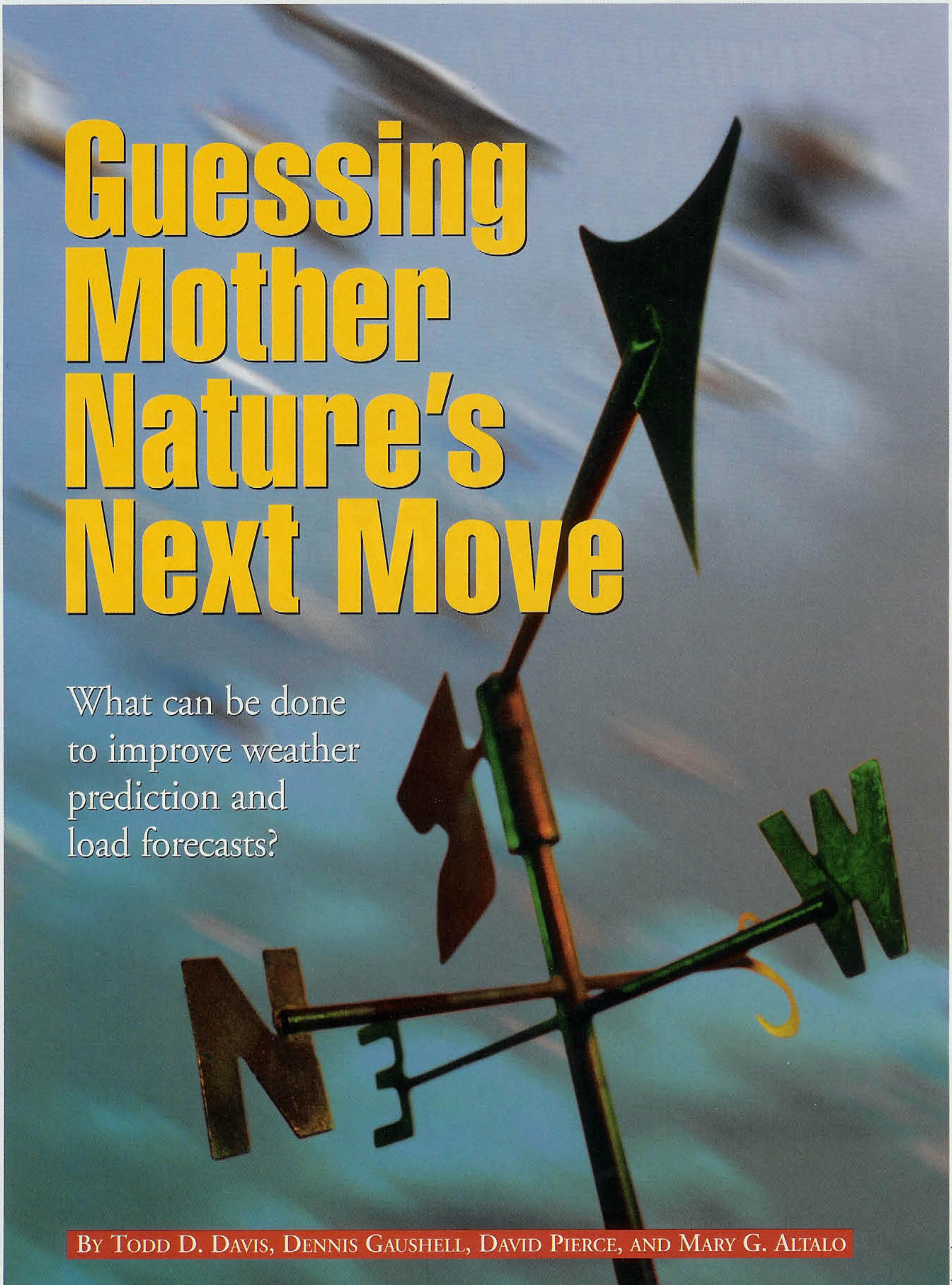
Guessing Mother Nature's Next Move

What can be done to improve weather prediction and load forecasts?



By Todd D. Davis, Dennis Gaushell, David Pierce, and Mary G. Altalo

Illustration by Alex Stephen

Weather forecast error, whether large or small over a year, can result in substantial added costs to energy organizations. With today's movement toward real-time pricing and sub-hourly markets, it is critical to be as accurate as possible with load forecasts.

