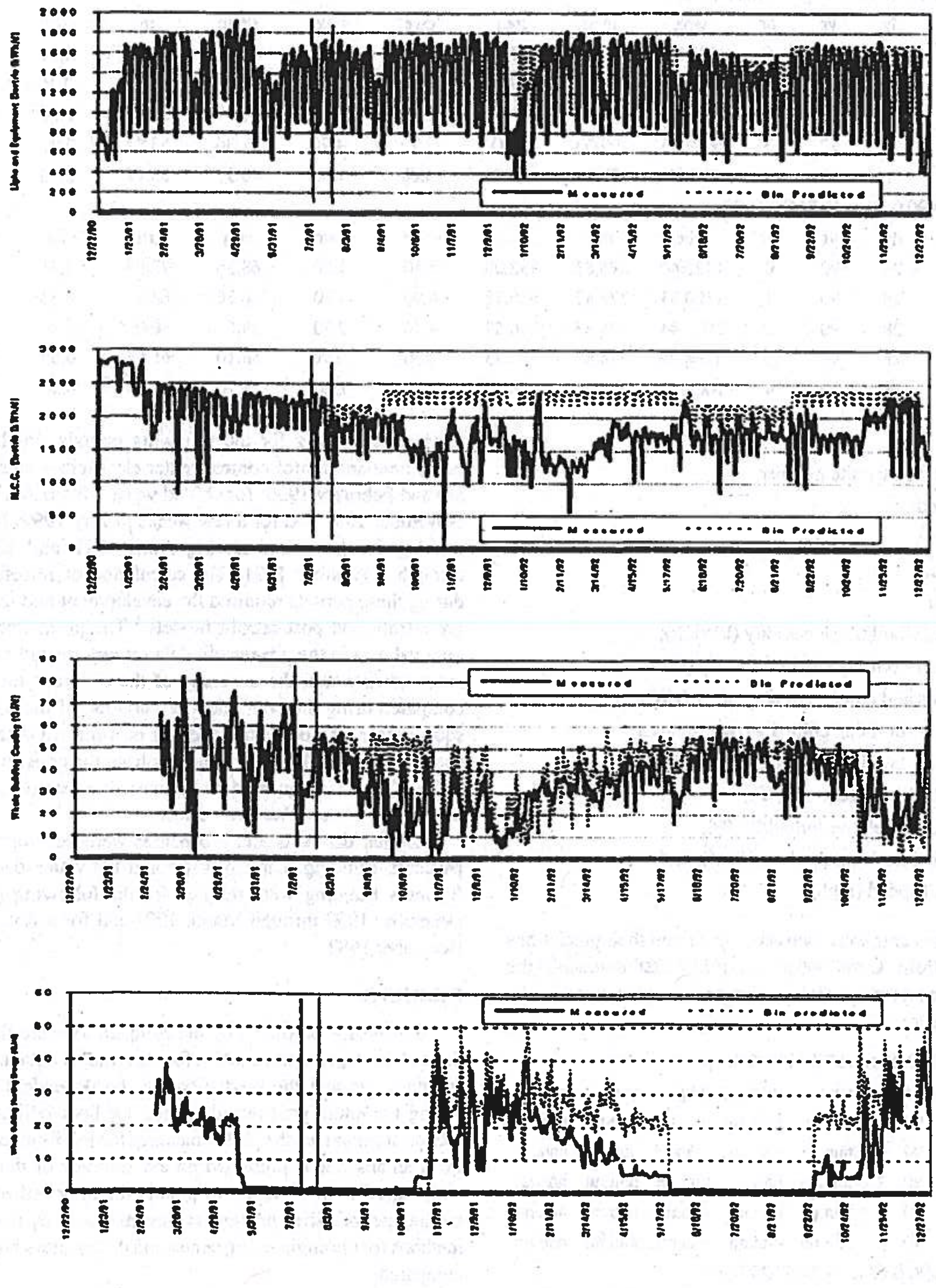

***Figure 4** Daily time series plots of the Business Building (BUS). Beginning with the upper figure, daily lights and equipment electricity use is shown followed by motor control center electricity (MCC), chilled water and heating energy use. Data shown are for the pre-retrofit training period, construction and post-retrofit periods. Data also were removed for the training period for test purposes. Missing data occurred for three months in 1991 for chilled water and hot water and for a few weeks in December 1992. In the post-retrofit period the daily energy use predicted by a bin model is also shown.*

C.TRN data set (1/1/90 0:00 to 11/27/90 23:00):

site	mo	dy	yr	hr	wbe	mcc	lteq	cwe	hwe	temp	rh	sol	wspeed
1	1	1	90	0	883.00	-99.00	-99.00	3.70	4.70	42.99	53.32	2.00	4.88
1	1	1	90	1	877.30	-99.00	-99.00	3.70	4.70	41.80	55.48	2.10	4.86
1	1	1	90	2	878.00	-99.00	-99.00	3.60	4.70	41.49	54.02	2.20	4.61
1	1	1	90	3	880.80	-99.00	-99.00	3.70	4.70	41.36	53.22	2.20	4.88
1	1	1	90	4	879.80	-99.00	-99.00	3.60	4.80	40.99	53.17	2.10	4.34

C.TST (11/28/90 0:00 to 12/31/92 23:00):

site	mo	dy	yr	hr	wbe	mc	lteq	cwe	hwe	temp	rh	sol	wspeed
1	11	28	90	0	1126.40	375.83	632.08	5.20	1.60	68.36	72.95	5.91	4.87
1	11	28	90	1	1100.03	376.42	606.78	4.90	1.70	66.36	68.95	6.53	7.08
1	11	28	90	2	1081.44	376.33	586.52	4.70	2.30	59.63	80.06	6.84	8.35
1	11	28	90	3	1064.36	374.57	571.73	4.50	2.70	56.10	66.54	6.53	9.41
1	11	28	90	4	1061.10	375.09	568.40	4.40	2.80	53.70	67.70	6.53	5.65

where

site = an arbitrary site number,
mo = month,
dy = day,
yr = year,
hr = hour,
wbe = whole-building electricity (kWh/h),
mcc = motor control center electricity (kWh/h),
lteq = lights and equipment electricity (kWh/h),
cwe = whole-building chilled water (MBtu/h),
hwe = whole-building hot water (MBtu/h),
temp = ambient temperature (°F),
rh = ambient relative humidity (%),
sol = global horizontal solar (W/m^2), and
wspeed= wind speed (mph).

The contestants were then asked to submit their predictions for the pre-retrofit "C.trn" and post-retrofit "C.tst" data sets in the following format (where their predictions were inserted into the "nnn.nn" values):

C.TRN (1/1/90 0:00 to 11/27/90 23:00):

site	mo	dy	yr	hr	wbe	mcc	lteq	cwe	hwe
1	1	1	90	0	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	1	1	90	1	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	1	1	90	2	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	1	1	90	3	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	1	1	90	4	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn

C.TST (11/28/90 0:00 to 12/31/92 23:00):

site	mo	dy	yr	hr	wbe	mcc	lteq	cwe	hwe
1	11	28	90	0	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	11	28	90	1	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	11	28	90	2	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	11	28	90	3	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn
1	11	28	90	4	nnn.nn	nnn.nn	nnn.nn	nnn.nn	nnn.nn

Data were missing for the following periods: for lights and equipment and motor control center electricity—during January and February 1990; for chilled water—from April 1991 to November 1991 and for a few weeks in July 1992. Data were missing for hot water during April 1991 and from June through November 1991. The calculation of retrofit savings during these periods required the development and use of both pre-retrofit and post-retrofit models.[4] The predictions for the removal data in the pre-retrofit data set were tested against the removed data and the accuracy of the different models was compared using the coefficient of variation of the root-mean-square error. The predictions for the post-retrofit data set were used to calculate the energy savings from the conservation retrofits. These savings predictions were compared against the predictions of the other contestants.

