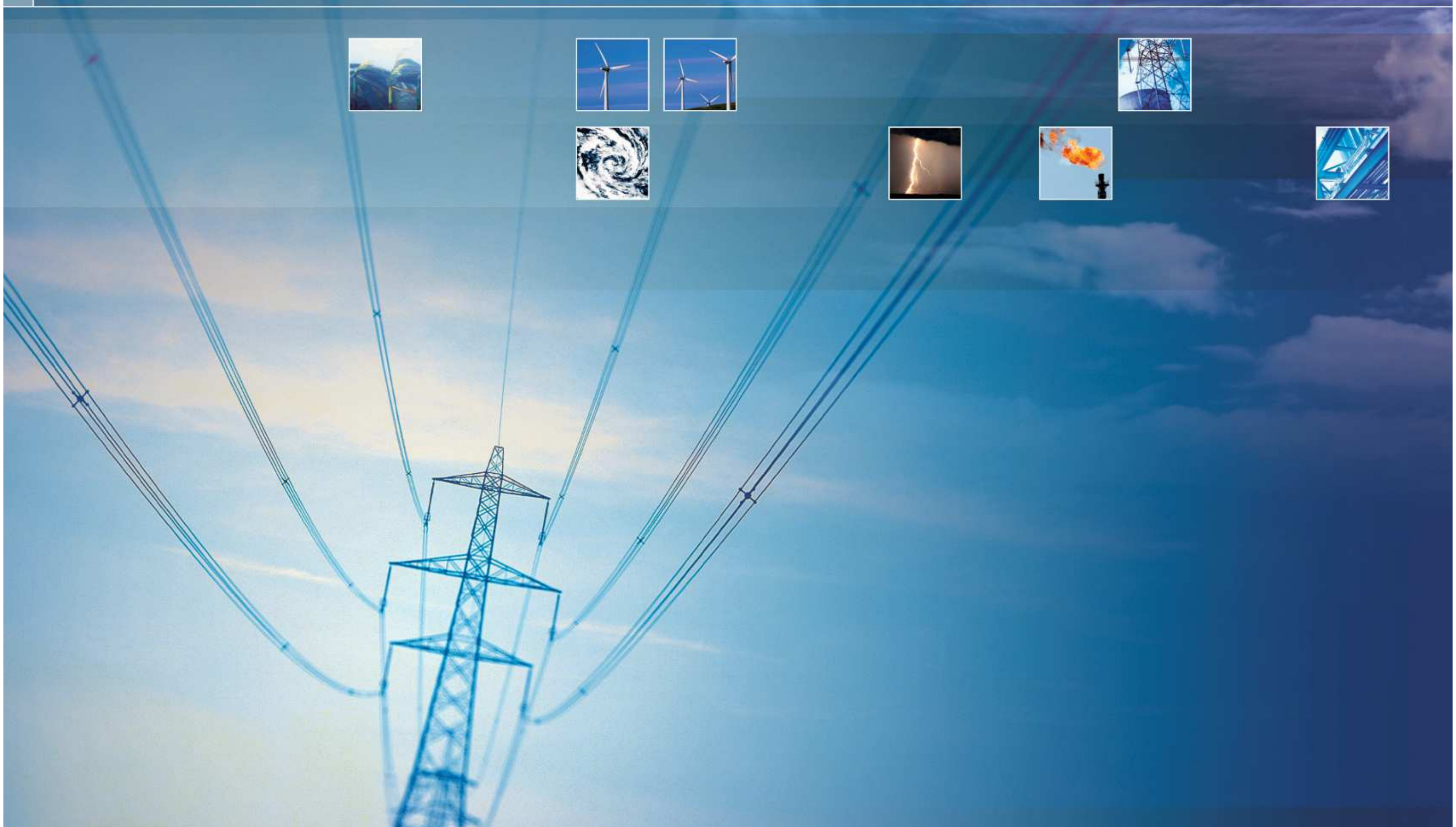
SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

The Economic Benefit of Incorporating Weather and Climate Forecasts Into Western Energy Production Management

Deliverable 5



Submitted To:

Scripps Institution of Oceanography (SIO)
La Jolla, CA

July 15, 2004

SAIC®
An Employee-Owned Company

FINAL REPORT

The Economic Benefit of Incorporating Weather and Climate Forecasts Into Western Energy Production Management

Deliverable 5

FINAL REPORT

Rescinded version

July 15, 2004

A Project Sponsored By the National Oceanographic and Atmospheric Administration (NOAA)

This Page Intentionally Left Blank

Contributors and Acknowledgements

This report is sponsored by the National Oceanographic and Atmospheric Administration (NOAA) under contract no.: P.O. 1022-8562

Key contributors to the report include:

Scripps Institution of Oceanography

- Dr. Tim Barnett
- Dr. Eric J. Alfaro
- Dr. Alexandar Gershunov
- Dr. David Pierce
- Dr. Anne Steinnemann

Science Applications International Corporation

- Dr. Mary G. Altalo
- Mr. Todd D. Davis
- Ms. Monica Hale
- Dr. Lenny Smith

PacifiCorp

- Mr. Reed Davis

The authors deeply appreciate the support, cooperation and contributions of the Cal ISO, PacifiCorp and the National Oceanographic and Atmospheric Administration (NOAA).

This Page Intentionally Left Blank

Table of Contents

List of Figures and Tables

Page

Figures

Tables

List of Figures and Tables (continued)

Page

Tables

1. Climate Forecasting of Irrigation Power Loads in Climate Division 9 and The Economic Value to PacifiCorp

1.1 Introduction and Overview[1]

The objective of this case study is to demonstrate the scientific and economic value of a climate derived irrigation load forecast for a section of the PacifiCorp service territory and the economic value in making such a forecast. Predicting irrigation pump loads is a critical seasonal summer forecasting issue in order to anticipate when the pumps will be turned on in the spring or early summer. There are considerable economic advantages in predicting in advance the timing and duration when irrigation will commence as this has significant impacts on the costs to PacifiCorp in terms of the amount of peak demand that should be planed for. In addition, the supply scheduling and timing of the load is also important. Generally, the more accurate such predictions are in terms of timing and duration, the more economical power supply purchases will be and the greater the opportunity for PacifiCorp to plan for grid power flow and disposition of the unused energy on its system.

Historically, PacifiCorp has not done any forecasting of irrigation pump loads. From a climate point of view, many factors contribute to the timing of the onset of load demands, including soil moisture value, the crops planted, seasonal precipitation, etc. This report relies on a climate and statistical regression model for predicting when irrigation load will occur. The economic value of predicting this load is also presented.

1.2 PacifiCorp Service Area

PacifiCorp service area extends for 136,000 square miles with 15,000 miles of transmission, 44,000 miles of overhead distribution and 12,000 miles of underground cable. PacifiCorp's 53 hydropower facilities are located in Washington, Oregon, Idaho, Utah and Montana. With a total generating capacity of 1,078 megawatts (mw) of electricity,[2] PacifiCorp has more than 8,300 megawatts of generation capacity from coal, hydro, renewable wind power, gas-fired combustion turbines and geothermal and serving more than 1.5 million customers. PacifiCorp operates as Pacific Power in Oregon, Washington, Wyoming and California; and as Utah Power in Utah and Idaho (**figure 1-1**).

Figure 1-1. Map of PacifiCorp's Service Area

Source: PacifiCorp

1.3 Data

The data used for irrigation loads are from PacifiCorp internal records over the period 1997-2003. One of the limitations of the analysis is the short record of irrigation pump load data. The air surface temperature data used was from two

[1] This section is derived from an earlier technical paper entitled, "Some relationships between the irrigation pumping loads and the local climate in Climate Division 9, Idaho", written by Eric J. Alfaro, A. Gershunov, David Pierce, and Anne Steinnemann of the Scripps Institution of Oceanography, with selected edited contributions by Science Applications International Corporation (SAIC).

[2] Meeting with PacifiCorp staff, including Mr. R. Davis

sources: a) Idaho's Climate Division #9, Upper Snake River Plains, (Record: 1932 – 2003, Lat: 43.83° N, Lon: 112.82° W) and b) station Idaho Falls 46W, (Record: April 1st, 1954 – December 31st, 2001, ID: 104460, Lat: 43.52° N, Lon: 112.93° W, Elev: 1505 masl) from the National Climatic Data Center (NCDC) first order and cooperative observed data (NCDC, 2003). Soil moisture data was also analyzed for the same climate division (accessed at ftp://ftp.ncep.noaa.gov/pub/cpc/wd51jh). Soil moisture is estimated by a one-layer hydrological model (for details about the soil moisture data see Huang et al., 1996, van den Dool et al., 2003).