Short-term load forecast error (*i.e.*, percent variations high and low within a four-day period) typically is caused by load measurement error, weather forecast error, and load forecast model error.

Benefits of improving the day-ahead weather and load forecast by just 1-degree Fahrenheit include:

- About $20 million to $25 million per year for a Northeast regional transmission authority; and
- About $1 million to $2 million per year for a large regional distribution company.

Con Edison reported that its earnings per share in 2003 vs. 2002 were positively affected by $0.15 per share. Earlier it reported -$.05 earnings per share due to weather effects—demonstrating that over a short period of time, earnings can swing from -$11 million to +$36 million—a $47 million shift.

NOAA Research: Looking for Diamonds in the Data

What can be done to reduce forecast error, and what are the costs/benefits in making these improvements?[1] NOAA sponsored three projects conducted by SAIC that benchmarked the accuracy of energy organization weather and load forecasts and estimated the value of a 1-degree improvement in forecast accuracy on selected energy enterprise operations. The case study covered two regional transmission organizations interested in short-term (*i.e.*, less than five days) forecasting issues and two distribution utilities providing both natural gas and electricity. After discovering the source of the error and the cost, NOAA determined the economic value of reducing weather and load forecast error.

Effects of the California Delta Breeze

Not many energy systems have a weather phenomenon that can result in a 4,000-MW shift in a very short period of time. In summer, Northwest winds, west of the Pacific coastline, are drawn into the interior through the Golden Gate and over the lower portions of the San Francisco Peninsula. Immediately to the south of Mount Tamalpais, the winds accelerate considerably and come more nearly from the West as they stream through the Golden Gate (*see Figure 1*). This channeling of the flow produces a jet that sweeps eastward but widens downstream, producing southwest winds at Berkeley and northwest winds at San Jose; a branch curves eastward through the Carquinez Straits and into the Central Valley.

Figure 2 shows the possible error that may occur when a delta breeze abruptly stops and weather models underforecast temperature, causing load shortfalls. The delta breeze is strongly influenced by large-scale synoptic weather patterns that move into California from the North Pacific Ocean.

The analysis attempted to build a better predictive model of the delta breeze. Statistical analysis of observed weather data from more than100 weather stations in California during the latest 10 years showed that the breeze indicates large-scale winds that affect both Northern and Southern California. Furthermore, the analysis showed that accurate forecasts of the delta breeze required a large-scale weather condition: the North Pacific High.

The statistical forecast model developed by Scripps for delta breeze uses three variables: latest-hour wind speed/direction at Fairfield, previous hour's wind speed/direction at Fairfield, and latest pressure gradient between Merced, Calif., and Portland, Ore. (measuring the strength of North Pacific High). The accuracy of the statistical model was compared with the accuracy of the National Weather Service (NWS) GFS MOS model for Fairfield. The resulting forecast accuracies are shown in Table 1.

The statistical model is much more accurate for today's forecast (made at 7 a.m.). However, for tomorrow's forecast (made at 11 a.m.), the statistical and NWS models are virtually the same. Therefore, a combination of statistical model and NWS model can be used for optimum accuracy.

As shown in Figure 2, the peak load was underforecast by Cal-ISO by 4,724 MW. In addition, the load-serving entities had a similar total underforecast, which resulted in their underscheduling of load by an even greater amount. The underlying cause of the underforecast was the failure to forecast the cessation of the delta breeze. This resulted in a major increase in prices from a typical average of $45/MWh to $110/MWh during the peak 12 hours.

Additional costs were a minimum of $1 million.