Similar data sets and procedures were developed for the Business Building. Chilled-water and hot-water data for the Business Building were missing for the following periods—December 1990 through March 1991 and for a few weeks in December 1992.

RESULTS

The results produced by the competitors were the predictions of the dependent variables for the removal data in the training data sets and the predictions of the dependent variables during the entire post-retrofit period for both buildings. This section summarizes the performance of the top four contestants. First, an analysis is presented on the accuracy of their models with regard to their ability to predict the removed data in the training period. Then energy savings calculated by the different methods for the engineering center and the business building are compared.

Evaluation of the Accuracy of the Models

Table 1 gives the statistics for the models produced by the contestants for the removal data for the training period for the

[4] Post-retrofit models were created by the contest organizers using an inverse bin method to fill in the missing data (Thamilseran and Haberl 1995).

TABLE 1 The Predictor Shootout II Competition Results

		Coefficient of Variation - CV					Mean Bias Error - MBE					
		E1	E2	E3	E4	E5		E1	E2	E3	E4	E5
	wbe	3.1205	13.2116	8.6475	2.9032	3.1647	wbe	0.2722	-1.8049	-6.5555	-0.0907	-0.1237
EC	mcc	3.2803	3.4728	3.2811	3.5751	3.1796	mcc	0.4817	0.5929	0.4871	0.3613	0.2165
	lteq	4.456	25.6786	3.4384	4.2476	4.3546	lteq	0.5106	-2.6949	0.3971	-0.7068	-0.4096
	cwe	7.1312	8.2585	8.877	7.0312	9.4499	cwe	-0.8917	-3.0309	-3.4214	-1.3372	-1.689
	hwe	21.2758	39.2016	35.3721	16.5914	24.6541	hwe	-3.0988	-15.3095	-10.9781	-2.1426	-2.874
EC		7.85276	17.96462	11.92322	6.8697	8.96058						
	wbe	17.0462	26.8642	20.8693	16.5368	14.9939	wbe	2.7898	4.2215	-1.6086	2.2887	1.1941
	mcc	17.5371	41.0527	17.7616	22.2904	23.5215	mcc	4.4054	-0.1351	4.1446	0.3206	0.4455
BUS	lteq	21.3187	43.9609	21.1616	17.0489	17.0553	lteq	-3.4806	14.6371	-3.4868	0.6829	0.2512
	cwe	55.9467	53.0195	40.9747	42.0513	51.805	cwe	-3.6721	-16.5221	-8.7733	-11.4627	-19.2526
	hwe	46.0517	47.3523	24.7441	36.8548	33.7135	hwe	-11.1313	-8.31	-4.8605	-12.2519	-6.6527
BUS		31.58008	42.44992	25.10226	26.95644	28.21784						
Overall		19.7164	30.2073	18.5127	16.9131	18.5892		-1.3815	-2.8356	-3.4655	-2.4338	-2.8894
Rank		3	4	2	1							

NOTE: This table contains a comparison of the overall accuracy of the models developed by the top four contestants for the two buildings in the competition. A fifth entry by the contest organizers is shown for comparative purposes. Values are shown for the prediction of the whole-building electricity use (wbe), motor control center electricity use (mcc), lights and receptacles electricity use (lteq), chilled water (cwe), and heating energy use (hwe) for both the Engineering Center, and Business Building.

Engineering Center and Business Building. Contestants are indicated by entry number[5] for the two statistics used to evaluate the contestant's models for the Engineering Center and the Business Building. The overall ranking, shown at the bottom of the table, was determined from the coefficient of variation of the root-mean-square-error. The mean bias error was to be used as a secondary parameter in the event of a tie.

Table 2 describes the methods used by the four contestants as reported in their respective papers that describe their work (Dodier and Henze 1996; Katipamula 1996; Chonan et al. 1996; Jang et al. 1996). Figures 5, 6, and 7 give a graphical look at the results of the predictions for the Engineering Center. Figures 8, 9, and 10 show the results for the Business Building. Additional information concerning a detailed look at the accuracy of the models can be found in Thamilseran and Haberl (1995).

The overall winner of the second contest was E4, who scored a global CV(RMSE) of 16.91% and an MBE of -2.43%. This contestant used a combination of 10 neural networks (i.e., one for each of the target variables) with two hidden layers of 25 units each. The accuracy of the model by winner 1 varied from an average of 6.87% for the Engineering Center to an average of 26.95% for the Business Building. In both the Engineering Center and the Business Building the dependent variables that were weather dependent were more of a challenge to predict than the electricity use. All the dependent channels in the Business Building were more difficult to predict than those in the Engineering Center, which is a consequence of the Business Building's on/off energy-use characteristic. The on/off characteristic can clearly be seen as the increased scatter in the temperature-dependent plots and as a larger diurnal variation in the whole-building electricity plots. Typically, the modeling of on/off energy use requires either a dummy variable (to signal the on/off) or some other method of predicting when the systems are on or off.

Winner 2 was close behind winner 1 with a CV(RMSE) of 18.51% and an MBE of -3.46%. In contrast to the other winners, this contestant used a non-neural, net-based statistical daytyping routine for the weather-independent channels and weekday-weekend, hourly multiple regressions for the weather-dependent channels. This is interesting to note because it shows that a finely tuned statistical regression model can perform as well as a neural network

TABLE 2 Methods Used by the Predictor Shootout II Entrants

Identification	Analysis Method Used
Winner #1 (E4)	10 neural networks, 2 hidden layers of 25 units, than activations, single output layer with linear activation, Wald test used for input variable selection.
Winner #2 (E3)	Statistical daytyping routine for weather-independent channels, hourly weekday/weekend statistical multiple regression models for each hour of the day.
Winner #3 (E1)	Bayesian nonlinear regression with multiple hyperparameters. Manual removal of outliers.
Winner #4 (E2)	Feed-forward, autoassociative neural networks. Hyperbolic tangent transfer function.

[5] E1—Waseda University (Chonan et al. 1996), E2—Iowa State University (Jang et al. 1996), E3—Battelle/PNNL (Katipamula 1996), E4—University of Colorado (Dodier and Henze 1996), E5—inverse bin method (Thamilseran and Haberl 1995).