The model uses this equation for water balance: dw/dt=P-E-R-G, where the equation is applied locally and w is the soil moisture in a single column of 1.6 m depth, P is precipitation, E is evaporation, R is runoff and G is loss of ground water. The model parameters are estimated using observed precipitation, temperature and runoff in Oklahoma (1960-1989) and then applied to the entire United States. The calculations are done using monthly data for 344 U.S. Climate Divisions during 1932 to present. The NOAA-Climate Prediction Center monitors U.S. soil moisture using this model (http://www.cpc.ncep.noaa.gov/soilmst/; downloadable at: ftp://ftp.ncep.noaa.gov/pub/cpc/wd51jh.)

Calculations were also carried out using PDSI data with very similar results. The Palmer Drought Severity Index (PDSI) is a standard way to quantify the severity of drought conditions using a supply and demand model for the amount of moisture in the soil. The value of the PDSI is reflective of the how the soil moisture compares with normal conditions. A given PDSI value is usually a combination of the current conditions and the previous PDSI value, so the PDSI also reflects the progression of trends, whether it is a drought or a wet spell. That means that a single PDSI value is not representative of just the current conditions, but also of recent conditions.

The combination of various soil moisture conditions, precipitation, and non-climate factors (such as the mix of planted crops and the cost of irrigation) contribute to large year to year variability in pump loads, as is shown from PacifiCorp data in Figure 1-2. Our primary goal is to predict the average summer season pump load.

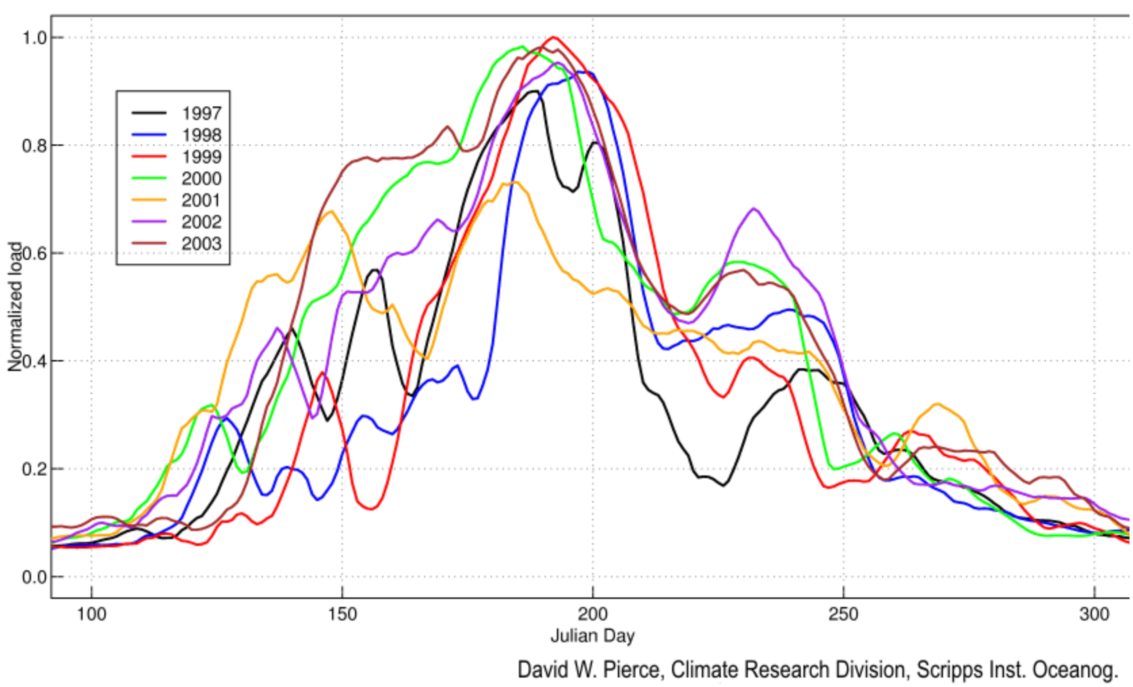


Figure 1-2. Normalized Irrigation Pump Loads Verses Day of the Year From Idaho Falls ID (Data from PacifiCorp)

1.4 Results

Several authors have described the relationship between the air surface temperature and soil moisture and also its potential use in prediction (e. g. Huang et al., 1996; van den Dool et al., 2003). **Figure 1-3** shows the annual time series for soil moisture and the mean temperature in climate division 9, Idaho for different seasons. There are negative correlations between previous (panel b) and simultaneous (panel a) soil moisture values and May, June, July, and August (MJJA) mean air temperature (Tmean; **figures 1-3a** and **1-3b**). There is a positive correlation between February–March (FM) and MJJA soil moisture (**figure 1-3c**).

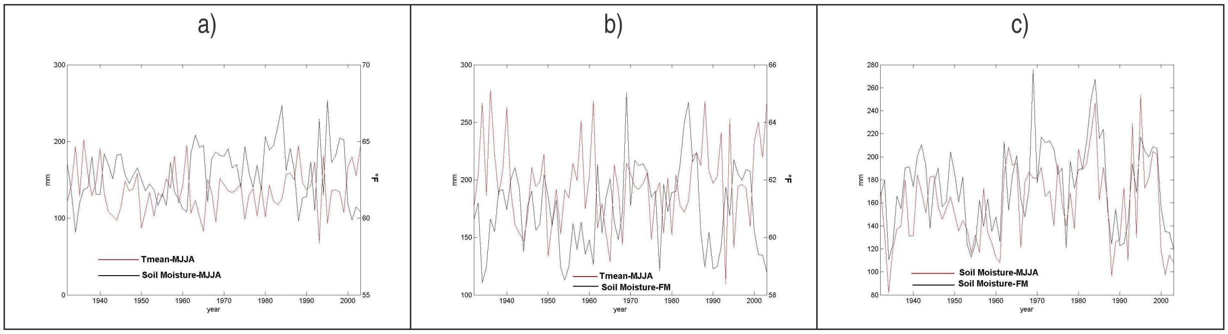


Figure 1-3. Time Series Plotting for Soil Moisture and Mean Temperature in Climate Division 9, Idaho for Different Seasons. *a) MJJA-Soil Moisture and MJJA-Tmean, r = -0.64. b) MJJA- Soil Moisture and MJJA-Tmean, r = -0.32. c) FM and MJJA Soil Moisture, r = 0.70. All the correlation values have statistical significance greater than the 95% level.*

Table 1-1 shows the contingency analysis between soil moisture and Tmean in Idaho climate division 9. Panel a) shows a prediction of summer temperatures given spring soil moisture conditions. The most likely scenario given below (above) normal FM-soil moisture conditions is above (below) normal MJJA-Tmean conditions (see the upper right and lower left values in **table 1-1a**). This relationship can be understood by examining **table 1-1**, panels b) and c). Panel b) shows that there is a strong simultaneous relationship between dry soil conditions and hot temperatures, while panel c) shows that dry spring conditions tend to persist through mid summer. Taken together, the physical picture indicated is that dry spring conditions persist to mid summer, at which time the dry land cannot moderate temperatures on extreme summer hot days. Similar results were obtained using the station Idaho Falls 46W, finding even stronger relationships between the MJJA-Tmean and the previous FM and MJJA soil moisture, but for the period 1954-2001 (**figure 1-4** and **table 1-2**).