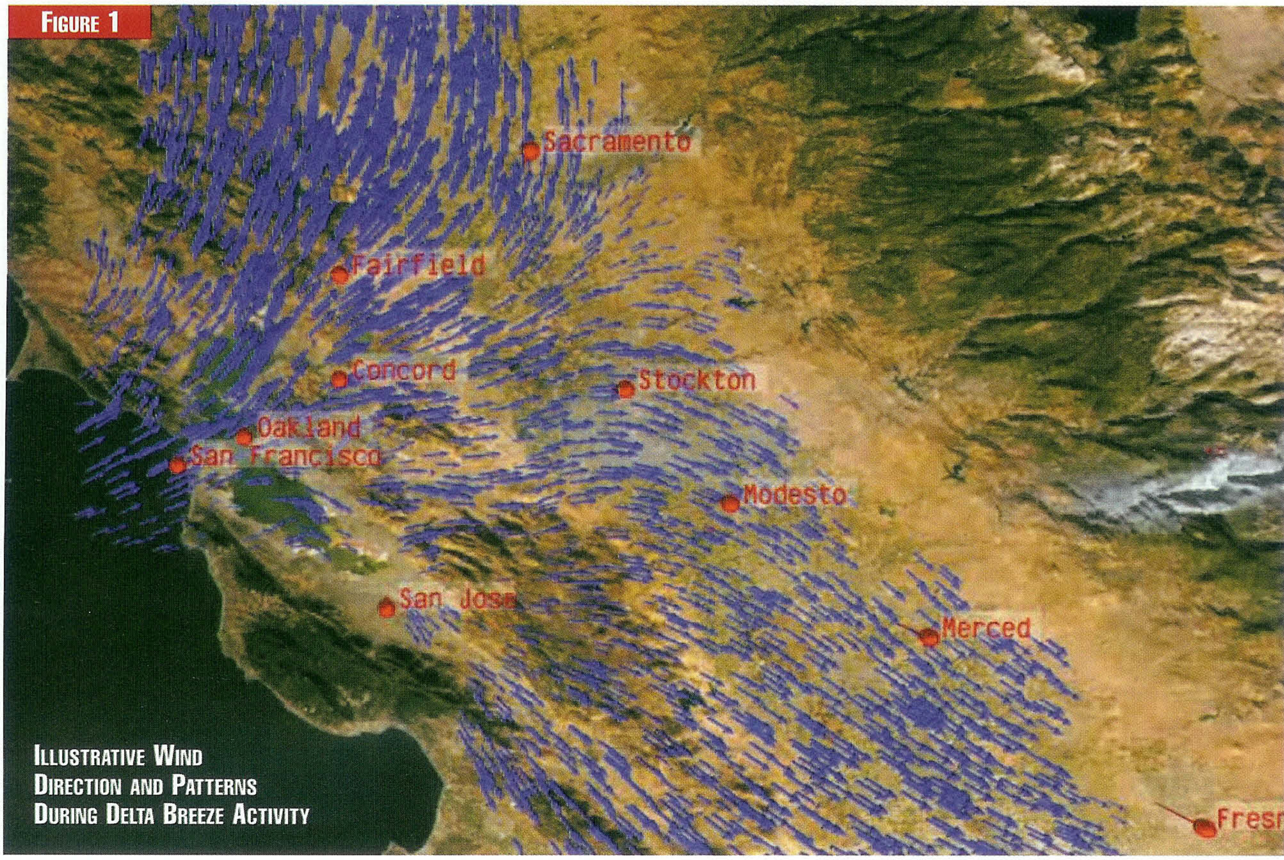
New England ISO: Relying on Biased Weather Stations?

NOAA also investigated a weather-load forecast improvement

TABLE 1 COMPARISON OF STATISTICAL AND NWS FORECAST MODELS FOR DELTA BREEZE

	Delta Breeze Forecast Accuracy		Non Delta Breeze Forecast Accuracy	
Forecast For	Statistical	NWS GFS MOS	Statistical	NWS GFS MOS
Today	94	84	97	73
Tomorrow	73	75	70	67

FIGURE 1



ILLUSTRATIVE WIND DIRECTION AND PATTERNS DURING DELTA BREEZE ACTIVITY

Source: Dr. David Pierce, Scripps Institution of Oceanography, La Jolla, California

opportunity with ISO New England (ISO-NE). The NOAA Northeast Energy Project aimed to:

- Explore improvements in the ISO-NE weather station selection and weather station weighting in the ISO-NE ANNSTLF model calibration;
- Recommend ways to improve the accuracy of weather forecasts during extreme peak-load weather periods when the value to ISO-NE is the highest;
- Recommend ways to improve the forecast in time and periods ranging from the day-ahead gas market to periods of four days.