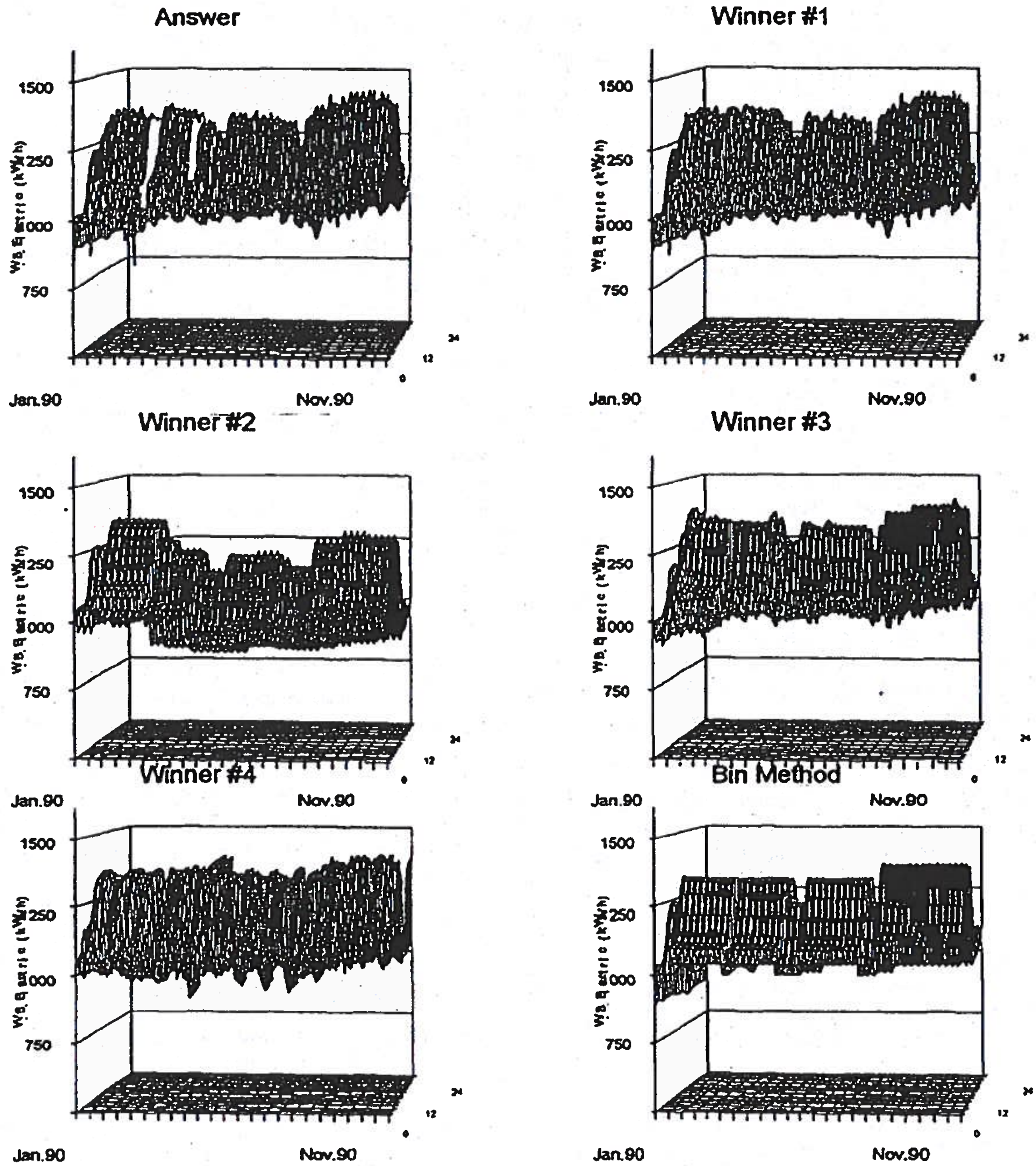

Figure 5 Actual and predicted whole-building electricity use for the Engineering Center. In this figure data are shown as three-dimensional time series plots where the axis into the page is the hour of the day (i.e., 0 to 24) and the x-axis is the day of the year for the training data set. The actual data are shown in the graph labeled "answer" and the contestants' data are shown in their respective graphs. The Bin Method graph represents the data from an analytical method developed by the authors.

The remainder of the models also performed well. Winner 3 scored a CV(RMSE) of 19.71% and an MBE of −1.38% using a Bayesian nonlinear regression with multiple hyperparameters after removal of outliers. Winner 4 had an overall CV(RMSE) of 30.21% and an MBE of −2.83% using a feed-forward autoassociative neural network. Entry 5, which is shown for comparative purposes only, was developed by the authors using an inverse bin method. This method produced an overall CV(RMSE) of 18.59% and an MBE of −2.89%. This method uses a nonlinear, weekday-weekend inverse bin model for weather-dependent variables and weekday-weekend 24-hour profiles for weather-independent loads. Additional information about the method can be found in Thamilseran and Haberl (1995).

The accuracy of the models is remarkable considering the fact that there is a built-in inaccuracy in the method used to evaluate the models. This test inaccuracy is due to the fact that the days that were removed from the training period for testing purposes do not exactly represent the training period. This can be seen in Table 5, where the removed data for the Engineering Center and Business Building are compared against the training data.[6] In the case of the Engineering Center, the average CV(RMSE) comparison of the removed data against the training data was 2.66%, varying from a high of 6.6% for the hot water data to a low of 0.75% for the whole-building electricity use. For

6 The method used to compare the removed data against the training set data is an inverse bin method described in Thamilseran and Haberl (1995).

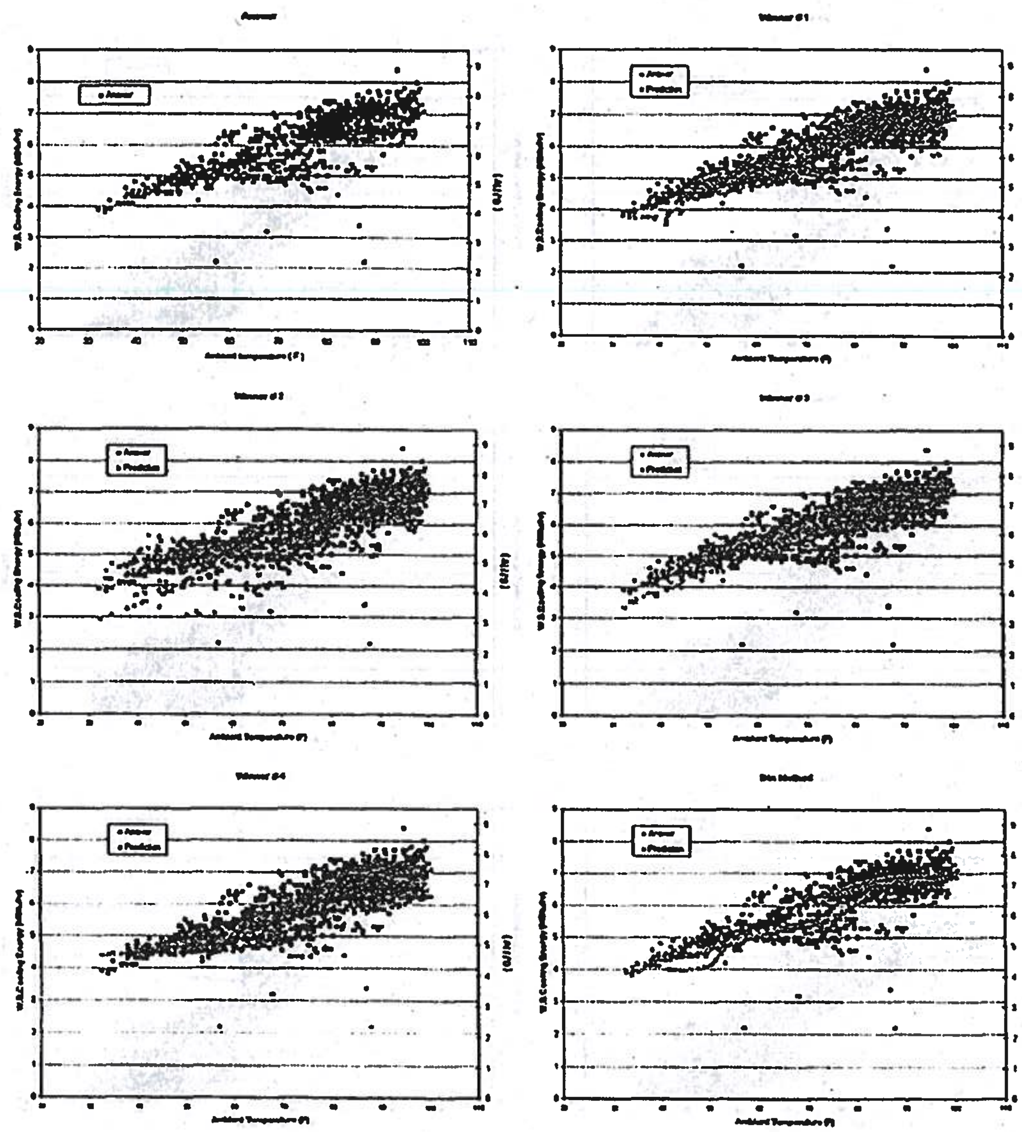

Figure 6 Actual and predicted whole-building cooling for the Engineering Center.

the Business Building, the comparison of the removal data against the training data contained quite a bit more uncertainty, with an average CV(RMSE) of 9.59% that includes a low of 1.95% for the whole-building electricity to a high of 21.45% for the whole-building cooling.

