Subject: **RPU Wholesale & Retail Load Forecasting Methodologies**

CEC 2011 IEPR Form 4 Report

Participant: **City of Riverside, Riverside Public Utilities (RPU)**

Date: **April 21, 2011**

Contact: **Scott M. Lesch, Senior Utility Resource Analyst, RPU**

slesch@riversideca.gov

1. Overview & Introduction

RPU uses regression based econometric models to forecast both its total expected GWh system load and system MW peak on a monthly basis. Regression based econometric models are also used to forecast expected monthly retail loads (GWh) for our four primary customer classes. These models are calibrated to historical load and/or sales data extending back to January 2003. The input variables to these econometric models include various monthly weather summary statistics, specific calendar effects and two econometric input variables for the Riverside – San Bernardino – Ontario metropolitan service area; annual per capita personal income (PCPI) and monthly non-farm employment (EMP) estimates. The monthly forecasts produced by these models are normally used to project RPU wholesale gross and peak loads and retail sales five years forward. However, for this application we have extended the weather, calendar and econometric variables through 2022 to produce the CEC requested forecast information.

RPU does not currently produce forecasts of the following variables; customer counts associated with any specific customer class, peak loads associated with any specific customer class, or future electrical rates for any customer class and/or tier rate structure. Since both our wholesale and retail forecast models are calibrated to historical load data, the corresponding forecasts implicitly capture the demand side effects of all active Riverside Public Utility DSM programs at their current funding and implementation levels. Additionally, we have no ESP's in our service territory and we do not anticipate either losing any existing load or gaining any new service territory over the next 10 years.

2. Forecasting Approach

2.1. General modeling methodology

The following load based metrics are modeled and forecasted by the RPU Resources department:

- Hourly system loads (MW),
- Total monthly system load (GWh),
- Maximum monthly system peak (MW),
- Total monthly retail loads for our Residential, Commercial, Industrial and Other customer classes (GWh).

Additionally, dynamic-regression (time series) models are used to simulate the following seasonal weather information (UCR CIMIS Weather Station data) for the Riverside electrical service area:

- Riverside average daily temperature (°F)
- Riverside max-min temperature differential (°F)

These daily weather data simulations are used in our hourly system load equations (to produce prospective, simulated hourly system loads). These daily temperature simulations are also summarized into monthly cumulative cooling and heating degree indexes; the average value of these indexes are in turn used as prospective weather input values for our monthly load forecasting equations, respectively.

All primary monthly forecasting equations are statistically developed and calibrated to 7-8 years of historical monthly load data. The parameter estimates for each forecasting equation are updated every 6 months; if necessary, the functional form of each equation can be updated or modified on an annual basis. Please note that this report <u>only</u> summarizes the methodology and statistical results pertaining to our monthly forecasting equations. (Section 3 of this report describes our monthly system load and system peak equations, while section 4 discusses our class-specific, retail load models.)

2.2. Input variables

The various weather, economic and structural input variables used in our monthly forecasting equations are defined in Table 2.1. Note that all weather variables represent functions of the average daily temperature (ADT, °F) expressed as either daily cooling degrees (CD) or extended heating degrees (XHD), where these indices are in turn defined as

$$CD = \max\left[ADT - 65, 0 \right] \qquad \text{Eq. 2.1}$$

$$XHD = \max\left[55 - ADT, 0 \right]. \qquad \text{Eq. 2.2}$$

Thus, two days with average temperatures of 73.3° and 51.5° would have corresponding CD indices of 8.3 and 0 and XHD indices of 0 and 3.5, respectively. Additionally, low order Fourier frequencies are used in the regression equations to help describe structured seasonal load (or peak) variations not already explained by other predictor variables. These Fourier frequencies are formally defined as

$$Fs(n) = Sin\left[n \times 2\pi \times \{(m-0.5)/12\}\right], \qquad \text{Eq. 2.3}$$

$$Fc(n) = Cos\left[n \times 2\pi \times \{(m-0.5)/12\}\right], \qquad \text{Eq. 2.4}$$

where *m* represents the numerical month number (i.e., 1 = Jan, 2 = Feb, .., 12 = Dec). Note that low order Fourier frequencies are also used to describe seasonal variation in the residual variance component of our system (wholesale) total and peak load equations.

Table 2.1 Weather, economic and structural input variables used in RPU monthly forecasting equations (SL = system load, SP = system peak, RL = retail load(class specific)).

Effect	Variable	Definintion	Forecasting Eqns.		
			SL	SP	RL
Economic	PCPI	Per Capita Personal Income ($1000)	X	X	X
	EMP	Non-farm Employment (100,000)	X	X	X
Calendar	SumMF	# of Mon-Fri (weekdays) in month	X		
	SumSS	# of Saturdays and Sundays in month	X		
	Xmas	Retail (residential) indicator variable for Christmas effect (DEC = 1, JAN = 1.5, all other months = 0)			X
Weather	SumCD	Sum of monthly CD's	X	X	X
	SumXHD	Sum of monthly XHD's	X		X
	MaxCD3	Maximum concurrent 3-day CD sum in month		X	
	MaxHD	Maximum single XHD value in month		X	
Fourier terms	Fs(1)	Fourier frequency (Sine: 12 month phase)	X	X	X
	Fc(1)	Fourier frequency (Cosine: 12 month phase)	X	X	X
	Fs(2)	Fourier frequency (Sine: 6 month phase)	X	X	X
	Fc(2)	Fourier frequency (Cosine: 6 month phase)	X	X	X
	Fs(3)	Fourier frequency (Sine: 4 month phase)		X	
	Fc(3)	Fourier frequency (Cosine: 4 month phase)		X	
Lag function	Lag(X[i])	Produces value of X for month i-1			X

2.3. Historical and forecasted inputs: economic and weather effects

The annual values of our historical and forecasted economic indices are reported on Demand Form 2.1 in our 2011 CEC IEPR submission packet. Annual PCPI data have been obtained from the US Bureau of Economic Analysis (http://www.bea.gov), while monthly employment statistics have been obtained from the CA Department of Finance (http://www.labormarketinfo.edd.ca.gov). As previously stated, both sets of data correspond to the Riverside-Ontario-San Bernardino metropolitan service area. Graphical plots of these input data are also shown in Figures 2.1 and 2.2, respectively.