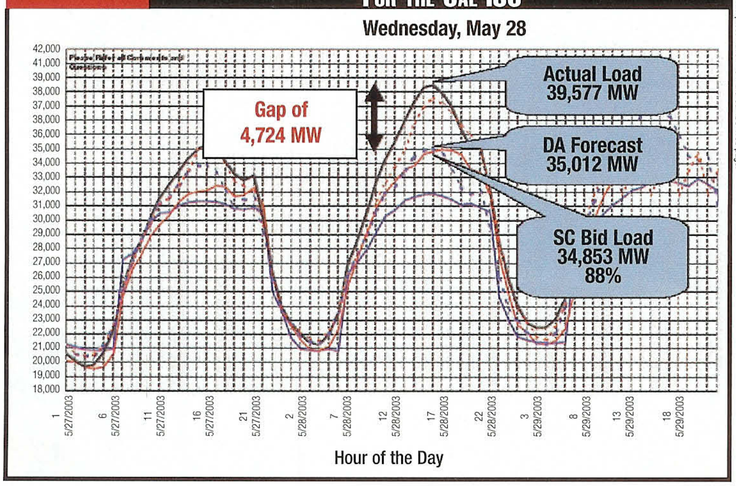
ISO-NE provided a list of parameters (temperature, relative humidity, and wind speed) used in the load model and the formula for combining the parameters into indexes for load model input.

The methodology included using three data sets for:

1. Identifying the actual system load.
2. Estimating the perfect load forecast. The actual pseudo-temperatures were calculated for the system to simulate the perfect forecast (actual = forecast). The actual system pseudo-temperatures used as input for the perfect forecast load were calculated by ISO-NE.
3. Estimating the day-ahead hourly forecast.

AVN MOS (00Z) hourly forecasts of temperature, relative humidity, and wind speed were combined for eight New England weather stations (BDL, BDR, BOS, BTV, CON, ORH, PVD, PWM) using the ISO-provided formulas to create a systemwide pseudo-temperature forecast. EarthSat prepared the system hourly forecast, and ISO-NE ran it through the

FIGURE 2 **UNDER FORECAST OF MAY 28, 2003, DELTA BREEZE EVENT FOR THE CAL-ISO**



Source: Cal-ISO presentation from the Cal-ISO Web page.

ANNSTLF load model. Load data for all three data sets were calculated/observed on an hourly basis for the year 2002. Data was then aggregated by day to simplify the comparison of data. To isolate the component of the total error associated with the ANV MOS forecast, the total load error was partitioned into forecast-related error—or the error stemming from the accuracy of the previous day system-wide hourly pseudo-temperature forecast—and the model-related error, which contains errors associated with the fundamental capabilities of the ANNSTLF model itself and included the selection of stations used to develop systemwide forecasts, and stations weight determination (*see Fig. 3*).

The conclusion show:

- Forecasting error accounts, on average, for less than 40 percent of the total load forecast error. The majority of the error appears to lie within the models' load-assessment capabilities;
- The forecast component of the total load error and the total load error exhibit minimal correlation ($r^2 = 0.1900$). This means that under-forecasting the temperature profile does not necessarily imply under-forecasting the load, and vice-versa; and
- The model component of the total load error and the total load error possessed a high degree of correlation ($r^2 = 0.8564$) (*see Figure 4*).

Figure 5 provides a probability distribution histogram for the fraction of absolute forecast error amongst the total absolute error. The figure shows that AVN MOS absolute load forecast error in this analysis is approximately 40 percent of the total absolute load error.

Two model improvement runs then were conducted. The first model improvement run tested the improvement of the load model given a 1-degree improvement in temperature forecast. Because the ISO-NE load model uses an index combining dry bulb temperature and relative humidity in the summer and dry bulb temperature and wind speed, the improvement run was actually conducted using a one-increment improvement in the index. It also should be noted that on hours when the original index was equal to the observed, no improvement was possible. Thus, the total improvement over the course of the run averaged 0.7 rather than 1.0 incremental index units. Since temperature is the major contributor to the index, this approximation should not negate the temperature improvement objective.