In Figures 5 and 8 the predictions of the whole-building electricity are shown for the Engineering Center and the Business Building, respectively. Figures 6 and 9 compare the accuracy of the hourly cooling models and Figures 7 and 10 illustrate the accuracy of the hourly heating models. There are several features worth pointing out in these figures. First, in the Engineering Center, in the pre-retrofit and post-retrofit periods almost all systems were run 24 hours per day. This is in contrast to the Business Building, where systems are turned on during the day and turned off in the evenings (Figures 9 and 10).

Second, as shown in Figures 2 and 4, the energy savings at the Engineering Center represent a larger portion of the building's energy use versus that of the Business Building. That means that a more accurate model is needed to capture the hourly schedule-dependent loads in the Business Building. This on-off schedule dependence is also apparent in Figure 9, which shows almost no temperature dependence in the cooling energy use, and in Figure 10, which only shows a modest temperature dependence in the heating energy use. Figure 2 also reveals that there were significant missing data in the post-retrofit period for the Engineering Center due to malfunctioning flow sensors. This

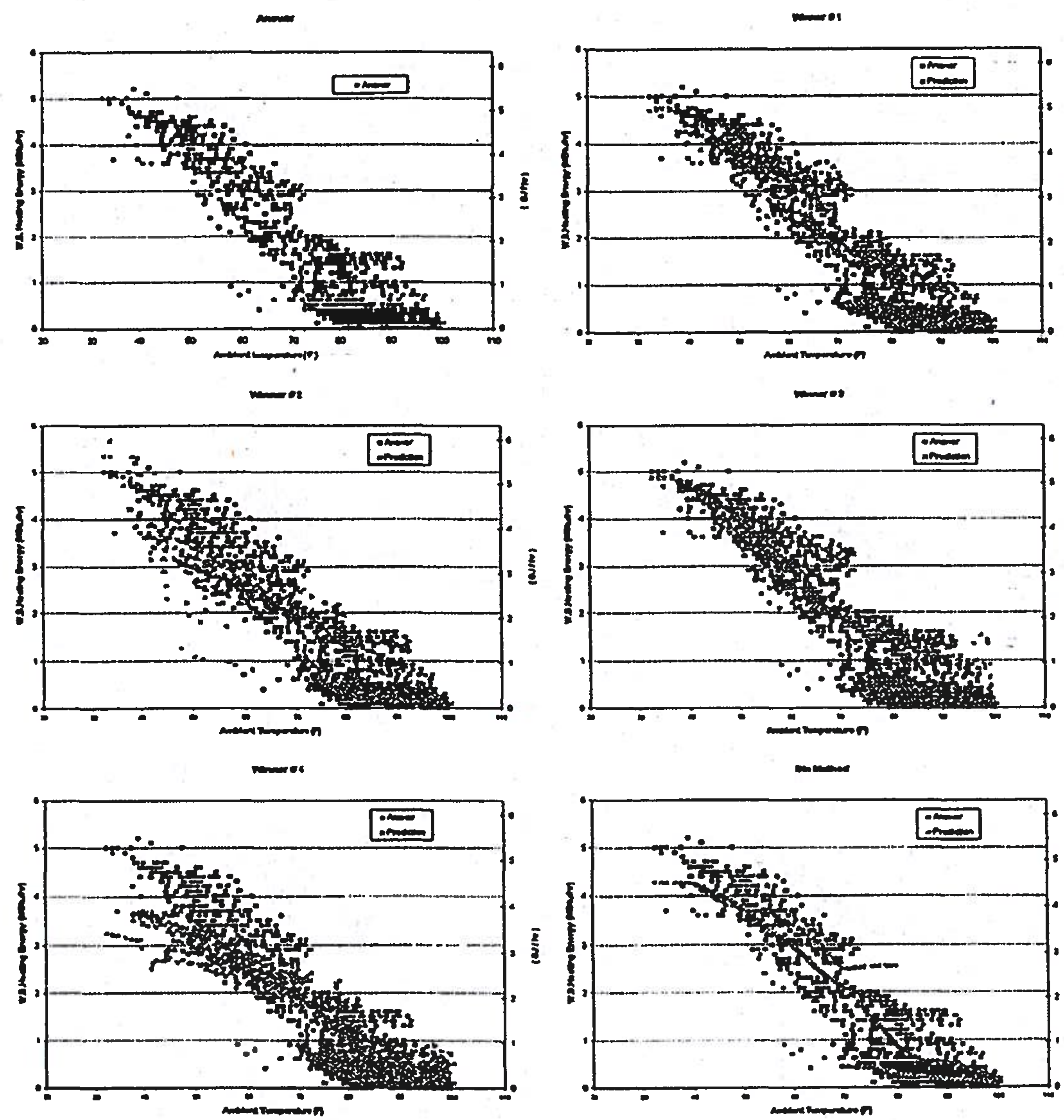

Figure 7 Actual and predicted whole-building heating for the Engineering Center.

required the development and use of both pre-retrofit and post-retrofit models to calculate and compare the savings for the entire 12-month post-retrofit period.

Evaluation of the Savings Predicted by the Models

In general, the evaluation of the accuracy of the four models showed that their ability to predict the removed data was remarkably similar, with the top three models having an overall CV(RMSE) of less than 20%. Unfortunately, the comparison of the savings calculations by the four models for both buildings was mixed. Tables 3 and 4 provide the cost savings predicted by the four models. Figures 11 and 12 provide a graphical comparison of the data presented in Tables 3 and 4

The retrofits implemented for the Engineering Center building (C data set) included variable-speed drives for the motors in the AHUs (which are the major load on the MCC) and conversion of the air-handling units from constant air volume to variable air volume (CAV to VAV). The savings were calculated by comparing the projection of the pre-retrofit models into the post-retrofit period to the actual measured post-retrofit period energy use. For the Engineering Center, the channels affected by the retrofit included the MCC, cooling, and heating channels. Differences in the lights and receptacles (LEQ) and whole-building electricity (WBE) channels, although significant, could not be entirely attributed to the retrofit. Therefore, the savings for the Engineering Center considered only the MCC, cooling, and heating channels.[7] Table 3 and Figure 11 display the savings for the MCC, cooling, and heating channels as calculated by the five methods. Savings for the whole-building electricity and lights and equipment are shown in Figure 11 but are not included in the savings calculations shown in Table 3.

Several features are worth pointing out in Table 3 and Figure 11. First, the average of the total savings for all five methods was remarkably close at $163,058; this represented 38.8% of the post-retrofit energy use, which is substantially above the model's prediction accuracy or CV(RMSE). Second, the highest savings, calculated by E5, were $189,655, which is only 16.3% above the average of all five methods. Third, the lowest savings, calculated by E2, were $83,399, which is 48.9% below the average. Finally, when normalized by conditioned area, the savings varied from $\$0.26/ft^2$ by contestant E2 to $\$0.56/ft^2$ by contestant E4 for the Engineering Center.

In cases where the MCC electricity savings are large compared to the whole-building electricity, the whole-building electricity can be used to measure the savings from the MCC retrofit. However, the WBE channel contains a potentially confounding factor that is hidden in the form of lights and equipment electricity use. This L&E channel represents the electricity use of personal computers, task lighting, laboratory equipment, and photocopying machines—the number and use of which can vary from day to day. In the case of the Engineering Center, all methods predicted little change in lights and equipment savings,

[7] It is possible to model the MCC retrofit at the WBE level. However, significant differences in the semester, nonsemester WBE electricity use that are not related to the MCC complicated the modeling process. This is why the MCC channel was provided to the contestants separately.