Historical SumCD and SumXHD weather indices for the Riverside service area (2003-2009) are shown in Figure 2.3. Likewise, historical MaxCD3 and MaxHD weather indices for 2004-2009 are shown in Figure 2.4. Both sets of data have been calculated from historical average daily temperature levels recorded at the UC Riverside CIMIS weather station (http://wwwcimis.water.ca.gov/cimis). Forecasted average monthly weather indices have been derived from a detailed simulation study using our dynamic time series temperature models (back-calibrated to six years of CIMIS weather data); these forecasted monthly indices are shown in Table 2.2. Note that these average monthly values are used as weather inputs for all 2011-2022 forecasts.

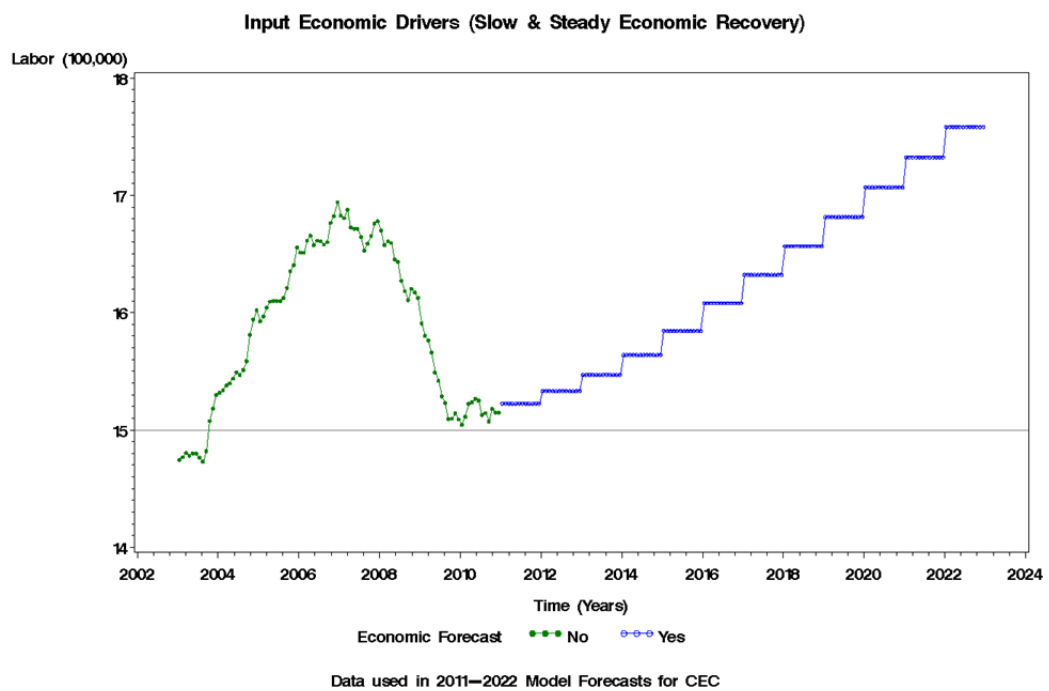

Figure 2.1. Riv/Ont/SB monthly historical employment levels (Jan 2003 - Nov 2010) and annual 2011-2022 forecasts.

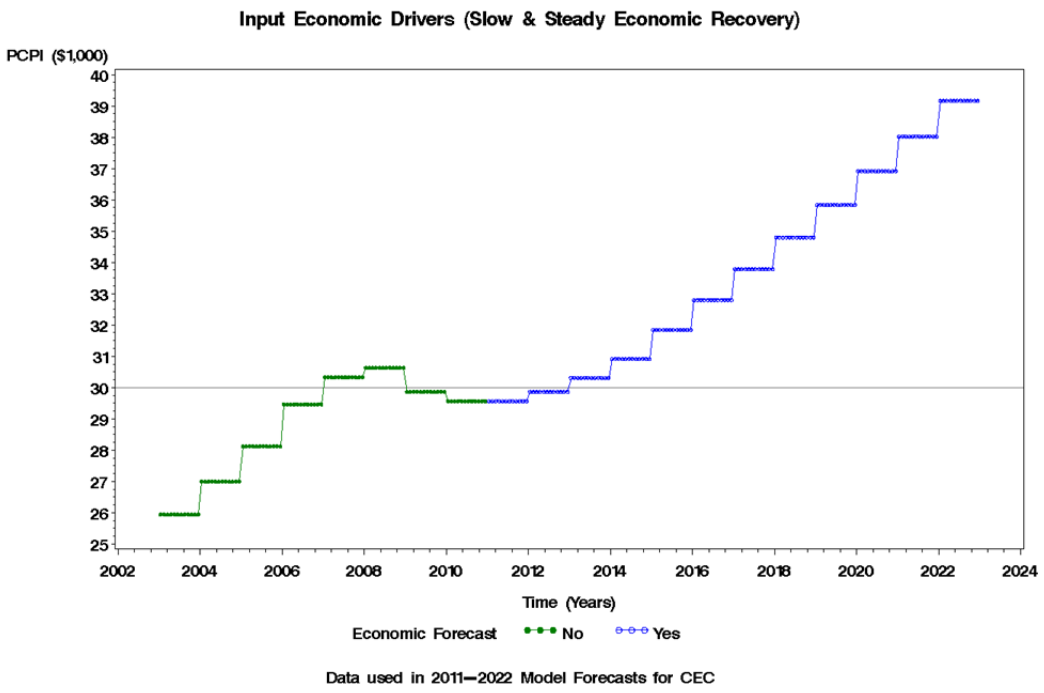

Figure 2.2. Riv/Ont/SB annual historical PCPI levels (2003-2009) and annual 2010-2022 forecasts.