FIGURE 3 **ISO-NE SYSTEM LOAD COMPARISONS**

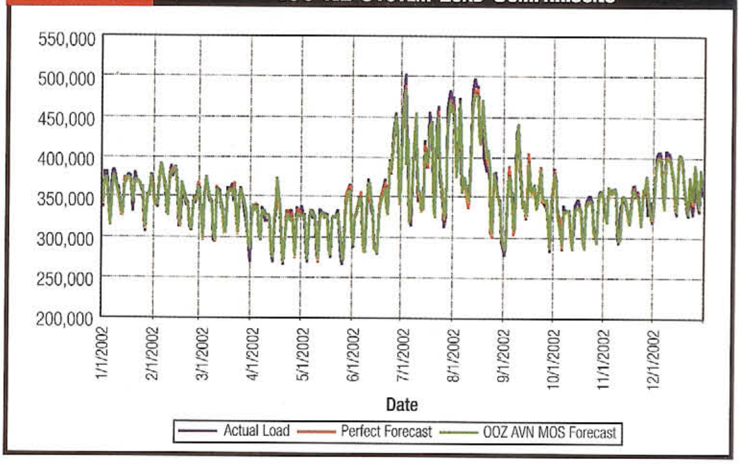


FIGURE 4 **DAILY TOTAL LOAD TRACE FOR ACTUAL LOAD, PERFECTLY FORECASTED LOAD, AND AVN MOS FORECASTED LOAD.**

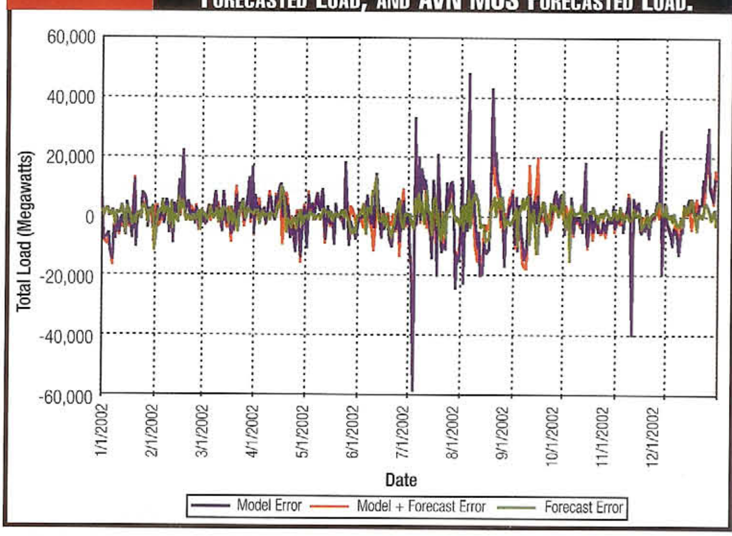
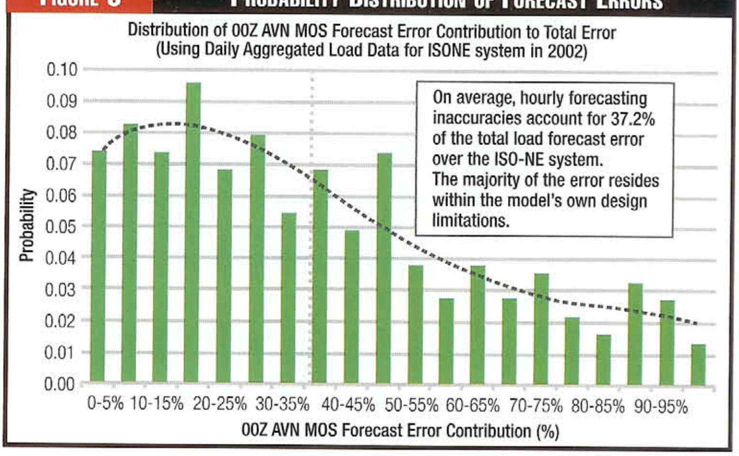


FIGURE 5 **PROBABILITY DISTRIBUTION OF FORECAST ERRORS**



Source: The SAIC Northeast Energy Network Performance Analysis Report to NOAA (with contributions from ISO NE, EarthSat and AWS)

The second improvement run was conducted substituting Bedford for Boston in an attempt to indicate the potential model improvement of a more representative station.

Mean Absolute Errors (MAE) were calculated for each of the model runs and compared to the baseline and perfect model-run MAE. These results are shown in Table 2.

The Bedford replacement of Boston showed virtually no improvement. The 6.41 percent improvement resulting from the 0.7-incremental index improvement was somewhat less than initially anticipated.