TABLE 3 Annual Savings Calculated by the Four Contestants for the Engineering Center for 1992

Entry#	Engineering Center Savings (C.TST)			Total
	mcc	Cool	Heat	
E1	$45,928	$75,808	$56,487	$178,224
E2	$1,043	$19,922	$62,434	$83,399
E3	$45,933	$54,574	$81,562	$182,069
E4	$45,879	$73,034	$63,028	$181,941
E5-BM	$45,836	$71,330	$72,490	$189,655
AVG	$36,924	$58,934	$67,200	$163,058
Total 1993	$237,044 (15.6%)*	$162,409 (36.3%)	$18,519 (36.2%)	$419,972 (38.8%)†

NOTE: In the Engineering Center savings were calculated using the following costs: $0.0278/kWh electricity consumption, $4.67/MBtu cooling energy use, and $4.75/MBtu heating energy use. The last row in the table shows the 1992 annual energy use as reported in the LoanSTAR Annual Energy Consumption Report (AECR) and ratio of the average savings to the 1992 annual utility expenditure by end-use

* 1992 total electricity use shown is for whole-building electricity use. Percentages shown represent the average mcc, cooling, heating, and total savings divided by the respective 1992 post-retrofit energy use.

† 1992 total utility costs includes whole-building electricity, chilled water and hot water as published in the 1993 LoanSTAR Annual Energy Consumption Report (AECR).

TABLE 4 Annual Savings Calculated by the Four Contestants for the Business Building for 1992

Entry#	Business Building Savings (D.TST)				
	mcc	L & E	Cool	Heat	Total
E1	$6,478	$111	($28,927)	$13,264	($9,074)
E2	$20	($218)	$2,655	$13,217	$15,673
E3	$5,332	$1,345	$10,464	$5,682	$22,822
E4	$8,120	($494)	($32,920)	$2,071	($23,223)
E5-BM	$5,968	$767	($3,423)	$14,294	$17,606
AVG	$5,184	$302	($10,430)	$9,706	$4,761
Total 1993	$70,034 (7.4%)*	(0.43%)	$59,371 (17.5%)	$18,338 (52.9%)	$147,743 (3.2%)†

NOTE: In the Business Building the savings were calculated using the following costs: $0.02931/kWh electricity consumption, $4.417/MBtu cooling energy use, and $3.64/MBtu heating energy use. Electric demand savings were not calculated for either building since they are located on university campuses and receive their services from a central plant.

* 1992 total electricity use shown is for whole-building electricity use. Percentages shown represent the average mcc, L&E, cooling, heating, and total savings divided by the respective 1993 post-retrofit energy use.

† 1992 total utility costs includes whole-building electricity, chilled water, and hot water as published in the 1992 LoanSTAR Annual Energy Consumption Report (AECR).

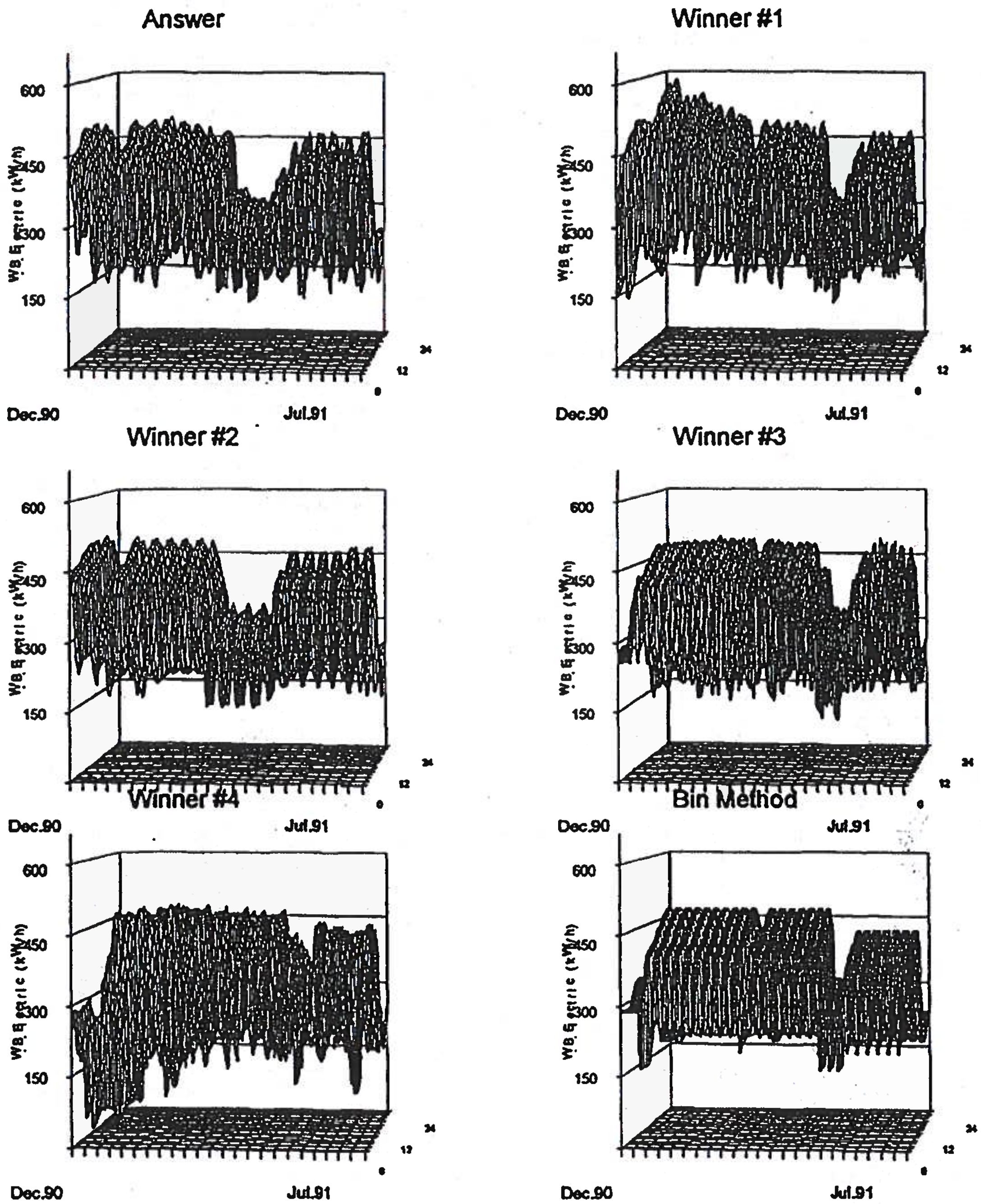

Figure 8 *Actual and predicted whole-building electricity use for the Business Building. In this figure data are shown as three-dimensional time series plots where the axis into the page is the hour of the day (i.e., 0 to 24) and the x-axis is the day of the year for the training data set. The actual data are shown in the graph labeled "answer" and the contestants' data are shown in their respective graphs. The Bin Method graph represents the data from an analytical method developed by the authors.*

which shows small changes in the L&E from "pre-" to "post-" periods. Methods E1, E3, E4, and E5 appear to have similar predictions for the MCC and cooling savings. All methods had similar savings for the heating energy use.