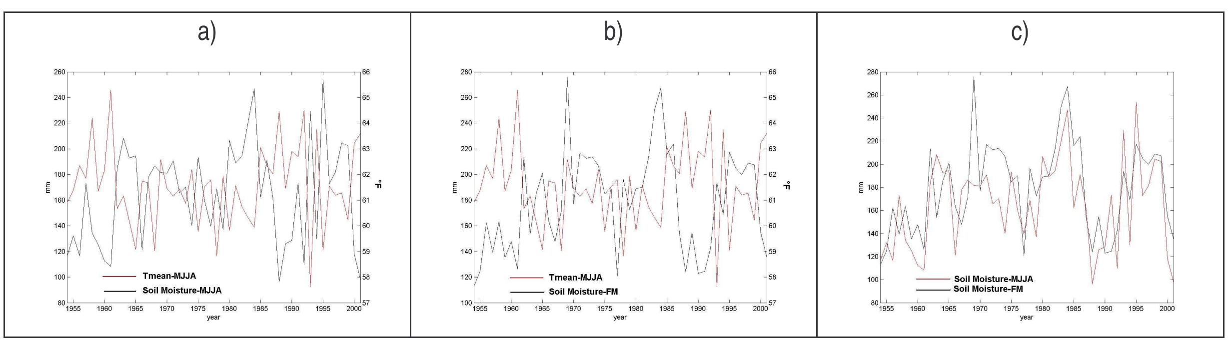


Figure 1-4. Time Series Plotting for Soil Moisture and Mean Temperature for Different Seasons. *a) MJJA-Soil Moisture and MJJA-Tmean, r = -0.72, b) MJJA-Soil Moisture and MJJA-Tmean is -0.43 (Figure 1-1b) and 1-1c) FM and MJJA Soil Moisture, r = 0.70. Soil moisture data are for climate division 9 and Tmean data are for the station Idaho Falls 46W, 1954-2001. All the correlation values have statistical significance greater than the 95% level.*

Table 1-1. Contingency Analysis Between a) FM-Soil Moisture, and MJJA-Tmean, b) MJJA-Soil Moisture and MJJA-Tmean and c) FM-Soil Moisture and MJJA-Soil Moisture. *All data are for climate division 9, 1932-2003 (α = 0.01 => ***, 0.05 => **).*

a)		(< 61.4 °F)	Tmean-MJJA	(> 62.2 °F)
		BN	N	AN
Soil Moist.	BN (< 156.0 mm)	21**	33	46**
FM	N	33	38	29
	AN (> 190.8 mm)	46**	29	25

b)		(<61.4 °F)	Tmean-MJJA	(>62.2 °F)
Soil Moist.		BN	N	AN
Clim. Div. 9	BN (<137.0 mm)	17***	20**	63***
MJJA	N	20**	51**	29
	AN (>178.7 mm)	63***	29	8***

c)		(< 137.0 mm)	Soil Moist.-MJJA	(> 178.7 mm)
Soil Moist.		BN	N	AN
Clim. Div. 9	BN (< 156.0 mm)	67***	25	8***
FM	N	29	37	34
	AN (> 190.8 mm)	4***	38	58***

Table 1-2. Contingency Analysis Between a) FM-Soil Moisture and MJJA-Tmean, b) MJJA-Soil Moisture and MJJA-Tmean and c) FM-Soil Moisture and MJJA-Soil Moisture. *Soil Moisture data are for climate division 9 and Tmean data are for the station Idaho Falls 46W, 1954-2001 (α = 0.01 => ***, 0.05 => **).*

a)		(< 61.2 °F)	Tmean-MJJA	(> 62.0 °F)
Soil Moist.		BN	N	AN
Clim. Div. 9	BN (< 154.9 mm)	6***	44	50**
FM	N	38	37	25
	AN (> 200.2 mm)	56***	19	25

b)		(< 61.2 °F)	Tmean-MJJA	(> 62.0 °F)
Soil Moist.		BN	N	AN
Clim. Div. 9	BN (< 138.9 mm)	6***	31	63***
MJJA	N	19**	50**	31
	AN (>185.1 mm)	75***	19**	6***

c)		(< 138.9 mm)	Soil Moist.-MJJA	(> 185.1 mm)
Soil Moist.		BN	N	AN
Clim. Div. 9	BN (< 154.9 mm)	63***	31	6***
FM	N	37	25	38
	AN (> 200.2 mm)	0***	44	56***

There is also a simultaneous relationship between the sum of the MJJA normalized electrical load values (i.e., total summer electrical load) and the Idaho's climate division 9 Tmean/Soil Moisture values for the 1997-2003. **Figure 1-5** shows that when negative (positive) Soil

Moisture (Tmean) anomaly values were observed the loads were above or equal to the median and when positive (negative) Soil Moisture (Tmean) anomaly values were observed the loads were under the median.

The results shown in **figure 1-5** and **tables 1-3 and 1-4**, suggest that previous soil moisture or Tmean data could be used as predictors for the loads associated within pumping during MJJA. PacifiCorp has expressed that the achievement of predictive relationships by the beginning of April are desirable. For this purpose, a stepwise routine was used to identify predictive linear regression models between various parameters. For predictors, previous values of various climate indices (for example DJF-Pacific Decadal Oscillation and Southern Oscillation Index) and variables (for example FM-Tmean and Soil Moisture) were examined. For the predictand, the sum of the MJJA normalized load data was used. Only soil moisture anomalies during the previous FM and February were retained as predictors for the MJJA loads estimation. These models could be summarized by the following equations:

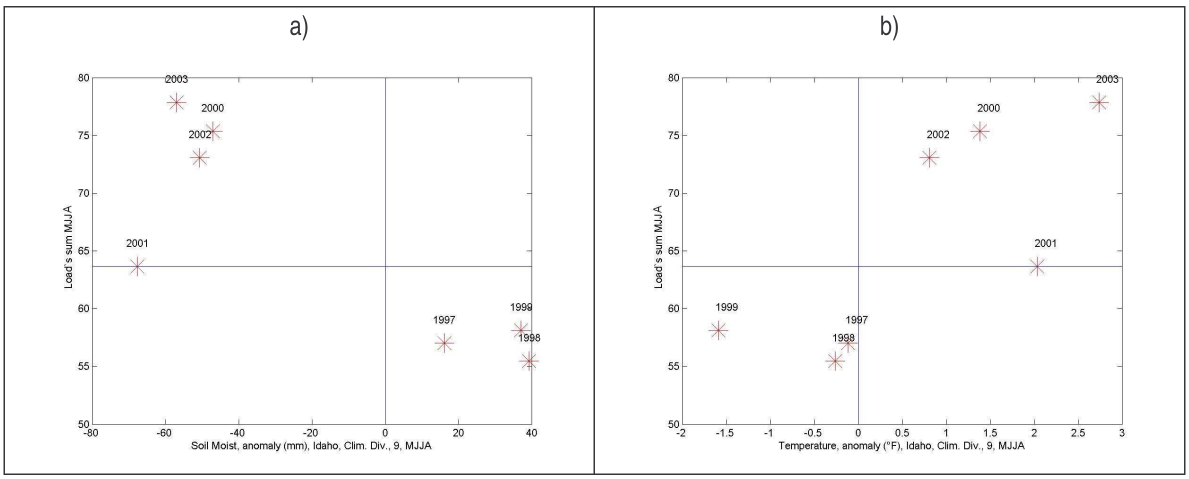


Figure 1-5. Scatter Plots Between the Sum of the MJJA Normalized Load Values Versus a) MJJA-Soil Moisture and b) MJJA-Tmean (red asterisks, 1997-2003). *Vertical blue lines are for the zero anomaly value and the horizontal ones are for the load's median value.*

$$\hat{Y} = 62.863 - 0.211\, Soil\ Moist.(FM) \text{ (1), and,}$$

$$\hat{Y} = 63.917 - 0.202\, Soil\ Moist.(Feb.) \text{ (2),}$$

where $\hat{Y}$ is the estimated value of MJJA loads. The model statistics are summarized in **table 1-3**. Equations (1) and (2) show models with negative correlations between the soil moisture anomalies and the load values, so negative (positive) soil moisture anomalies during the previous FM or February tends to be related with the latest (earliest) load data as is presented in Fig. 4. This is also in agreement with the simultaneous relationship (**figure 1-5a**). This signifies that under a wet soil moisture scenario, less water is pumped as the crops do not need to be irrigated as much as under a dry soil moisture scenario. The reason for this could be again persistence of the FM-soil moisture conditions through MJJA. A similar result is obtained if MJJA-soil moisture anomalies are used in regression as an independent variable (eq. 3, see also **figure 1-6a**):

$$\hat{Y} = 62.757 - 0.164\, Soil\ Moist.(MJJA) \text{ (3),}$$

where $\hat{Y}$ is the estimated value of MJJA loads as before. **Figure 1-6b** also suggests that positive (negative) MJJA-Tmean anomalies tend to be related to the latest (earliest) load data. This means that under a wet soil moisture scenario, less water is pumped as there is less evapotranspiration from the plantations due to the below than normal temperatures and under a dry soil moisture scenario, evapotranspiration is elevated due to above normal temperatures (see **tables 1-1** and **1-2**).