To evaluate the improvement that resulted in the high error cases when temperatures usually were outside the norm, an error analysis was conducted for the top 30 load-model error days in the data set. As shown in Table 2, most of the high-error days occurred during the summer, when power demand is high in the ISO-NE service region.

The results of the top 30 error days analysis is shown in Table 3.

The Bedford replacement of Boston still showed very little improvement, indicating this change in the model would not be useful. This analysis does indicate an increasing importance of accurate forecasts in high energy usage days because the perfect forecast improvement and the 0.7 incremental index forecast improvement more than doubled in the top 30 error day analysis compared to the error for all cases.

California ISO: A Western View Of WeatherStations

Cal-ISO took a similar approach to the selection of improved weather stations. For the Los Angeles area, the existing model used Los Angeles airport, Orange County airport, L.A. Civic Center, Ontario, and Riverside. The regression of weighted average temperature from these stations vs. load had an R-squared of 0.75 (1.0 is perfect correlation).

Analysis of individual stations showed that the two coastal stations (L.A. airport and Orange County airport) strongly were influenced by local sea breezes. However, these local sea breezes generally did not penetrate far enough inland to affect load significantly. This led to development of a new model using L.A. Civic Center, Fullerton airport (8 miles inland), Ontario, Lancaster, and Palms Springs. These latter two stations are far inland and represent high area air-conditioning loads. Riverside was eliminated because its temperatures were very highly correlated with Ontario and thus did not add any new information. The weighted average temperature from these stations had an R-squared of 0.90 vs. load. This approach, applied to all weather stations in Cal-ISO load areas, resulted in a 16 percent drop in large (>5 percent of load) hourly load forecast errors from 2002 to 2003. This result shows the extreme importance of proper weather station selection and weather station weighting.

Table 2 Improvements in ISO-NE Load Model.

All days – 2002		
Run	MAE (MW)	Improvement
Baseline	6089.2	-
Perfect Forecast	5365.4	11.89%
Bedford Swap	6082.9	0.12%
One-Degree Improvement	5699.1	6.41%

Improvement is defined as the percentage decrease in load MAE over the baseline

Table 3 Improvements in ISO-NE Load Model for Perfect Forecast, Bedford Swap, and One-Degree Improved Forecast for 30 Error days.

Top 30 Forecast Error Days		
Run	MAE (MW)	Improvement
Baseline	10042.9	-
Perfect Forecast	7491.5	25.40%
Bedford Swap	9914.5	1.71%
1-Degree Improvement	8480.6	15.76%

Improvement is defined as the percentage decrease in load MAE over the baseline

Weather Event Analysis

A case-by-case analysis on the synoptic weather events in the ISO-NE region was completed to identify the standard synoptic situations causing the high weather error cases. As would be expected, the high error cases occurred when there were active weather events in the area. In three of the cases, the load-model errors could not be readily explained. In the remaining 27 cases, the load-model errors could be explained by weather data forecast errors associated with reoccurring weather situations. The key synoptic weather features that were noted included:

- A frontal boundary in the area that was moving slower or faster than expected;
- Easterly or Northeasterly maritime wind influences resulting in temperatures below those that were forecast;
- Strong westerly winds flowing downhill from the mountains to the coastal plain in the high-weighted population areas around Boston, Mass., and Providence, R.I. The westerly flow resulted in compression and provided higher surface temperatures than forecast; and
- Unexpected afternoon thundershowers resulting in temperatures lower than forecast (*see summary in Table 4*).

A review of the forecast errors that were identified in the summer of 2002 and the associated market prices at the time of the error show significant forecast errors during periods of high market prices. This information shows an opportunity for marketers to pinpoint key load and temperature events along with the need for ISO-NE to work to reduce load forecast error during critical weather synoptic events. A comparison of the baseline forecast with the perfect forecast indicated that the weather forecast component in the model was responsible only for approximately 40 percent of the overall total error. A subsequent test of the highest 30 days of error revealed that when load error was high, forecast weather was responsible for a substantially larger percentage of the load-model error. Future research must determine the reason for the other 60 percent of the load-model error that remains unexplained.

The index forecast improvement provided a marginal improvement of 6.41 percent for the entire model run. This improvement more than doubled to 15.76 percent when the analysis was conducted on the 30 largest error days. These conclusions indicate the potential value of improving weather forecast on the load model forecast.