The retrofits implemented for the Business Building (D data set) included CAV to VAV conversion of the air-handling units (AHU) and motion sensors for the lighting. As with the Engineering Center building, the savings were calculated by comparing the projection of the pre-retrofit or baseline models into the post-retrofit period against the actual measured energy use from these channels during the post-retrofit period. For the Business Building the channels affected by the retrofit included the lights and equipment (LEQ) channel, the MCC channel, and the cooling and heating channels. In a similar fashion to the Engineering Center, differences in the Business Building's whole-building electricity (WBE) channels could not be attributed to the retrofit. Therefore, the savings for the Business Building considered only the LEQ, MCC, cooling, and heating channels. Table 4 and Figure 12 display the savings for the LEQ, MCC, cooling, and heating channels as calculated by the five methods for the Business Building. Savings for the whole-building electricity and lights and equipment are shown in Figure 12 but are not included in the totals indicated in Table 4.

The prediction of savings for the Business Building was not as consistent as those for the Engineering Center in several ways. First, only three of the models predicted positive savings from

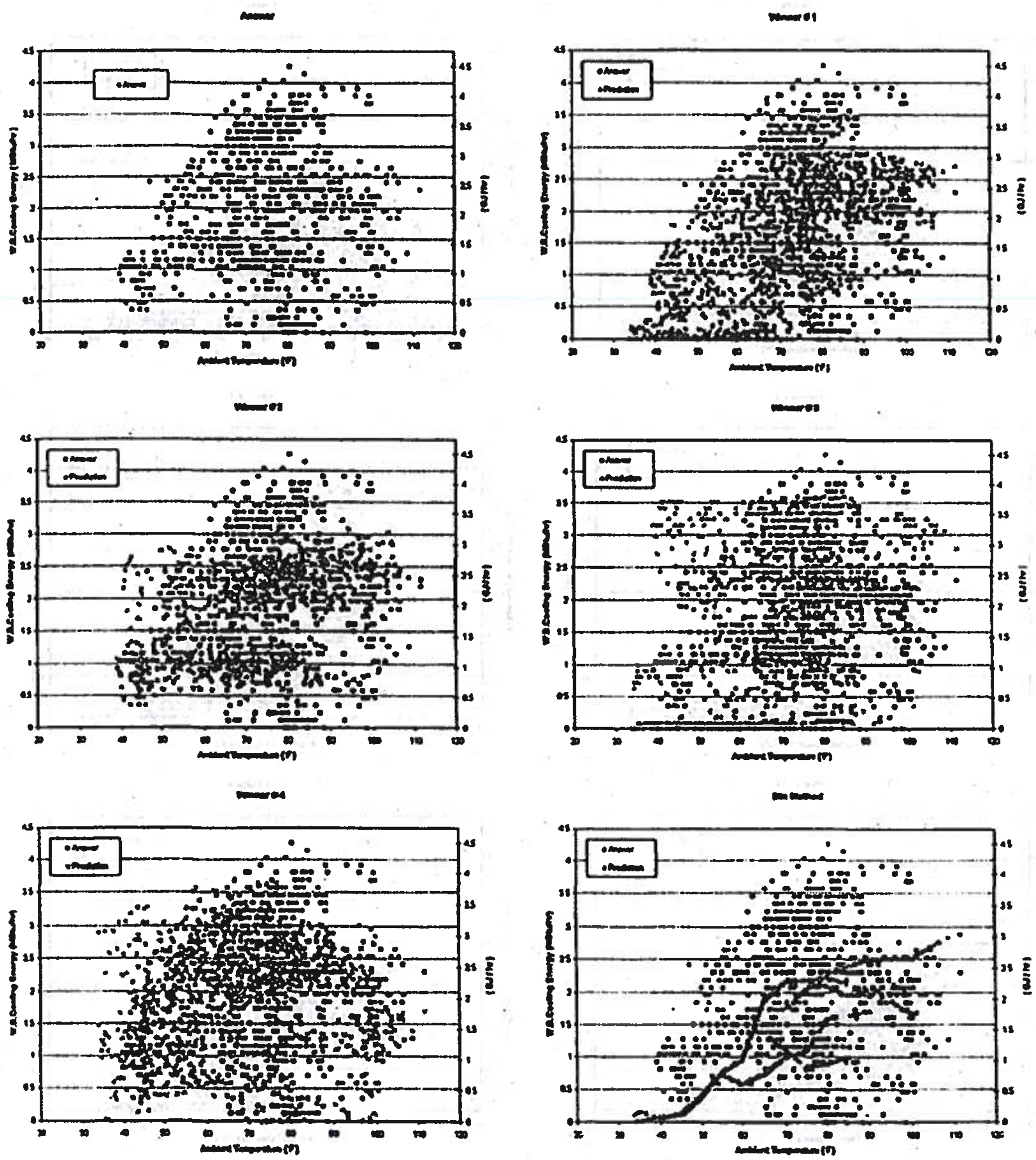

Figure 9 Actual and predicted whole-building cooling for the Business Building.

the retrofits to the Business Building. None of these models was the most accurate model in predicting the training period data. Specifically, the model with the closest fit in the training period (winner 1, contestant E4) was one of the two models that predicted negative total savings for the Business Building.[8] Models E2 (winner 4), E3 (winner 2) and E5 (the bin method) predicted positive savings of $15,673, $22,822, and $17,605, respectively. The variable that seemed to cause the major difference in the predictions was the cooling energy-use savings. Both models E1 and E4 underpredicted the cooling energy savings when compared to the other models. One reason for this underprediction may be the strong on/off (occupied/unoccupied) operation of systems in the building during this period. Such on/off operation makes the savings more difficult to predict since the inefficient systems were not being run during evenings and weekends, which significantly reduced the potential savings..

This would seem to indicate that factors other than the accuracy of the model can play an important role in the model's ability to predict energy use into the future. In the Engineering Center and the Business Building one of the major influencing factors was an assumption that was made about semester and nonsemester schedules. Also, in contrast to the Engineering Center, the predicted savings in the Business Building were much smaller when compared to the total annual utility cost. The