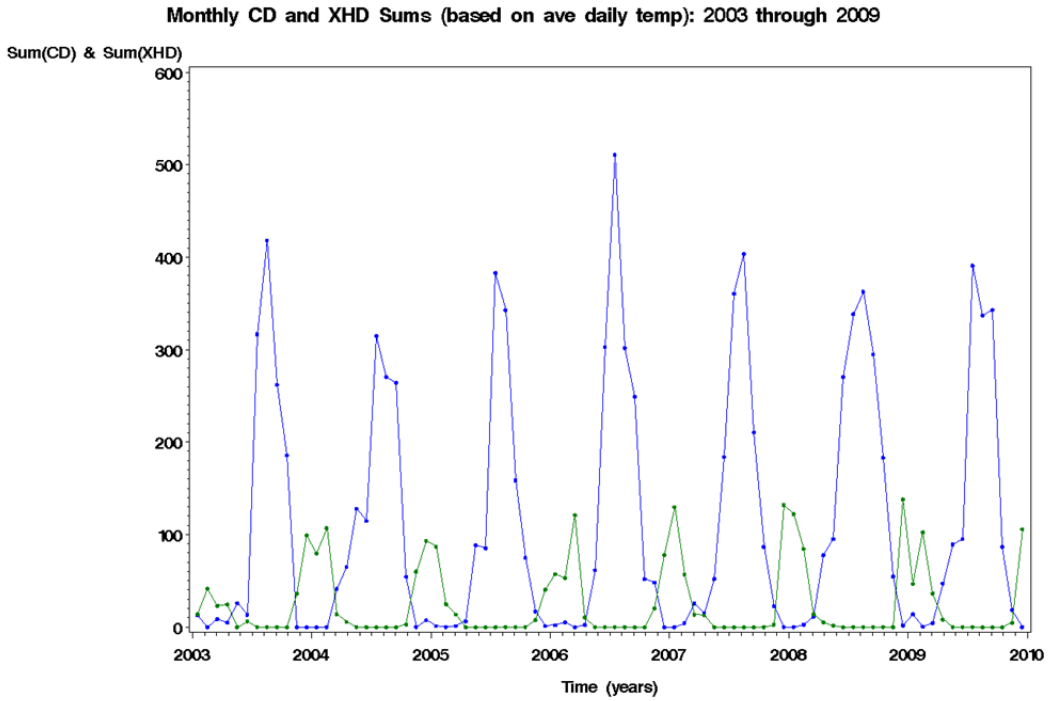

Figure 2.3. Monthly historical SumCD and SumXHD weather indices (2003-2009) for the Riverside service area.

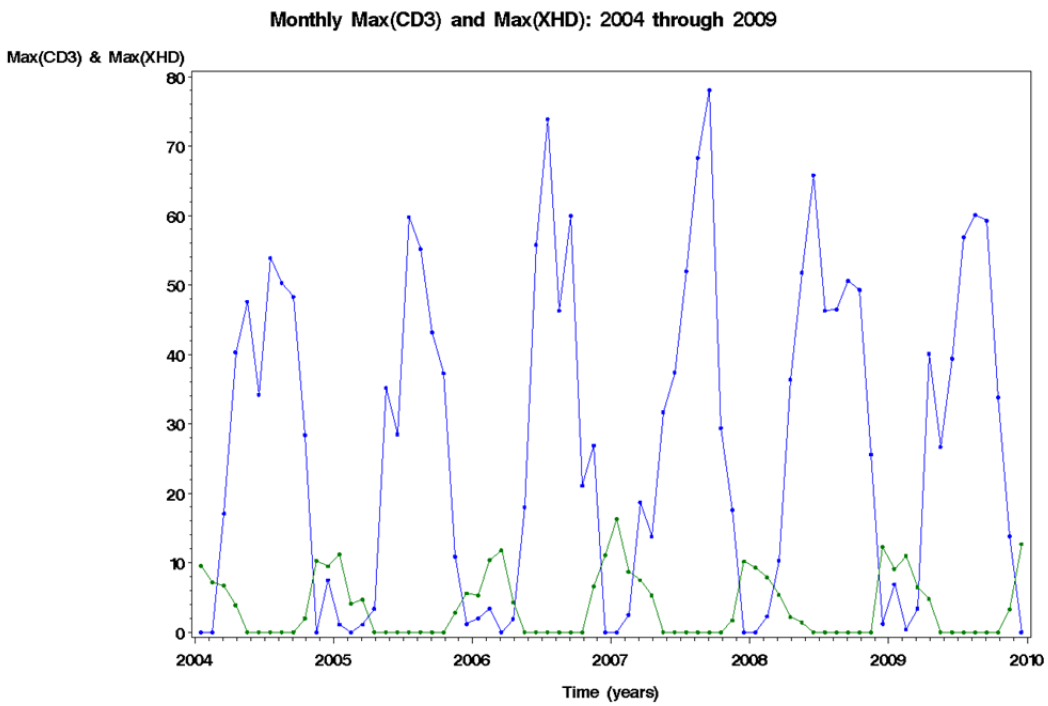

Figure 2.4. Monthly historical MaxCD3 and MaxHD weather indices (2004-2009) for the Riverside service area.

Table 2.2. Expected average values (forecast values) for 2011-2022 monthly weather indices; see Table 2.1 for weather index definitions.

Month	SumCD	SumXHD	MaxCD3	MaxHD
JAN	1.6	98.3	1.4	11.6
FEB	2.2	66.8	2.0	9.9
MAR	7.4	41.4	5.4	7.9
APR	26.8	14.4	13.9	4.6
MAY	88.7	2.1	28.2	1.1
JUN	212.1	0.1	45.5	0.1
JUL	340.8	0.0	57.0	0.0
AUG	362.4	0.0	59.8	0.0
SEP	243.7	0.1	50.2	0.0
OCT	93.0	2.7	30.9	1.3
NOV	14.6	27.4	10.4	6.7
DEC	2.7	77.1	2.5	10.4

3.0 System Load and Peak Forecast Models

3.1 Monthly system total load model

The regression component of our monthly total system load forecasting model is a function of our two economic drivers (PCPI and EMP), two calendar effects that quantify the number of weekdays (SumMF) and weekend days (SumSS) in the month, two weather effects that quantify the total monthly cooling and extended heating degrees (SumCD and SumXHD), and four low order Fourier frequencies (Fs(1), Fc(1), Fs(2) and Fc(2)). Additionally, the heterogeneous residual variance (mean square prediction error) component is defined to be a function of two low order Fourier frequencies (Fs(1) and Fc(1)). Mathematically, the model is defined as

$$\begin{aligned} y_t &= \beta_0 + \beta_1[PCPI_t] + \beta_2[EMP_t] + \beta_3[SumMF_t] + \beta_4[SumSS_t] + \beta_5[SumCD_t] \\ &\quad + \beta_6[SumXHD_t] + \beta_7[Fs(1)_t] + \beta_8[Fc(1)_t] + \beta_9[Fs(2)_t] + \beta_{10}[Fc(2)_t] + h_t\sigma^2 \end{aligned}$$

Eq. 3.1

where

$$h_t = \exp\{\alpha_1[Fs(1)_t] + \alpha_2[Fc(1)_t]\}\,. \qquad \text{Eq. 3.2}$$

In Eq. 3.1, y_t represents the RPU monthly total system load (GWh) for the calendar ordered monthly observations and forecasts (*t*=1 → Jan 2003, *t*=240 → Dec 2022) and the seasonally heterogeneous residual errors are assumed to be Normally distributed and temporally uncorrelated. Eqs. 3.1 and 3.2 were simultaneously optimized using restricted maximum likelihood estimation (SAS AutoReg Procedure).