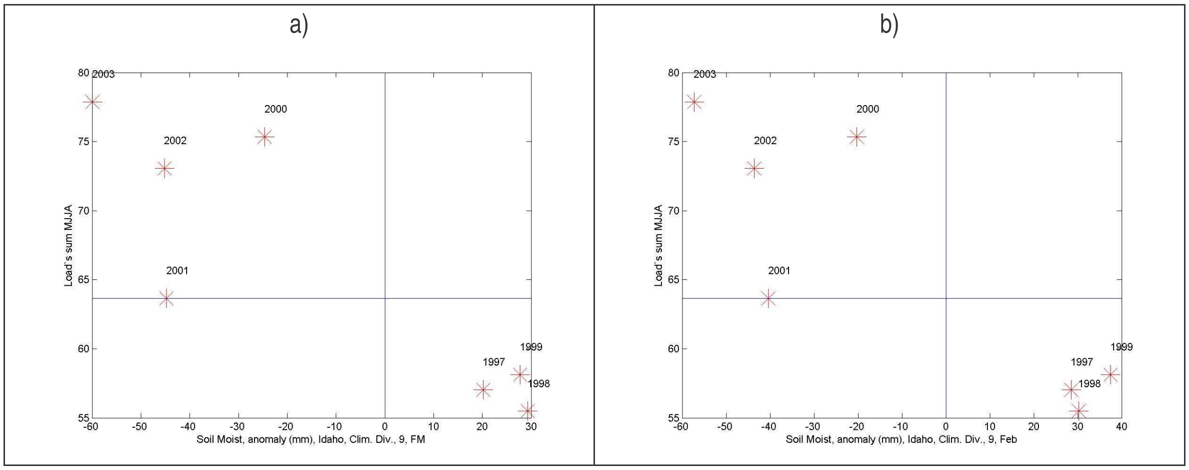


Figure 1-6. Scatter Plots Between the Sum of the MJJA Normalized Load Values Versus a) FM-Soil Moisture and b) Feb.-Soil Moisture (Red Asterisks, 1997-2003). *Vertical blue lines are for the zero anomaly value and the horizontal ones are for the load's median value.*

Table 1-3. Statistics Associated to the Models Described in the Equations (1)-(3). The Skill, Mean Absolute Deviation and Maximum Absolute Deviation values where obtained by cross validation. All the Skill values have statistical significance greater than the 95% level.

	Equation Number		
Statistics	**(1)**	**(2)**	**(3)**
R	0.86	0.86	0.82
R^2	0.74	0.74	0.67
Skill	0.76	0.76	0.64
Mean Absolute Deviation (MAD)	4.37	4.56	5.60
Maximum Absolute Deviation	11.52	11.08	15.06

The observed and estimated values for all the models are plotted in **figure 1-7**. Notice that the maximum absolute deviations described in **table 1-2** are for the year 2001. In the **figures 1-7a** and **1-7b** the estimate for the year 2004 is also included. Both estimated values are greater than the median for the 1997-2003 period, related to the dry conditions observed during last February and March, but the confidence interval for this estimate is large, mainly due to the small sample size used (7 years).

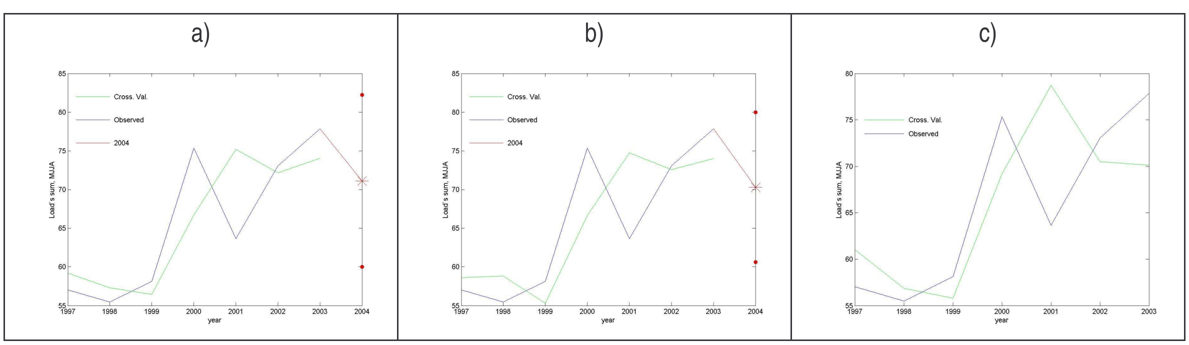


Figure 1-7. Observed and Estimated Values for the Models Described in the Equations a) (1), b) (2) and c) (3). *The red dots in a) and b) are the 95% statistical confidence levels for the 2004 estimation.*

1.5 Economic Benefit of Pump Load Forecast

The net economic benefits of the pump load forecasts would accrue by being able to purchase power contracts in advance of June 1 for usage in May and June of that same year. For the purposes of this study, we evaluated the benefit of power contracts purchased 1 to 2 months ahead, using forecasts produced by April 1 for the ramp-up date for the following spring-summer (usually in April, May, or June). This analysis assumes "perfect" action based on forecast information, meaning that appropriate actions will be taken by decision-makers following the forecast output.

The net economic benefit of forecast information is calculated to be approximately $X per year, using the following assumptions. If decision-makers knew by April 1 that the ramp-up date would be at the beginning of May (or earlier) rather than toward the end of May (or later), then they could purchase contracts 1 or 2 months ahead (for energy usage in May and June). The difference between the contracts and the spot market price (i.e., if they waited to buy the power in May and June) is approximately $X/MWh. This calculation assumes that the 1-month and 2-month ahead contracts are $X/MWh, and the spot market price ranges from $X/MWh to as high as $X/MWh, with an average of around $X/MWh. The cost of a forecast of an early May ramp up-date, but with actual ramp-up occurring in late May, is the same as (a), namely $X/year, as are the benefits and cost of a forecast of a later ramp-up date (i.e., zero, relative to existing information and standard operating procedures).

1.6 Summary

This analysis found that there is a statistical basis for predicting the controlled pump start date based on simultaneous precipitation and the antecedent soil moisture (**figure 1-8**), but that there is no predictive skill for springtime precipitation in that region. For this reason, there is no good climatological prediction of pump start date.

Conversation with the stakeholder revealed that they were more interested in total summer load than actual pump start date. More success was found with this predictive skill. That is spring soil moisture conditions have a strong tendency to persist through the summer. There is then a strong (negative) correlation between summer soil moisture conditions and summer temperatures, i.e., wet soil tends to moderate the summer temperature extremes. Lastly, there is a strong relationship between summer soil moisture and temperature and total summer pump loads. As a result, there is reasonable predictive skills of summer pump loads based on spring soil moisture conditions in southeastern Idaho.

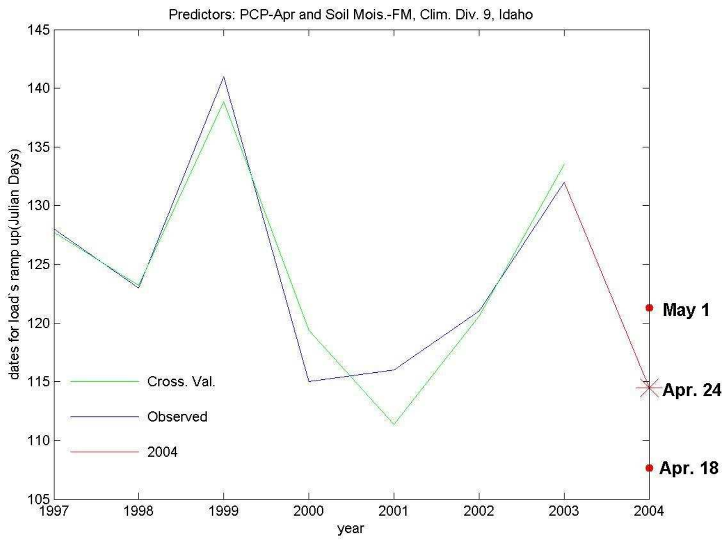


Figure 1-8. Observed and Estimated Values for the Pump Start Dates (Skill = 0.95). *The red dots are the 95% statistical confidence levels for the 2004 estimation.*

Using these forecasts was estimated to be able to save PacifiCorp $X, per summer irrigation season.