8 The other model that predicted negative savings was that of contestant E1—winner 3.

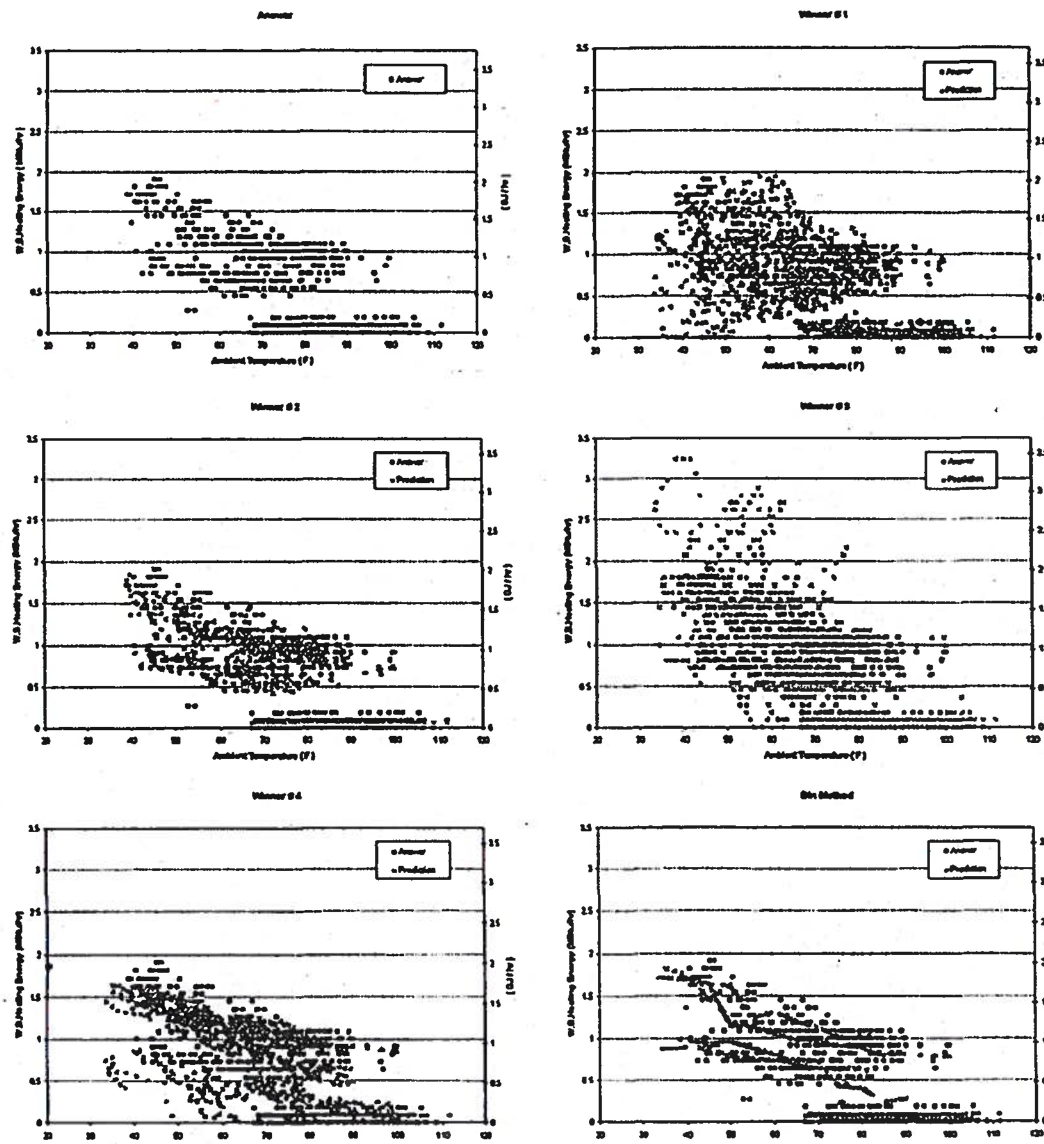

Figure 10 Actual and predicted whole-building heating for the Business Building.

TABLE 5 Average Difference for the "Removed" Data Versus the "Training" Data for the Engineering Center and the Business Building

Channel	Engineering Center (%)	Business Building (%)
WBE	0.75	1.96
MCC	0.11	3.29
L&E	1.17	5.96
Cooling	4.65	21.45
Heating	6.65	15.30
Average	2.66	9.59

total average savings from all the models represented only 3.2% of the annual energy use for the building. With the exception of the heating energy savings, none of the average savings at the Business Building rose above the inherent noise in the models (i.e., the noise predicted by the average CV(RMSE) = 30.8%). When normalized by conditioned area, the savings varied from $-\$0.15/ft^2$ by contestant E4 to $\$0.15/ft^2$ by contestant E3.

SUMMARY AND DISCUSSION

A second predictor shootout contest was developed and conducted to evaluate the most effective models for predicting hourly whole-building energy use for purposes of measuring savings from energy conservation retrofits. This second contest utilized measured hourly pre-retrofit and post-retrofit data from two buildings participating in the Texas LoanSTAR program. The accuracy of the contestants' models was evaluated by determining their ability to predict data that were purposefully removed from the training (or pre-retrofit) period. A comparison of the savings predicted by the models was also presented.

The results from the second contest reconfirm certain of the results from the first contest and have also provided some exciting new insights about the use of such baseline models for calculating the savings from energy conservation retrofits. First, the results of the second contest show that neural networks again provide the most accurate model of a building's energy use. However, the accuracy of the neural network entries varied according to the assumptions that the contestants made about the training data sets and how skilled the contestants were in choosing and assembling their networks. One of the surprising results of the second contest was the fact that cleverly assembled statistical models appear to be as accurate or, in some cases, more accurate than some of the neural network entries.

The second surprise was the fact that when these models were used to forecast the baseline use into the post-retrofit period, large variations in the savings occurred in certain buildings, particularly for the cooling energy savings in the Business Building. These variations appear to be due to the model's ability (or inability) to predict savings in buildings with significant on-off schedule characteristics and assumptions that the contestants made about the post-retrofit energy use, specifically, the periods for semester, nonsemester schedules that influence the on-off operation.

In general, all four models and the bin method predicted similar savings for the Engineering Center building. However, in the Business Building the savings predictions represented a smaller fraction of the annual energy costs and remained well within the noise of the models. It would appear that only contestants E2, E3, and the bin method made similar predictions of savings for the Business Building. The differences in the savings predictions at the Business Building also seem to indicate the importance of assessing the noise in the models and paying close attention to assumptions about schedules, etc.

One of the other findings that can be inferred from the second contest regards the accuracy of the models and their ability to predict savings. If one assumes that the model's prediction ability is indicated by the CV(RMSE), then only those savings for the Engineering Center and the Business Building that fall out of this range can be deemed as being larger than the inherent noise in the model. In the case of the Engineering Center, it appears as though all the models adequately predicted the savings.

However, in the Business Building the ratio of the estimated savings to the post-retrofit utility costs were considerably smaller than the models' CV(RMSE), which should indicate that the savings at the Business Building are more difficult to predict. This uncertainty in the models' predictions should raise a flag of caution for those who blindly invest in energy conservation without regard for how the savings are evaluated. Furthermore, it would seem to indicate that additional studies are needed to determine when it is necessary to use additional end-use metering and/or highly accurate models vs. more simplified analytical methods.

ACKNOWLEDGMENTS

The authors would like to recognize the significant input to this paper by Professor Jan Kreider at the University of Colorado. Also, thanks to the 50 persons who visited the anonymous ftp site where the Predictor Shootout II contest instructions and data were posted, especially the four entrants who successfully completed their predictions. These four entrants are listed below according to their model's goodness of fit:

- (Winner 1, Entry E4)—R. Dodier, G. Henze, Joint Center for Energy Management, Dept. of Civil Engineering, University of Colorado at Boulder, Boulder, CO 80309; Tel: 303-492-3915, e-mail: dodier@bechtel.colorado.edu, henze@bechtel.colorado.edu.
- (Winner 2, Entry E3)—S. Katipamula, Battelle Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352; Tel: (509) 372-4592, e-mail: s_katipamula@pnl.gov.
- (Winner 3, Entry E1)—Y. Chonan, K. Nishida, and T. Matsumoto, Department of Electrical Engineering, Waseda University, 3-4-1 Ohkubo Shinjuku-ku, Tokyo 169, Japan; Tel/Fax:01-3-3702-4735, e-mail: chonan@matsumoto.elec.waseda.ac.jp.
- (Winner 4, Entry E2)—K. Jang, E. Bartlett, and R. Nelson, Dept. of Mechanical Engineering, H.M. Black Engineering Bldg., Iowa State University, Ames, IA 50011; Tel: (515) 294-6886 or Fax: (515) 294-3261, e-mail: ronn@iastate.edu.