All input observations that reference historical time periods are assumed to be fixed (i.e., measured without error) during the estimation process. For forecasting purposes, we treated the forecasted economic indices as fixed variables and the forecasted weather indices as random effects. Under such an assumption, the first-order Delta method estimate of the forecasting variance becomes

$$Var(\hat{y}_t) = \hat{h}_t\hat{\sigma}^2_{MSPE} + Var\left\{\hat{\beta}_5[SumCD_t] + \hat{\beta}_6[SumXHD_t]\right\} \qquad \text{Eq. 3.3}$$

where $\hat{\sigma}^2_{MSPE}$ represents the model calculated mean square prediction variance and the second variance term captures the uncertainty in the average weather forecasts. Note that the second variance term can be conveniently approximated via simulation, once the parameters associated with the SumCD and SumXHD weather effects have been estimated.

3.2 System load model statistics and forecasting results

Table 3.1 shows the pertinent model fitting and summary statistics for our total system load forecasting equation. The equation explains approximately 99% of the observed variability associated with the monthly 2004-2010 system loads and all input parameter estimates are statistically significant below the 0.01 significance level.

The estimates for the seasonal variance components are shown at the bottom of Table 3.1. These components define how the model mean square error (MSE) changes across the calendar months. A graphical display of the monthly root MSE estimates is shown in Figure 3.1; note that the root MSE ranges from about 2 to 4 GWh with the greatest uncertainty occurring during the Q3 summer months. This result is intuitively reasonable since summer load predictions tend to be more uncertain, even after adjusting for known weather indices. An analysis of the model residuals suggests that these errors are also Normally distributed, devoid of outliers and temporally uncorrelated; implying that our modeling assumptions are likewise reasonable.

The regression parameter estimates shown in the middle of Table 3.1 indicate that monthly system load increases as either/both weather indices increase (SumCD and SumXHD); note that an increase in one cooling degree raises the forecasted load four times as quickly as a one heating degree increase. Additionally, weekdays contribute slightly more to the monthly system load, as opposed to Saturdays and Sundays (i.e., the SumMF estimate is > than the SumSS estimate). Finally, RPU system load is expected to increase as either the area wide PCPI and/or employment indices improve over time (i.e., both economic parameter estimates are > 0).

Figure 3.2 shows the observed (blue points) versus calibrated (green line) system loads for the 2004-2010 timeframe. Nearly all of the calibrations fall within the calculated 95% confidence envelope (thin black lines) and the observed versus calibrated load correlation exceeds 0.995. Figure 3.3 shows the forecasted monthly system loads for 2011 through 2022, along with the corresponding 95% forecasting envelope. This forecasting envelope encompasses both model and weather uncertainty, while treating the projected economic indices as fixed inputs. Note that there is considerable uncertainty associated with summer forecasts due to the increased uncertainty surrounding summer weather patterns.

Table 3.2 shows the forecasted monthly RPU system loads for 2011, along with their forecasted standard deviations. Once again, these standard deviations quantify both model and weather uncertainty. The 2011 forecasts project that our annual system load should be 2,160.4 GWh, assuming that the RPU service area experiences typical weather conditions throughout the year.

Table 3.1 Model summary statistics for our monthly total system load forecasting equation.

Gross Monthly Demand Model (Jan 2003 - Nov 2010): GWh units
Forecasting Model: includes Weather & Economic Covariates (w/Fourier Effects)

Dependent Variable: GWhload Load (GWh)

Number of Observations Read	240
Number of Observations Used	95
Number of Observations with Missing Values	145

Weight: ht_1 (structured seasonal pattern)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	69011	6901.13456	765.94	<.0001
Error	84	756.84202	9.01002		
Corrected Total	94	69768			

Root MSE	3.00167	R-Square	0.9892
Dependent Mean	171.56878	Ordinary R-Sq	0.9907
Coeff Var	1.74954		

Parameter Estimates

Regression Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-147.3160	12.8003	-11.51	<.0001	0
PCPI	PCPI ($1,000)	1	2.6716	0.2368	11.28	<.0001	1.58788
Emp_CC	Labor (100,000)	1	4.4894	0.5420	8.28	<.0001	1.57977
sumMF	# Mon-Fri	1	5.3375	0.3706	14.40	<.0001	1.59262
sumSS	# Sat-Sun	1	4.9463	0.4441	11.14	<.0001	1.45406
sumCD	Sum CD's	1	0.1698	0.0075	22.66	<.0001	8.33066
sumHD	Sum XHD's	1	0.0475	0.0118	4.02	0.0001	2.87564
Fs1	Sin ff1	1	-4.4689	0.8268	-5.40	<.0001	3.73217
Fc1	Cos ff1	1	-7.2618	1.1539	-6.29	<.0001	6.78939
Fs2	Sin ff2	1	2.1398	0.6838	3.13	0.0024	2.66847
Fc2	Cos ff2	1	2.4448	0.5122	4.77	<.0001	1.47729

Variance Effect	Label	DF	Parameter Estimate	Standard Error
Fs1	Sin ff1	1	-0.2184	0.2878
Fc1	Cos ff1	1	-0.5233	0.2679