Table 5 shows the incidence of average and extreme forecasts errors in the summer period from May through August of 2002. Of the 325 total hours during the peak periods of 1 p.m. to 5 p.m., close to one-third of the forecast error exceeded 3 percent. Some forecast errors were as high as 15 percent during critical peak days.

It is very difficult to estimate ISO-NE's cost or a true social cost for this error. However, during some of the severe error days, and one that even included an outage of a major base-load plants that resulted in expensive imports from Quebec, costs for this power approached $1,000/MWh.

TABLE 4 — COMMON EVENTS FOR HIGH FORECAST ERRORS IN ISO-NE SERVICE REGION

Weather Type Event	Forecast Error Direction	Number Of Events
Frontal boundary	Either high or low	10
Maritime flow	Forecast too high	6
Strong west winds	Forecast too low	6
Unforecast afternoon showers	Forecast too high	5
Unexplained	Either high or low	3

TABLE 5 — INCIDENCE OF FORECAST ERROR (MAY–AUGUST 2002)

Error Brand	May	June	July	August	Total
0-1%	24	12	13	6	55
-1.01 to -2%	11	8	7	13	39
-2.01 to -3	9	8	1	10	28
0-1%	6	10	4	2	22
> -3	8	4	8	7	27
1.01-2.0%	12	11	3	7	33

How Accurate?

At Cal-ISO, the accuracy of 13 commercial and public weather forecasting services was evaluated during the heavy load period from July through September of 2003. The accuracy was evaluated in terms of impact on electricity load forecasting; for example, the impact of a 1-degree error at 75 degrees is far less than a one degree error at 95 degrees. Therefore, average absolute megawatt error (at the time of highest daily temperature) was utilized as the measurement of forecast accuracy. Impacts of the load forecast model were removed, so that only weather forecast errors were considered.

The results of Table 6 showed a wide variation in error in both commercial services (730 MW to 3,397 MW average absolute error) and public services (616 MW to 922 MW average absolute error). In addition to the comparison below, several consensus forecasts of the best performers were evaluated, in which forecasts were combined by averaging. However, the consensus forecast accuracy was worse than the individual forecasts.

Several conclusions can be drawn from this analysis: 1) Due to the wide variation in forecast accuracy, forecasting services should be evaluated before selection for electricity-load forecasting; 2) Weather forecast accuracy should be continuously monitored and services improved or changed if necessary; and 3) Consensus forecasts apparently do not improve forecast accuracy.

Moreover, the commercial forecasters do not necessarily provide higher accuracy than public forecasts; however, they do provide many value-added services, such as hourly forecasts (instead of every 6 hours), data formatting for direct computer input, and hourly forecast resolution instead of public 3-hour resolution.

By analyzing the forecast errors on a daily basis, the Cal-ISO also found that most of the commercial and public forecasts consistently under-forecast major temperature warm-up events by a few degrees on the daily maximum. These few degrees are critical to electricity load forecasting. Therefore, the Cal-ISO built a "temperature spike predictor," which compensated for this deficiency. Used for the summer of 2004, this temperature forecast, combined with the forecast from our commercial forecaster, resulted in a significant improvement. The Cal-ISO published day-ahead load forecast accuracy (in terms of mean absolute percent error) dropped from 1.91 for 2003 to 1.66 for 2004, and large errors dropped another 19 percent. These overall accuracy figures include model error, weather error, and pumping error.

In conclusion, weather forecast error is significant and has costs to suppliers, transporters, and consumers. While practitioners take solace that their forecasts errors average in the range of 1 to 2 percent per year, these errors do add up to significant

Table 6 Comparison of Weather Forecast Services For Summer 2003 at Cal-ISO		
Weather Service Average Absolute MW Error		
Forecaster (11 a.m. DA Unless Marked)	July 10-Sept. 22	Aug. 15-Sept.22
NWS– AVN MOS	616	600
Commercial Forecaster A	730	671
Commercial Forecaster B	750	659
Commercial Forecaster B (7 a.m.)	784	745
Commercial Forecaster C	845	735
Commercial Forecaster D 7 a.m.	846	832
NWS – MRF MOS	846	857
Commercial Forecaster E	922	895
NWS – ETA MOS	922	955
Commercial Forecaster F	971	875
Commercial Forecaster G	1,135	1,187
Commercial Forecaster H	2,535	2,289
Commercial Forecaster I	3,397	3,497

costs over a year. In addition, extreme weather events can occur, as they do in most regions of the United States, and this often occurs during very high cost periods. Managing weather and load forecast error both of a small nature and for those larger events is very cost-effective. How cost-effective depends on the individual power system, weather phenomena, the load, and modeling tools used.