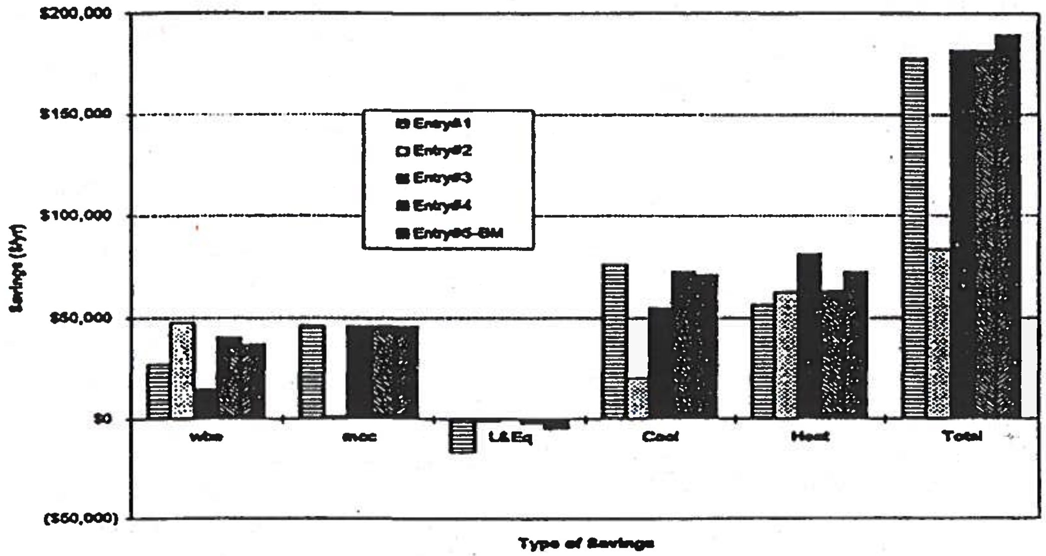

Figure 11 Annual savings comparisons for the Engineering Center for 1992.

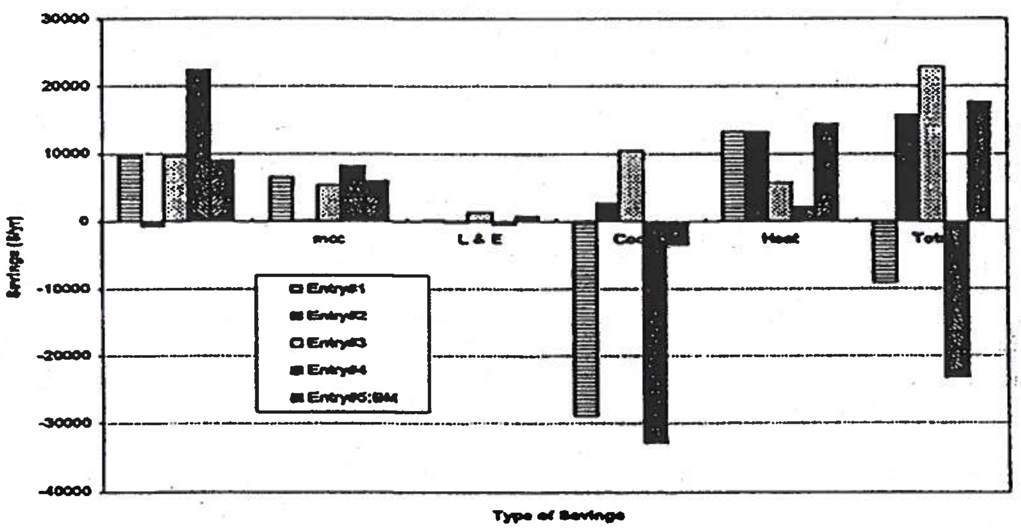

***Figure 12** Annual savings comparisons for the Business Building for 1992.*

Portions of this work were funded by the Texas State Energy Conservation Office as part of Texas A&M's LoanSTAR Monitoring and Analysis contract. 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, 1994). For further information and/or copies of the predictor data set, contact the authors.

REFERENCES

ASHRAE. 1994. Predicting hourly building energy use—The great energy predictor shootout, J. Haberl, ed. *ASHRAE Technical Data Bulletin* 10(5).

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 System.

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.

Chonan, Y., K. Nishida, and T. Matsumoto. 1996. A Bayesian nonlinear regression with multiple hyperparameters on the ASHRAE II time series data. *ASHRAE Transactions* 102(2).

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

Claridge, D., J. Haberl, M. Liu, J. Houcek, and A. Aather. 1994. Can you achieve 150% of predicted retrofit savings? Is it time for recommissioning? *Proceedings of the 1994 ACEEE Summer Study*, pp. 5.73-5.88 (August).

Dodier, R., and G. Henze. 1996. A statistical analysis of neural networks as applied to building energy prediction. *Proceedings of the 1996 ASME/JSME International Solar Energy Conference*, March.

Feuston, B.P., and J.H. Thurtell. 1994. Generalized nonlinear regressions with ensemble of neural nets: The great energy predictor shootout. *ASHRAE Transactions* 100(2): 1075-1080.

Haberl, J., D. Bronson, S. Hinchey, and D. O'Neal. 1993. Graphical tools to help calibrate the DOE-2 simulation program to non-weather-dependent measured loads. *ASHRAE Journal* 35(1): 27-32.

Haberl, J., D. Bronson, and D. O'Neal. 1995. An evaluation of the impact of using measured weather data versus TMY weather data in a DOE-2 simulation of an existing building in central Texas. *ASHRAE Transactions* 101(2).

Iijima, J., K. Takagi, R. Takeuchi, and T. Matsumoto. 1994. A piecewise-linear regression on the ASHRAE time series data. *ASHRAE Transactions* 100(2): 1088-1095.

Jang, K.-J., E. Bartlett, and R. Nelson. 1996. The great energy predictor shootout II: Measuring retrofit energy savings using autoassociative neural networks. *ASHRAE Transactions* 102(2).

Katipamula, S. 1996. The great energy predictor shootout II: Modeling energy use in large commercial buildings. *ASHRAE Transactions* 102(2).

Katipamula, S., and D. Claridge. 1992. Monitored air-handler performance and comparison with a simplified system model. *ASHRAE Transactions* 98(1): 341-351.

Kawashima, M. 1994. Artificial neural network backpropagation model with three-phase annealing developed for the building energy predictor shootout. *ASHRAE Transactions* 100(2): 1096-1103.

Kreider, J.F., and J.S. Haberl. 1994a. Predicting hourly building energy usage: The results of the 1993 great energy predictor shootout identify the most accurate method for making hourly energy use predictions. *ASHRAE Journal* 36(6): 72-81.

Kreider, J.F., and J.S. Haberl. 1994b. Predicting hourly building energy use: The great energy predictor shootout—Overview and discussion of results. *ASHRAE Transactions* 100(2): 1104-1118.

MacKay, D. 1994. Bayesian nonlinear modeling for the energy predictor competition. *ASHRAE Transactions* 100(2): 1053-1062.

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): 1063-1074.

Stevenson, W.J. 1994. Predicting building energy parameters using artificial neural nets. *ASHRAE Transactions* 100(2): 1081-1087.

Thamilseran, S., and J. Haberl. 1995. A bin method for calculating energy conservation retrofit savings in commercial buildings. *Proceedings of the ASME/JSME/JSES International Solar Energy Conference*, Maui, Hawaii, pp. 111-123.

APPENDIX

The evaluation was performed using two indicators—the coefficient of variation of the root-mean-square error (CV-RMSE) and the mean bias error (MBE). These were the statistics that were used in Predictor Shootout I (Kreider and Haberl 1994a, 1994b) with the exception of parameter p in the definition of CV (RMSE) and MBE, which indicates the total number of regression parameters in the model. For the purpose of this evaluation, this parameter was assigned an arbitrary value of 1. The definitions of these two statistics are given below.

Coefficient of Variation CV (%):

$$\mathrm{CV(RMSE)} = \frac{\sqrt{\dfrac{\sum_{i=1}^{n} (y_{pred,i} - y_{data,i})^2}{n-p}}}{\bar{y}_{data}} \times 100$$

Mean Bias Error, MBE (%):

$$\mathrm{MBE} = \frac{\dfrac{\sum_{i=1}^{n} (y_{pred,i} - y_{data,i})}{n-p}}{\bar{y}_{data}} \times 100$$

where

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

$y_{pred,i}$ = predicted dependent variable value for the same set of independent variables above,

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

n = number of data points in the data set, and

p = total number of regression parameters in the model (which was arbitrarily assigned as 1 for all models).