Durbin-Watson D	1.869
Number of Observations	95
1st Order Autocorrelation	0.046

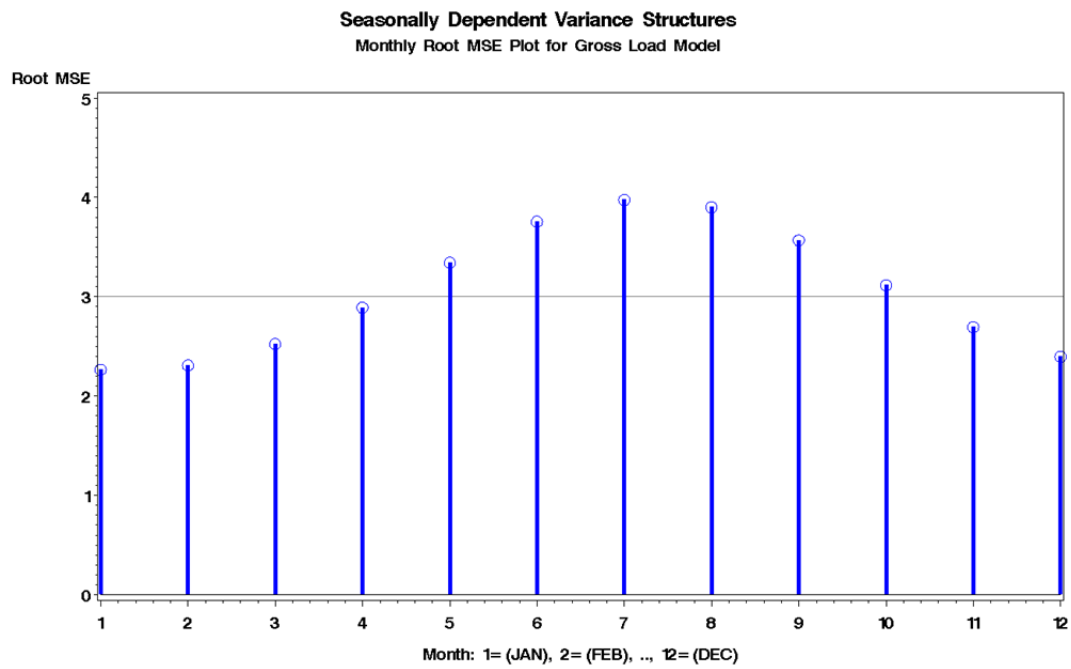

Figure 3.1. Monthly root MSE estimates for the RPU total system load equation.

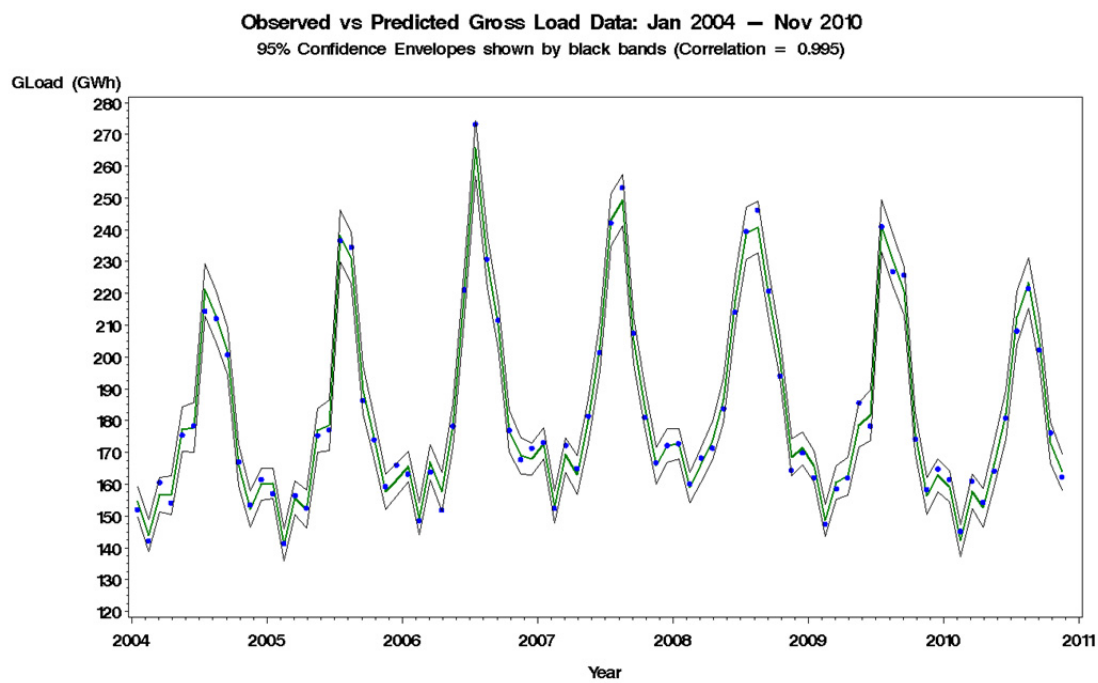

Figure 3.2. Observed and predicted total system load data (2004-2010), after adjusting for known weather conditions.

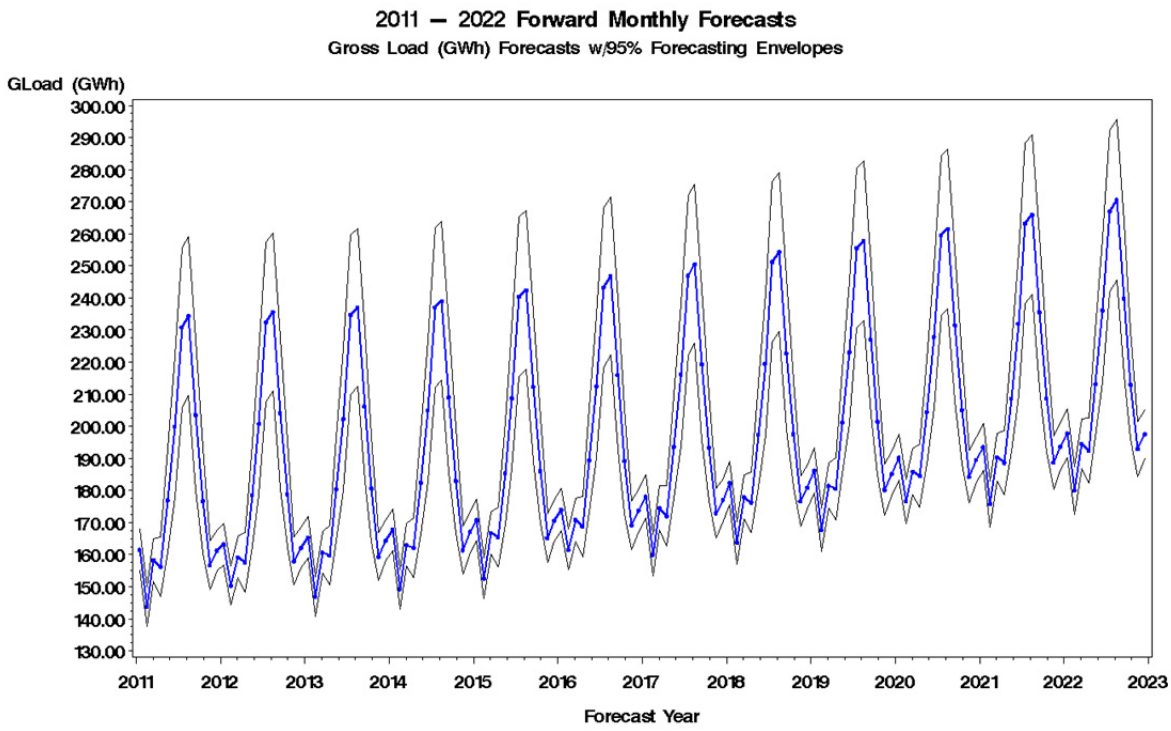

Figure 3.3. Forecasted monthly total system loads for 2011-2022; 95% forecasting envelopes encompass both model and weather uncertainty.