1.7 References

- Huang, J., H. van den Dool and K. Georgarakos, 1996: Analysis of Model-Calculated Soil Moisture over the United States (1931–1993) and Applications to Long-Range Temperature Forecasts. *J. of Climate*, **9**(6), 1350–1362.
- NCDC, 2003: Data documentation for data set 3200 (DSI-3200): Surface land daily cooperative summary of the day. National Climatic Data Center, Asheville, NC, 36 pp.
- van den Dool, H., J. Huang and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981-2001, *J. Geophys. Res.*, **108(D16), 8676**, doi:10.1029/2002JD003114.

2. Towards Operational Use of Probability Forecasts: Cross Validation and Real-time Updating

2.1 Introduction

Chapter 2 of the Cal ISO benchmark report of Deliverable 4 presented details of the use of multi-model ensemble forecasts to reduce the weather dependent component of Cal ISO load forecast error. Several probability forecast models were developed and evaluated on data from the summer of 2003. In this report, issues which arise in the real-time use of this approach are explored further. One day ahead (24 hour) forecasts are considered for the summers of both 2002 and 2003.

The significance tests in Deliverable 4 used "drop-one-out" verification to predict a given day in 2003 data from the entire summer of 2003 except the day (or days) that was used. This removes misleading results due to in-sample fitting, but is vulnerable to issues of changes in the NWP models from one year to the next. For example, to predict September 1, 2003, data from the last weeks in September would have been used. In a real-time system that data would not have been available, and only September data from previous Septembers would be known on 1 September 2003. In this report, true out-of-sample forecasts for 2003 (and 2002) were performed. Historical data provided by Quantum Weather allows both drop-one-out testing within 2002 and true out-of-sample cross validation. A model constructed from observations of the summer of 2003 was evaluated on the summer of 2002, and vice versa. To summarize:

1. Using the drop-one-out tests on 2002 data the model structure is shown to generalize well. There is no extreme degradation in skill when the tests previously performed on 2003 data are repeated on 2002 data. This implies that the model structures chosen in the previous report are effective.
2. Using true cross-validation indicates a significant bias is introduced with predicting 2003 temperatures from a model based on 2002 data (and vice versa). The bias is sufficiently large as to significantly degrade the results of a blind application of the technique across years.
3. A sliding window technique is demonstrated, which allows real-time true out-of-sample forecasting within a given season. True out-of-sample tests show that this approach exploits the successful model structure in a deployable, real-time method by daily refitting the model on the observations over the previous 90 days.

This chapter is structured as follows. The following section provides a brief review of the modeling procedure. A simple example of the input values, model parameters and results for a single day's forecast (1 September 2003) are given. In addition, this section includes a set of tables that summarize the various cross validation tests discussed in the remainder of the document.

In section 3, the drop-one-out results for 2002 are discussed. This section addresses the question of whether the modeling approach shown to be effective in the summer of 2003 is also effective over the summer of 2002. This is a test of whether or not the mathematical structure used to model Cal ISO's regional temperature is robust. It is demonstrated that it is robust. Results should be contrasted with those in Tables 2-2, 2-3, 2-4, and 2-5 of the original report, and demonstrate forecasts of similar quality.

In section 4, the annual out-of-sample cross-validation procedure is stated, and the results of the predicting 2003 using a model constructed from 2002 are discussed (which again mirrors Tables 2-2, 2-3, 2-4, and 2-5 of the original report). The question here is whether a forecast package trained on one summer could then be employed in another year. In this case large forecast biases were detected.

In section 5 the following question is explored. Is it possible to estimate parameters which could exploit operationally this successful model structure in real time forecasts of regional temperatures and demand? To determine this, a deployable moving window approach to estimate the parameter is defined and evaluated for September 2003. The performance of this model is then contrasted with the drop-one-out results and the 2002-based model in some detail. The approach suggests an algorithm that provides similar skill and could be used in practice.

To aid in distinguishing the three methods discussed below, they will consistently be referred to as (1) the drop-one-out method, (2) the annual out-of sample method and (3) the operational WIN method.

2.2 Packages and Tables of Results

This section demonstrates how a forecast is computed, using the NWP model output for each station. The table below shows all parameters and input variables required to compute the AVN-package forecast for the Bay area on 1 September 2003. The first two columns identify the input variable (station and model variable), the next four columns reflect the statistics of the AVN package coefficients (these are discussed in the relevant sections below). Column 7 gives the operational coefficient available when the forecast for 1 September was made. The last two columns presents the forecast.

Table 2-1: Input Parameter Values for the 1 September 2003 Forecast

AVN Package Bay Region Regression Coefficients and One Example Forecast for September 1, 2003								
Station	Model/Time	In Sample Median	In Sample Mean	In Sample Standard Deviation	In Sample Variational Coefficient	Operational Coefficient Set for September 2003	1 September 2003 Input	Component
KSFO	AVN 18:00	-1.1285	-1.1375	0.3807	-0.3347	-1.1317	82.15	-92.97
KSFO	AVN 21:00	-0.1664	-0.1488	0.4476	-3.0087	-0.1555	85.23	-13.25
KSFO	AVN MOS 18:00	-1.0652	-1.0649	0.0662	-0.0622	-1.0644	65	-69.18
KSFO	AVN MOS 21:00	0.7835	0.7822	0.0702	0.0898	0.7815	72	56.27
KOAK	AVN 18:00	7.3448	7.3531	0.9505	0.1293	7.3361	89.19	654.30
KOAK	AVN 21:00	-4.3936	-4.4214	1.0603	-0.2398	-4.4035	94.66	-416.83
KOAK	AVN MOS 18:00	0.8186	0.8142	0.0814	0.1000	0.8153	65	52.99
KOAK	AVN MOS 21:00	-0.2026	-0.1970	0.0758	-0.3847	-0.1973	73	-14.40
KCCR	AVN 18:00	-6.8466	-6.8318	-0.6226	-0.0911	-6.8217	93.5	-637.83
KCCR	AVN 21:00	4.5555	4.5517	0.7069	0.1553	4.5399	100.43	455.94
KCCR	AVN MOS 18:00	0.4300	0.4285	0.0765	0.1784	0.4274	77	32.91
KCCR	AVN MOS 21:00	-0.2127	-0.2158	0.0791	-0.3668	-0.2141	86	-18.42
KLVK	AVN 18:00	2.3067	2.3093	0.3605	0.1561	2.3099	93.62	216.25
KLVK	AVN 21:00	-1.8948	-1.8952	0.3986	-0.2103	-1.8948	100.56	-190.54
KLVK	AVN MOS 18:00	0.3106	0.3130	0.0737	0.2355	0.3122	79	24.67
KLVK	AVN MOS 21:00	0.1548	0.1527	0.0888	0.5816	0.1526	91	13.89
KSJC	AVN 18:00	-1.9467	-1.9483	0.1812	-0.0930	-1.9462	86.23	-167.82
KSJC	AVN 21:00	2.9708	2.9724	0.2116	0.0712	2.9797	89.49	264.85
KSJC	AVN MOS 18:00	0.7051	0.7109	0.865	0.1217	0.7093	74	52.49
KSJC	AVN MOS 21:00	-1.1372	-1.1431	0.0890	-0.0779	-1.1420	83	-94.79
Constant	Constant	-19.4824	-19.5098	1.1220	-0.0575	-19.4904		
							Forecast	90.05
							Observed	91.5

The models are constructed in the following way. For each day and for each region, the Cal ISO regional temperature defines the target which the model is used to forecast. The NWP values of the day are collected, re-weighted with a set of coefficients and combined to form the forecast.

How the parameters are determined depends upon which cross validation method is used. In each case, the historical observations are divided into a learning set and a test set. In the **drop-one-out method**, the learning set consists of all forecasts and observations from that summer (2002 when predicting 2002, 2003 when predicting 2003) *except* those days which are currently being forecast. The strength of this method is that dropping out the current target generates model parameters that are not artificially over-skillful (due to having seen the answer for which they are used to forecast). Nevertheless, the parameters used to forecast for 1 September depend on the observations from the last week in September, so while statistically valid, the method is not deployable operationally.

The **annual out-of-sample** cross validation method uses a learning set from a different summer than the target set. That is, the parameters which are computed from the summer of 2002 are frozen and then used to forecast the summer of 2003. This method is deployable, but it tests two separate aspects. In order to maintain high performance, the parameters estimated from one year must be similar to the parameters of the other year. Second, the model structure used in one year (the choice of variables, the selection of NWP models, etc.) must also be robust. The drop-one-out method tests the robustness of the method, the annual out-of-sample method then tests how well the parameters for one year can be estimated from data in some other year. There are several reasons why the optimal parameters might change. These include changes in the weather itself, changes in the NWP models (in particular in the MOS statistics), seasonal variations in the best parameters, and a lack of robustness in the original estimates due to insufficient data.