Given that load imbalances and congestion accumulate significant costs in competitive power markets, more and more attention will be devoted to benchmarking the errors and finding more advanced solutions to reduce their magnitude. Clearly, more accurate weather forecasting can make a significant improvement in load forecasting, as well as reduction of costs.

Todd D. Davis is an assistant vice president at SAIC, Dennis Gaushell is a consultant at the Cal-ISO, David Pierce is a staff scientist at Scripps Institution of Oceanography, and Mary G. Altalo is chief scientist, energy solutions, at SAIC.

Endnote

1. NOAA, The Northeast Energy Network Performance Analysis Report: Final Report (August 2003); NOAA, The Use of 30 Year Climate Forecasts To Improve Regional Long-Range Energy Infrastructure Planning for San Diego County); NOAA, The Economic Benefits of Incorporating Weather and Climate Forecasts into Western Energy Production Management.

* The work by Dennis Gaushell (*www.consultingforenergy.com*) was performed for the California ISO. The authors acknowledge the contribution of L. Heitkemper of EarthSat, Jan Dutton of AWS, and Alexander Gershanov of the Scripps Institution of Oceanography.

Scarcity

(Continued from p. 46)

Demand Outlook, March 2005, p. 4.

15. Since the ISO submission is unavailable, it is logical to assume that the forced outages were incorrectly included in the WECC tabulation. If so, this is simply a problem in the ISO's approach to the question, and not an actual adding error.
16. Summer 2005 Electricity Supply and Demand Outlook, CEC, March 2005, pp. 10-11.
17. Summer 2005 Electricity Supply and Demand Outlook, CEC, March 2005, p. 5.
18. Again, this is sorting the table back into the traditional reliability planning format—adding all resources and imports and then comparing that total to the 1-in-2 load forecast.
19. As noted above, the complex structure of the Cal-ISO has rewarded fraudulent outage reports in the past. Since the incentives for plant maintenance outside of the ISO are relatively simple (if the plant isn't running its owner bears the costs) and complex within the ISO (if the plant isn't running, its owner may actually be able to negotiate higher payments) most analysts would forecast lower outage rates at LADWP rather than higher outage rates.
20. Summer 2005 Electricity Supply and Demand Outlook, CEC, March 2005, p. 7.
21. 2005 Summer Operations Assessment, California Independent System Operator, March 23, 2005, p. 35.
22. Like many rules of thumb in the electric industry, this was selected on the basis of experience and has proven acceptable since it was adopted in the 1960s. For an "oral history" of this value, see Merrill Schultz's comments to the WECC as reported in *California Energy Markets*, March 11, 2005, p. 5.
23. The CEC's report assumed 9,903 MW while the ISO assumed 9,700 MW. Documentation for the 9,903 MW assumed by the CEC is on p. 12.
24. 10-Year Coordinated Plan Summary, WECC, September 2004, p. 28. Again, these estimates are made for "Adverse Hydro Conditions" as noted twice on this page.
25. Ibid., pp. 36 and 44. Again, these estimates are made for "Adverse Hydro Conditions" as noted four times on these two pages.
26. 2005 Summer Operations Assessment, California ISO, March 23, 2005, p. 26.
27. The ISO's SCIT estimate would also be required to accommodate "dynamic" resources to some degree. In either case, it seems unlikely that southwest imports would be blocked by transmission limitations.

Correction: On p. 60 of the June 2005 issue, in Table 3 accompanying the article *"CEO Pay Reflects Strong Stock Performance in 2004"* by Edward Metz of SNL Energy, Otter Tail Corp. CEO John D. Erickson's total compensation was overstated. Erickson would not have appeared in this table had John D. Erickson's compensation been correctly stated. Additionally, Entergy CEO J. Wayne Leonard did not exercise any stock options in 2004 and should not have appeared in Table 1. SNL Energy regrets the error and any confusion this may have caused.