Table 3.2. 2011 monthly total system load forecasts for RPU; forecast standard deviations include both model and weather uncertainty.

Month	Load (GWh)	Std.Dev (GWh)
JAN	161.55	3.25
FEB	143.75	3.10
MAR	158.35	3.27
APR	156.26	4.63
MAY	176.98	8.01
JUN	199.97	11.05
JUL	230.84	12.41
AUG	234.37	12.34
SEP	203.58	11.71
OCT	176.77	8.22
NOV	156.70	3.75
DEC	161.29	3.17
Annual TOTAL	2160.41	

3.3 Monthly system peak model

The regression component of our monthly system peak forecasting model is a function of our two economic drivers (PCPI and EMP), three weather effects that quantify the total monthly cooling needs, maximum three-day cooling requirements (i.e., 3-day heat waves) and the maximum single day heating requirement (SumCD, MaxCD3 and MaxHD, respectively), and six lower order Fourier frequencies (Fs(1), Fc(1), Fs(2), Fc(2), Fs(3) and Fc(3)). Once again, the heterogeneous residual variance (mean square prediction error) component is defined to be a function of low order Fourier frequencies (four frequencies in this model: Fs(1), Fs(2), Fc(1) and Fc(2)). Mathematically, the model is defined as

$$\begin{aligned} y_t &= \beta_0 + \beta_1[PCPI_t] + \beta_2[EMP_t] + \beta_3[SumCD_t] + \beta_4[MaxCD3_t] + \beta_5[MaxHD_t] \\ &\quad + \beta_6[Fs(1)_t] + \beta_7[Fc(1)_t] + \beta_8[Fs(2)_t] + \beta_9[Fc(2)_t] + \beta_{10}[Fs(3)] + \beta_{11}[Fc(3)] + h_t\sigma^2 \end{aligned}$$

Eq. 3.4

where

$$h_t = \exp\{\alpha_1[Fs(1)_t] + \alpha_2[Fc(1)_t] + \alpha_3[Fs(2)] + \alpha_4[Fc(2)]\}. \qquad \text{Eq. 3.5}$$

In Eq. 3.4, y_t represents the RPU monthly system peaks (MW) for the calendar ordered monthly observations and forecasts (*t*=1 → Jan 2004, *t*=228 → Dec 2022) and the seasonally heterogeneous residual errors are assumed to be Normally distributed and temporally uncorrelated. Eqs. 3.4 and 3.5 were again simultaneously optimized using restricted maximum likelihood estimation (SAS AutoReg Procedure).

As in the total system load equation, all input observations that reference historical time periods were assumed to be fixed. Likewise, we again treated the forecasted economic indices as fixed variables and the forecasted weather indices as random effects. Under such an assumption, the first-order Delta method estimate of the forecasting variance becomes

$$Var(\hat{y}_t) = \hat{h}_t\hat{\sigma}^2_{MSPE} + Var\left\{\hat{\beta}_3[SumCD_t] + \hat{\beta}_4[MaxCD3_t] + \hat{\beta}_5[MaxHD_t]\right\} \qquad \text{Eq. 3.6}$$

where $\hat{\sigma}^2_{MSPE}$ represents the model calculated mean square prediction variance and the second variance term captures the uncertainty in the average weather forecasts. As before, the second variance term can be conveniently approximated via simulation after the parameters associated with the SumCD, MaxCD3 and MaxHD weather effects have been estimated.

3.4 System peak model statistics and forecasting results

Table 3.3 shows the pertinent model fitting and summary statistics for our system peak forecasting equation. The equation explains approximately 98% of the observed variability associated with the monthly 2004-2010 system peaks.

The estimates for the seasonal variance components are shown at the bottom of Table 3.3. These components define how the model mean square error (MSE) changes across the seasons. A graphical display of the monthly root MSE estimates is shown in Figure 3.4; note that the root MSE ranges from about 6 to 30 MW with the greatest uncertainty occurring during winter-to-summer and summer-to-winter transition months (May and October-November). This result is intuitively reasonable because peak predictions during transition months tend to be more uncertain, even after adjusting for known weather indices. As with the system load residuals, an analysis of the peak model residuals suggests that these errors are Normally distributed, devoid of outliers and temporally uncorrelated. Hence, our modeling assumptions again appear to be reasonable.

The regression parameter estimates shown in the middle of Table 3.3 imply that monthly system peaks increases as each of the weather indices increase (SumCD, MaxCD3 and MaxHD), but the peaks appear to be primarily determined by the MaxCD3 index. (Recall that this index essentially quantifies the maximum cooling degrees associated with 3-day summer heat waves.) RPU system peaks are also expected to increase as either the area wide PCPI and/or employment indices improve over time (i.e., both economic parameter estimates are > 0). However, we note in passing that the employment parameter estimates is only weakly significant ($p=0.1157$). Additionally, not every individual Fourier frequency parameter estimate is statistically significant, although their combined effect significantly improves the forecasting accuracy of the model ($F = 8.32$, $p =0.0001$).

Figure 3.5 shows the observed (blue points) versus calibrated (green line) system loads for the 2004-2010 timeframe. Nearly all of the calibrations fall within the calculated 95% confidence envelope (thin black lines) and the observed versus calibrated load correlation exceeds 0.989. Figure 3.6 shows the forecasted monthly system peaks for 2011 through 2022, along with the corresponding 95% forecasting envelope. This forecasting envelope again encompasses both model and weather uncertainty, while treating the projected economic indices as fixed inputs. As with the system loads, there is considerable uncertainty associated with summer peak forecasts due to the increased uncertainty surrounding summer weather patterns.

Table 3.4 shows the forecasted monthly RPU system peaks for 2011, along with their forecasted standard deviations. Once again, these standard deviations quantify both model and weather uncertainty. The 2011 forecasts project that our maximum monthly system peak should be about 561 MW and occur in August, assuming that the RPU service area experiences typical weather conditions throughout the year. Note that this represents a 1-in-2 temperature forecast, respectively.

Table 3.3 Model summary statistics for our monthly system peak forecasting equation.