The **operational sliding-window (WIN) method** uses the last 90 days as the learning set to estimate the parameters. It will have a smaller learning set than the annual out-of-sample method, but this smaller learning set will be taken from a more relevant period (the recent past). Both the annual out-of-sample and the sliding window can be used operationally. Which method performs best depends on the details of the system and the model. There is no reliable rule as to which should be expected to forecast better.

Tables 2a-d report the bias, the root-mean-square (RMS) error, the standard deviation of the error (sdev) and the ignorance score for both the drop-one-out and the annual out-of-sample forecasts for 2003. There is one table for each region and in each case three packages are used: the AVN package, the ETA package, and the PF1 Package. The second column ("Learn Set") indicates which data was used as the learning set when building the package. The next three columns then give performance statistics for PF1, AVN and ETA packages. Taking the Bay Area as an example, note that the error standard deviations are similar between the two learning sets, the bias is much larger when the 2003 temperature forecasts are interpreted with the 2002 packages. In the drop-one-out case (that is, using the 2003 package) the bias is about 0.01, while it is about –1 in the annual out-of-sample case (that is, using the 2002 package). The difference is most dramatic in the SDGE area and the AVN package. The bias in NBAY remains relatively small. In each region, the bias significantly increases the ignorance score, decreasing the profitability of the forecasts.

Table 2-2a: Forecast error statistics in 2003 of the PF1, AVN, ETA packages in the Bay area

Bay Area 2003				
	Learn Set	PF1	AVN	ETA
Bias	2002	-1.13	0.05	0.64
	2003	0.01	0.01	0.01
Stdev	2002	3.46	3.69	3.78
	2003	3.59	3.61	3.88
Ignorance	2002	-1.07	-1.09	-1.00
	2003	-1.12	-1.12	-1.01

Table 2-2b: Forecast error statistics in 2003 of the PF1, AVN, ETA packages in the NBay area

Non Bay Area 2003				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.24	0.64	-0.16
	2003	0.04	0.03	0.03
Stdev	2002	1.94	2.03	2.66
	2003	1.58	1.70	1.77
Ignorance	2002	-1.93	-1.53	-1.54
	2003	-2.03	-1.92	-1.86

Table 2-2c: Forecast error statistics in 2003 of the PF1, AVN, ETA packages in the SCE area

SCE Area 2003				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.18	1.31	1.08
	2003	0.03	0.03	0.01
Stdev	2002	2.08	2.32	2.50
	2003	1.87	1.92	2.75
Ignorance	2002	-1.57	-1.02	-0.99
	2003	-1.75	-1.72	-1.20

Table 2-2d: Forecast error statistics in 2003 of the PF1, AVN, ETA packages in the SDGE area

SDGE Area 2003				
	Learn Set	PF1	AVN	ETA
Bias	2002	1.66	4.57	1.86
	2003	0.01	0.01	0.01
Stdev	2002	3.53	3.83	3.61
	2003	3.37	3.35	3.72
Ignorance	2002	-0.61	0.35	-0.60
	2003	-1.12	-1.13	-0.98

The data presented in **Tables 3a-d** shows the same forecast error statistic, this time for forecasting the summer of 2002.

Table 2-3a: Forecast error statistics in 2002 of the PF1, AVN, ETA packages in the Bay area

Bay Area 2002				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.03	0.06	0.00
	2003	0.15	-0.40	-0.50
Stdev	2002	3.16	3.60	3.33
	2003	3.11	3.61	3.18
Ignorance	2002	-1.31	-1.12	-1.24
	2003	-1.31	-1.09	-1.24

Table 2-3b: Forecast error statistics in 2002 of the PF1, AVN, ETA packages in the NBay area

NBay Area 2002				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.02	0.00	0.00
	2003	-2.03	-3.72	0.35
Stdev	2002	1.42	1.55	1.55
	2003	3.11	3.61	3.18
Ignorance	2002	-2.20	-2.07	-2.08
	2003	-0.89	1.71	-1.83

Table 2-3c: Forecast error statistics in 2002 of the PF1, AVN, ETA packages in the SCE area

SCE Area 2002				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.00	-0.03	0.01
	2003	-0.54	-0.51	-1.39
Stdev	2002	1.86	1.89	1.91
	2003	1.81	1.98	1.89
Ignorance	2002	-1.62	-1.60	-1.58
	2003	-1.60	-1.48	-1.25

Table 2-3d: Forecast error statistics in 2002 of the PF1, AVN, ETA packages in the SDGE area

Bay Area 2002				
	Learn Set	PF1	AVN	ETA
Bias	2002	0.00	0.06	0.04
	2003	-4.89	-3.69	-1.16
Stdev	2002	2.62	2.77	2.79
	2003	2.85	3.23	2.87
Ignorance	2002	-1.30	-1.21	-1.21
	2003	0.35	-0.14	-1.00

The origin of the bias of –4.89 for the PF1 predictor in the year 2002 trained with data on 2003 in the SDGE region is shown in Figure 1, a time series of the forecast and target values indicates that the problem is in fact one of drift in the bias. This figure shows the SDGE region forecasts for 2003 for a package with parameters based on the summer of 2002. The forecasts track the verifying observations, but are systematically too low. A scatter diagram of the errors for each day is shown in Figure 2 which confirms this result. This is discussed further in Section 4.

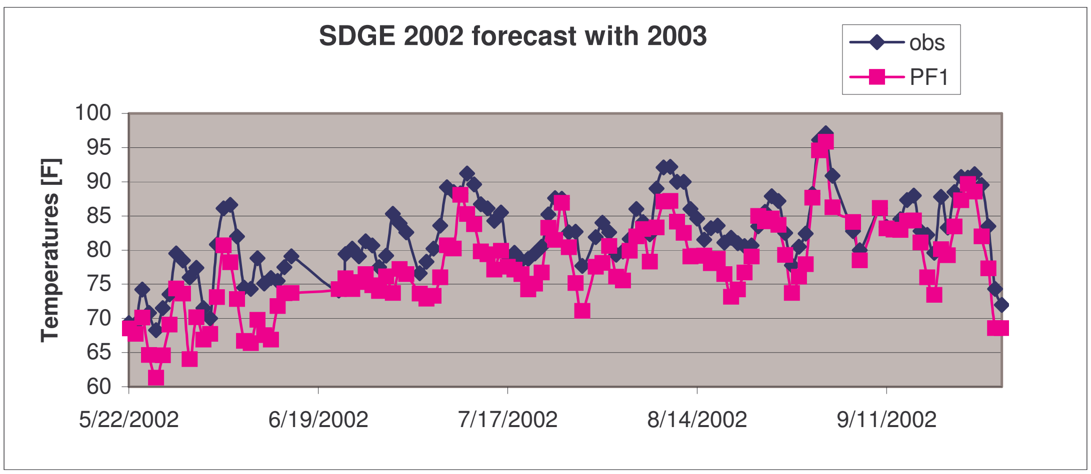


Figure 2-1. The SDGE region forecasts for 2003 for a package with parameters based on the summer of 2002

2.3 Drop-one-out results for the summers of 2002 and 2003

The new drop-one-out method results above show that the modeling approach works as well in 2002 as it was previously shown to work for the summer of 2003 as reported in Deliverable 4. This method forecasts one day in a summer by using all the data from that summer, except those day (or days) actually being predicted which are 'dropped out' of the learning set. For the Bay Area, the 2002 ignorance scores are in fact lower (better) than those of 2003 for each of the three packages. Similar results are found in the other three regions as well. Both the bias standard deviation of the errors of the 2002 drop-one-out forecasts are smaller (better) than 2003, indicating that to some extent 2002 was forecast more accurately than 2003.

2.4 Annual Out-of-sample Method

The tables in section 2 above are designed to allow easy comparison of the differences in forecast quality when a given target summer is forecast using each of the two summers as the learning set.

The details provided below indicate that there is a significant bias error in the cross-predictions, possibly due to changes in the NWP models. The drop-one-out forecasts of 2002 based on 2002 (Table 3a-d) show similar skill to the 2003-2003 drop-one-out forecasts in Table 2. This indicates that the statistical method holds. Looking at the cross-predictions, however, indicates that while the standard deviation of the errors is similar, there is a large increase in the systematic bias.