Monthly Peak Load Model (Jan 2004 - Nov 2010): MW units
Forecasting Model: includes Weather & Economic Covariates (w/Fourier Effects)

Dependent Variable: Peak (MW)

Number of Observations Read	228
Number of Observations Used	83
Number of Observations with Missing Values	145

Weight: ht_2 (structured seasonal pattern)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1346352	122396	581.91	<.0001
Error	71	14934	210.33515		
Corrected Total	82	1361286			

Root MSE	14.50294	R-Square	0.9890
Dependent Mean	342.67713	Ordinary R-Sq	0.9796
Coeff Var	4.23225		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	123.3610	37.6142	3.28	0.0016	0
PCPI	PCPI ($1,000)	1	3.8610	1.2130	3.18	0.0022	1.36911
Emp_CC	Labor (100,000)	1	3.7946	2.3824	1.59	0.1157	1.34969
sumCD	Sum CD's	1	0.1452	0.0467	3.11	0.0027	21.99381
maxCD3	Max 3-day CD	1	2.8042	0.2093	13.40	<.0001	13.69524
maxHD	Max XHD	1	1.4964	0.5944	2.52	0.0141	4.26435
Fs1	Sin ff1	1	-19.9081	4.6818	-4.25	<.0001	4.69848
Fc1	Cos ff1	1	-32.1605	6.5976	-4.87	<.0001	11.99192
Fs2	Sin ff2	1	5.9068	3.7357	1.58	0.1183	2.80015
Fc2	Cos ff2	1	-3.1017	2.8611	-1.08	0.2820	2.25387
Fs3	Sin ff3	1	4.3602	2.5575	1.70	0.0926	2.02863
Fc3	Cos ff3	1	7.9161	2.2763	3.48	0.0009	1.60158

Variance Effect	Label	DF	Parameter Estimate	Standard Error
Fs1	Sin ff1	1	-0.7667	0.3585
Fc1	Cos ff1	1	-0.3409	0.3149
Fs2	Sin ff2	1	-1.0763	0.3444
Fc2	Cos ff2	1	-0.2331	0.3442

Durbin-Watson D	2.077
Number of Observations	83
1st Order Autocorrelation	-0.052

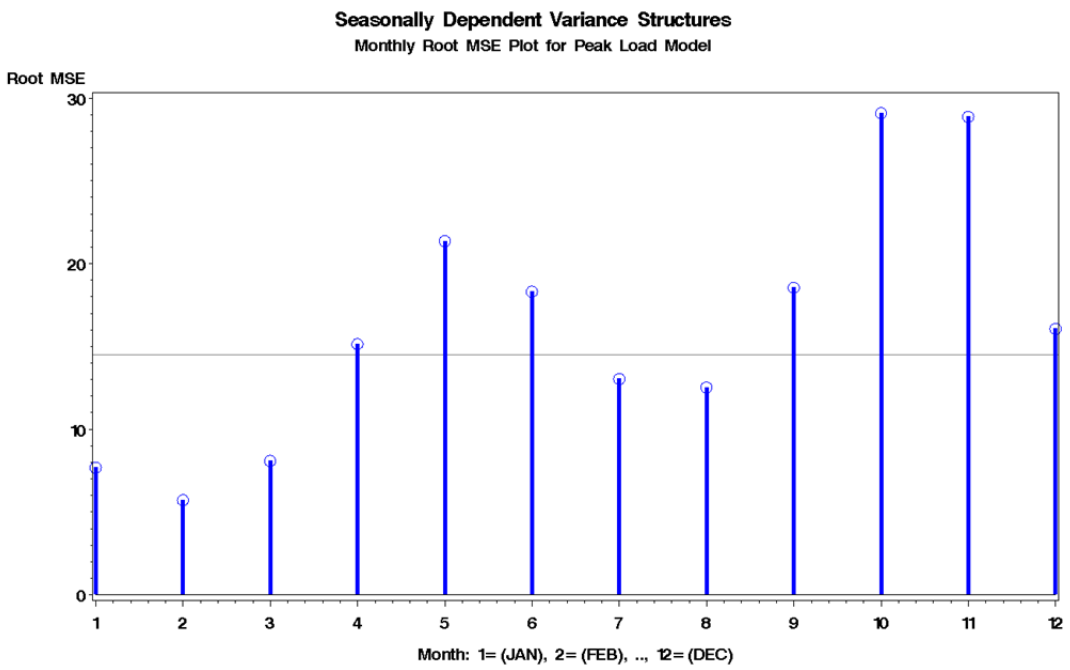

Figure 3.4. Monthly root MSE estimates for the RPU system peak equation.

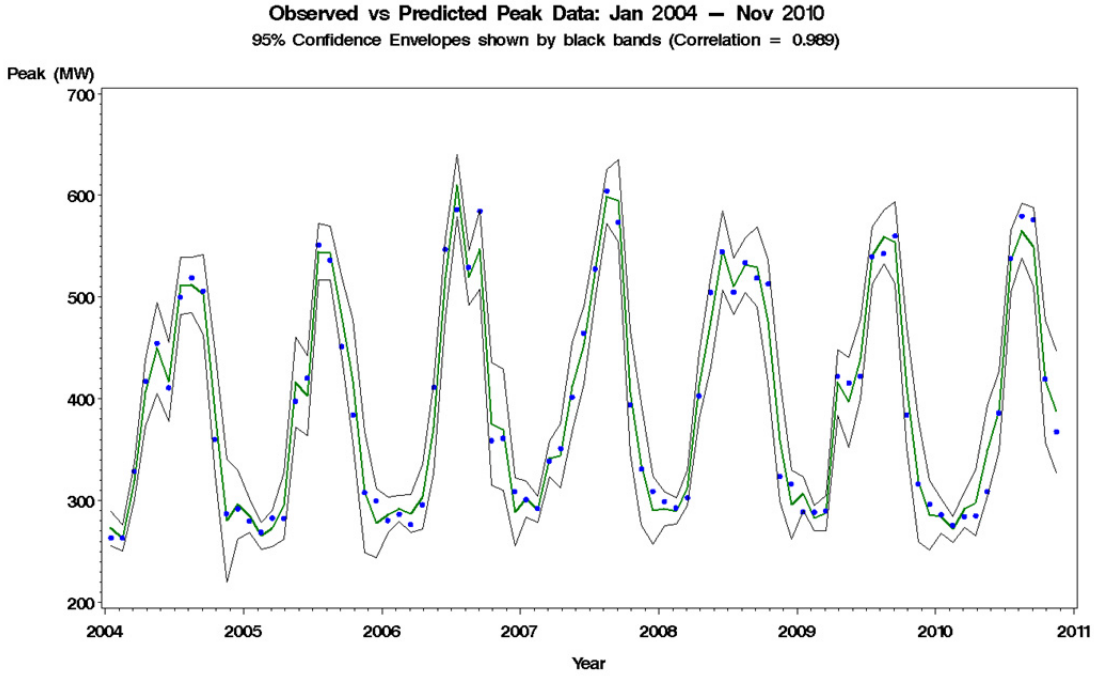

Figure 3.5. Observed and predicted system peak data (2004-2010), after adjusting for known weather conditions.