Table 2a of the Bay area in 2003, indicates that the bias is much larger when the learning set is for 2002 than when it is for 2003. Similarly in Table 3a for Bay Area in 2002, the bias using 2003 as the learning set is larger than when using the 2002 learning set. This theme runs through each of the regions and all three of the packages. For each measure of forecast skill the annual out-of-sample result is worse, often significantly, than the drop-one-out method.

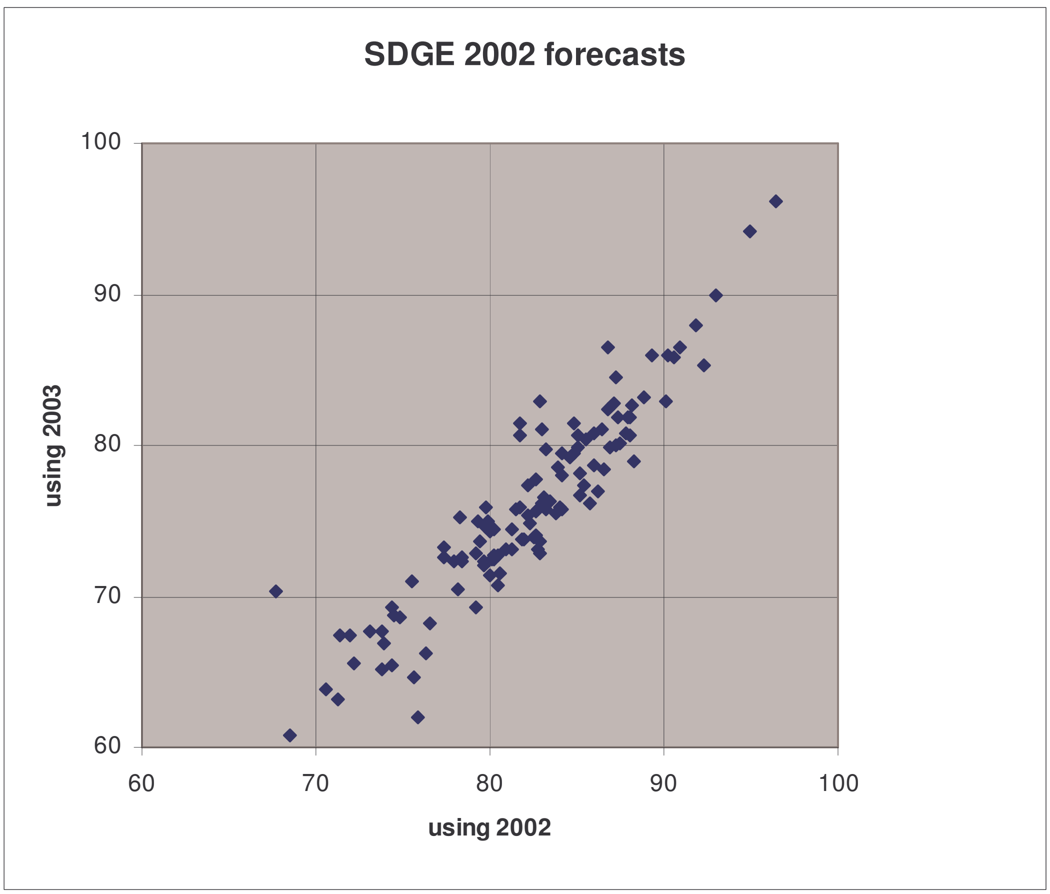


Figure 2-2: Scatter diagram of the two regional temperature forecasts for the SDGE region for each day in 2003 based on the learning set in 2003 and the learning set in 2003

The observed bias does not appear to be due only to inter-annual changes in the NWP models and differences between the weather observed in one summer and the next. In addition to these two known effects, examining bias within a month suggests some seasonality issues exist. The data set is too small to convey a large degree of confidence but is indicative as shown in the

AVN package trained on data up to 31 August 2003 and then used to forecast the month of September 2003 out-of-sample.

The day to day variations in bias and standard deviation of the forecast error in the median (not shown) are considered next. As expected from the small number of days used to estimate the bias in the first week of September, it moves about a good deal. After 10 September, however, there is a slow systematic move towards an increasing bias later in the season. The standard deviation of the error reflects a few large misses, but shows no clear systematic drift.

There are three effective remedies for the bias-between-years issue:

1. The introduction of a more flexible statistical modeling strategy
2. Real-time updating of the learning set and statistical model day by day.
3. Careful tracking of "upgrades" of the NWP models or MOS statistics.

Given the apparent drift in the bias in September 2003 above, a sliding window approach was investigated. This allows the learning set to be composed from the most recent observations and forecasts, thereby partially accounting for both changes in the NWP model inputs and changes in the season of the year. In the next section, this second approach is shown to provide an effective way forward.

2.5 Operational Sliding window method on September 2003

The question presented here is the following. How can seasonal changes and changes in the statistical properties of the NWP inputs be accounted for? One approach is to allow the training set to evolve in time, keeping the most recent, ideally most relevant, data. In this section a 90 day learning set is used- the 90 days being those immediately before the forecast date. Thus, this method is deployable in real-time and the coefficients are said to be dynamic (since they change every day based on what happened the previous day). There are a number of model parameters that should be verified before the model is, in fact, used. The length of the learning set is one such parameter. No tuning of this type has been done on the results presented below.

Since a 90 day learning set is required, only results for September(s) are presented. For clarity, only the AVN and ETA results are shown. The forecasts reflect the performance of the median. Kernels are not evaluated (the kernels do not change as quickly). Finally, one additional method is introduced for comparison. Consider the case where the learning set is fixed, containing the 90 days immediately prior to 1 September. This model is referred to as FROZEN. The purpose of contrasting FROZEN and WIN is to see potential seasonal effects independently of merely having the same calendar year. As shown below, WIN outperforms FROZEN, suggesting that the seasonal effects are at play.

First, the bias is examined between RMS and the standard deviation (sdev) of the WIN forecasts for each of the regions. These are shown in the Table 4a-d for September 2003.

Table 2-4a: Bay Area Sept 2003 out-of-sample WIN

Package	bias	rms	sdev
AVN	1.37	4.37	4.24
ETA	2.67	4.30	3.43

Table 2-4b: NBay Area Sept 2003 out-of-sample WIN

Package	bias	rms	sdev
AVN	0.69	2.37	2.32
ETA	0.34	2.46	2.49

Table 2-4c: SCE Area Sept 2003 out-of-sample WIN

Package	bias	rms	sdev
AVN	0.02	2.21	2.25
ETA	1.52	2.58	2.13

Table 2-4d: SDGE Area Sept 2003 out-of-sample WIN

Package	bias	rms	sdev
AVN	0.20	2.78	2.83
ETA	0.57	3.61	3.64

The September biases from WIN above show some improvement over the 2003 annual out-of-sample forecasts shown in Table 2. They also out perform FROZEN on September (tables not shown).

Figure 3 contrasts these different forecasts on a day by day basis through September 2003. The FROZEN predictions (in yellow) are systematically higher than the observations (dark blue). The dynamics window WIN forecasts (dark purple) approach those reported in Deliverable 4, which is the drop-one-out forecasts (in purple).

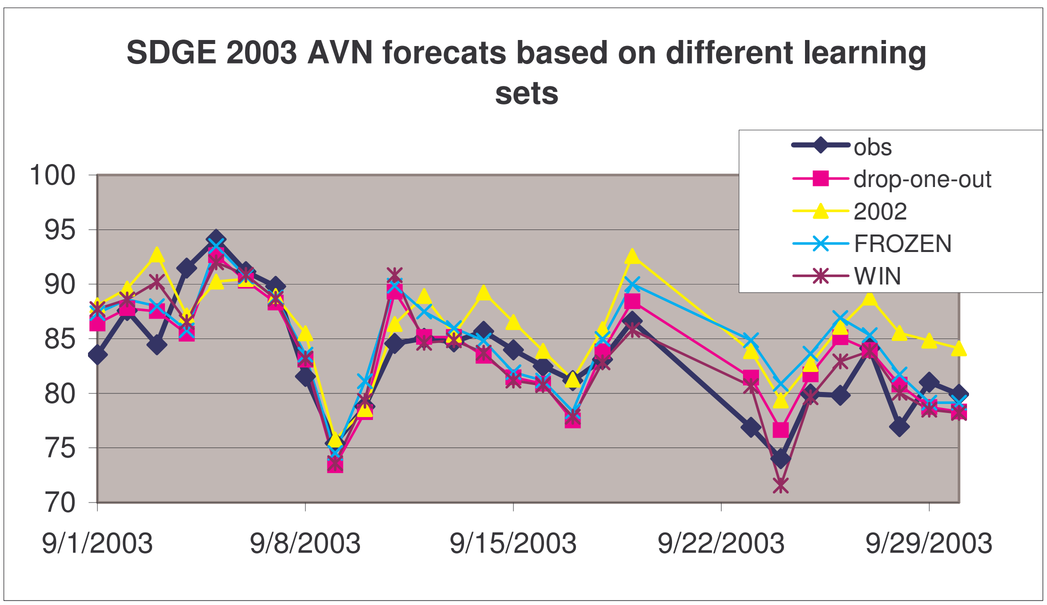


Figure 2-3: Forecasts of the SDGE regional temperatures for September 2003 based on the AVN package input using the drop-one-out method, the 2002 data, the FROZEN 2003 and the WIN method.

The WIN forecasts are clearly superior, being identical to the FROZEN on 1 September (by construction) and become more similar to the drop-one-out forecasts as the end of September approaches.

The procedure of this section is now repeated on the 2002 data. Table 5a-d and Figure 4 represent the results.

Table 2-5a: Bay Area Sept 2002 out-of-sample WIN

Package	bias	rms	sdev
AVN	-0.31	3.79	3.84
ETA	0.14	3.64	3.70

Table 2-5b: NBay Area Sept 2002 out-of-sample WIN

Package	bias	rms	sdev
AVN	0.00	1.67	1.70
ETA	0.38	1.64	1.62

Table 2-5c: SCE Area Sept 2002 out-of-sample WIN

Package	bias	rms	sdev
AVN	0.23	2.61	2.65
ETA	0.37	2.54	2.56

Table 2-5d: SDGE Area Sept 2002 out-of-sample WIN

Package	bias	rms	sdev
AVN	-0.63	2.89	2.87
ETA	-0.65	3.66	3.67

Comparing the WIN skill score set with that of FROZEN in September 2002 (not shown) shows much smaller differences than in September 2003, which may indicate that the coefficient are more stable in 2002 than in 2003.

Figure 4 shows the September 2002 forecasts for one region (in this case NBAY) using the drop-one-out method, FROZEN and the dynamic WIN method for one model (this time, AVN). FROZEN 2003 are the predictions using the annual out-of-sample method, while FROZEN 2002 use the 2002 data up to 31 August 2002. Frozen 2002 (magenta) is already significantly better than Frozen 2003 (yellow), WIN and the drop-one-out forecasts are very similar in 2002.

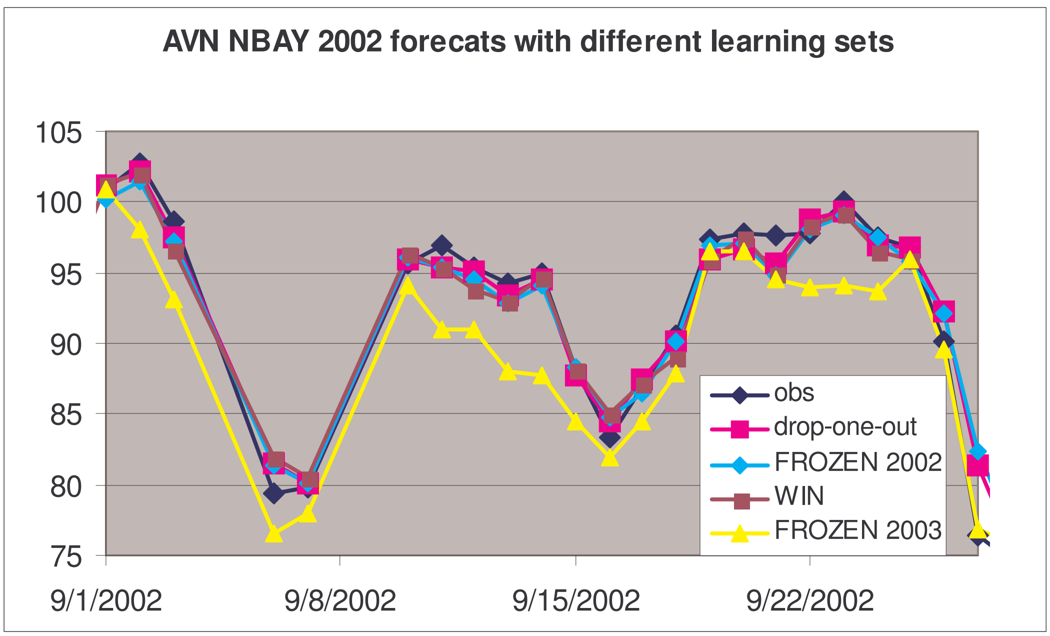


Figure 2-4: Forecasts of the Non - BAY area regional temperatures for September 2003 based on the AVN package using the drop-one-out method, the FROZEN 2002 and the WIN method. *In this case also the forecast based on the annual out-of-sample results (Frozen 2003) where the model coefficients are determined in 2003 and frozen are shown. Clearly, this is the worst of the 4 forecasts shown.*

Finally, September 2003 was considered and contrasted to the Cal ISO AVN forecast with those of the 90 day WIN methods. The bias and RMS error for each region is shown in Table 6. Although the counting statistics are small, but in the areas of large bias of RMS error (that is, BAY and SDGE) the WIN approach performs well, especially in terms of reducing the bias. Note from Table 1 that the WIN model does not have access to the AVN forecast of TMAX as this product was not available from the archive. It would be interesting to see how much improvement its inclusion into WIN would bring.

Table 2-6: Cal ISO and WIN forecast error statistics for all four regions and 27 days in September 2003

Sep-03	BAY		NBAY		SCE		SDGE	
	Cal_Iso	WIN	Cal_Iso	WIN	Cal_Iso	WIN	Cal_Iso	WIN
Bias	3.16	1.06	-0.34	0.63	0.53	-0.10	1.72	0.03
RMSE	5.91	4.36	2.70	2.32	1.90	2.27	3.84	2.80
Correlation	0.80	0.88	0.88	0.94	0.87	0.82	0.72	0.85

Note that these results are based on September 2003 omitting the three days 20-22nd September which have been dropped from both the Cal ISO forecast statistics to allow it a fair comparison. That is, the models are compared using forecasts on the same set of days.

The worst under-forecasts of MW are these 3 missing days. They are the largest MW under-forecasts in that month of September 2003. Unfortunately, it is not known how the WIN packages would have done. Another day with a large under-forecast of more than 1000 MW is the 13th of September. On this day (in all 4 regions) the AVN package WIN has an error of smaller magnitudes that the Cal ISO AVN forecast. Furthermore, rather than under-forecasting, the AVN package WIN over-forecasts in three out of the four regions.

2.6 Conclusions

The analysis of temperature forecasts for 2003 presented in Deliverable 4 has been repeated on data from 2002 and extended through the examination of cross forecasts. An explicit example of the parameters and inputs for the Bay region on 1 September 2003 has been given. The main conclusions of this reanalysis of 2003 data using the new 2002 data, and the 2002 data itself, are

1. The same modeling approach gives slightly better results for 2002, suggesting that the summer of 2002 may have been intrinsically easier to forecast than the summer of 2003.
2. That cross prediction, using 2002 data as the learning set when predicting 2003 (and vice versa) introduces a bias into the forecasts; this bias is large enough to significantly degrade the operational value of the system.
3. Using a moving window learning set consisting of the 90 days immediately prior to the forecast day provides a deployable approach to this difficulty.

Future work could include obtaining sufficient data such that the entire summer would have a 90-day data window before it, which would improve the statistical significance of the results. It would also be of use to determine the extent to which changes in the NPW system or output methods are impacting the results. Given information on the 2002 temperature and MW targets, the entire analysis of 2002 in terms both of temperature and of demand can be made and translated into likely value of the forecast to Cal ISO. While a number of questions have been raised within the report, it appears clear that the sliding window process provides similar performance to the drop-one-out methods, while avoiding non-operational aspects of the scheme.

These results do provide a more analytical basis of estimating conditional probabilities for estimating Cal ISO weather-load variations, for the asymmetric avoided costs tied to estimating the value of weather forecast error for the Cal ISO. This approach has shown an estimated $X benefit to making Cal ISO system forecasts more predictable subject to probabilistic forecast estimates. These economic benefits will likely increase with more historical test years being incorporated into the forecasts.