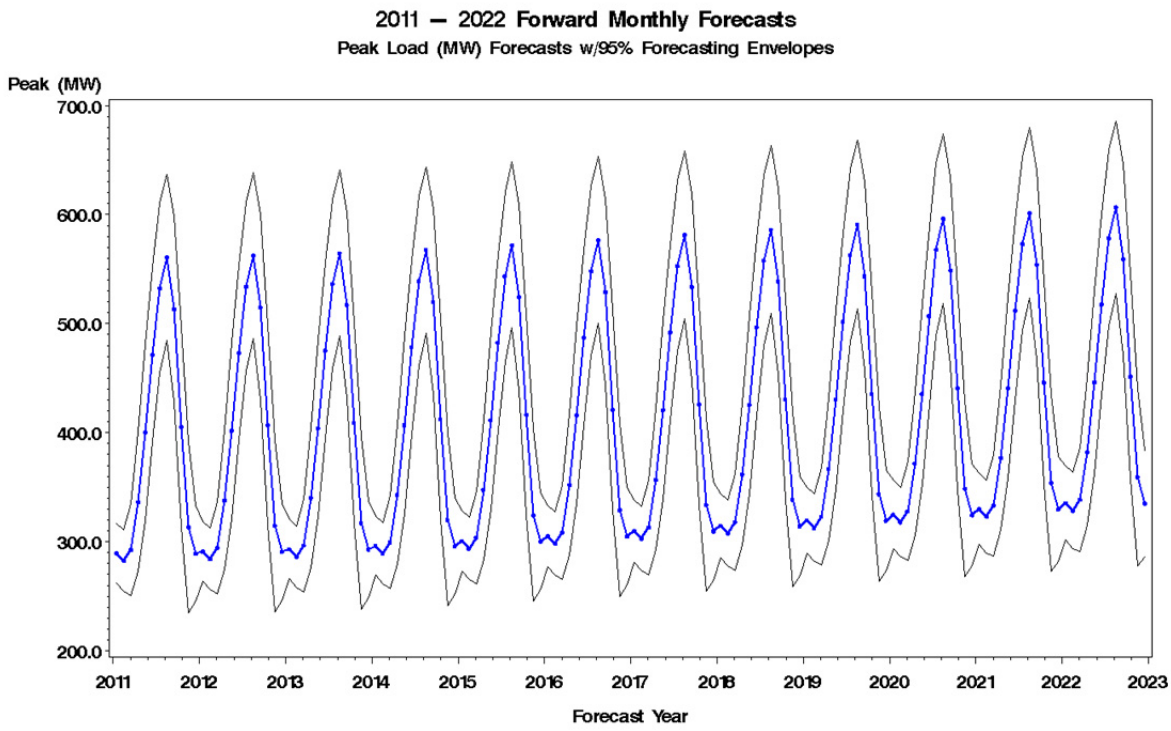

Figure 3.6. Forecasted monthly system peaks for 2011-2022; 95% forecasting envelopes encompass both model and weather uncertainty.

Table 3.4. 2011 monthly system peak forecasts for RPU; forecast standard deviations include both model and weather uncertainty.

Month	Peak (MW)	Std.Dev (MW)
JAN	289.6	13.5
FEB	282.6	14.0
MAR	292.7	21.3
APR	336.4	31.9
MAY	400.4	40.6
JUN	471.6	40.0
JUL	532.5	38.8
AUG	560.9	38.0
SEP	513.4	42.6
OCT	405.5	45.7
NOV	313.4	39.4
DEC	289.2	22.0

3.5 Peak demand weather scenario forecasts

After calculating all of the 2011-2022 monthly peak forecasts and their corresponding standard deviation estimates (that incorporate weather uncertainty), additional peak demand forecasts for more extreme weather scenarios were produced. Under the assumption that these $\hat{y}_t$ forecasts can be probabilistically approximated using a Normal distribution, the following formulas were used to calculate 1-in-5, 1-in-10, 1-in-20 and 1-in-40 forecast scenarios:

1-in-5 Peak: $\hat{y}_t + 0.842\left[Std(\hat{y}_t)\right]$ Eq. 3.7

1-in-10 Peak: $\hat{y}_t + 1.282\left[Std(\hat{y}_t)\right]$ Eq. 3.8

1-in-20 Peak: $\hat{y}_t + 1.645\left[Std(\hat{y}_t)\right]$ Eq. 3.9

1-in-40 Peak: $\hat{y}_t + 1.960\left[Std(\hat{y}_t)\right]$ Eq. 3.10

In Eqs. 3.7 through 3.10, the scale multiplier terms applied to the standard deviation represent the upper 80% (1-in-5), 90% (1-in-10), 95% (1-in-20) and 97.5% (1-in-40) quantiles of the Standard Normal distribution, respectively.

In the RPU service area, our maximum weather scenario peaks are always forecasted to occur in the month of August. Thus, for 2011, our forecasted 1-in-5, 1-in-10, 1-in-20 and 1-in-40 peaks are

560.9 + 0.842(38.0) = 592.9,

560.9 + 1.282(38.0) = 609.6,

560.9 + 1.645(38.0) = 623.4, and

560.9 + 1.960(38.0) = 635.4, respectively.

The weather scenario forecasts reported on our 2011 CEC Form 1.5 quantify these more extreme peak scenario projections through 2022.

4.0 Class-specific Retail Load Forecast Models

Our RPU retail load forecasting models are described in this section. However, before discussing each equation in detail, the following modeling issues require clarification. First, it is important to note that our retail sales data span convolved 30-day billing cycles and are subject to post-billing invoice corrections. As such, our retail load models tend to be inherently less precise and thus subject to more forecasting uncertainty. Additionally, all retail model MSPE terms are assumed to be constant (i.e., homogeneous) across the calendar year, since seasonal variance effects are difficult to identify and estimate in the presence of these increased signal-to-noise effects.

Second, RPU cannot currently analyze and estimate individual Commercial and Industrial forecasting models, because our Commercial versus Industrial classification schema was recently changed (over 2005 through 2007) by our Finance/Billing department. Instead, we have estimated a combined Commercial + Industrial load equation, produced combined forecasts using this equation and then split these forecasts into separate Commercial and Industrial predictions using a load ratio metric historically derived from our last 48 months of billing data (Jan 2007 through Dec 2010). This issue is discussed in more detail in section 4.2.

Finally, it is also important to note that we constrain the annual sum of our class specific, retail forecasts to be equal to 95% of our forecasted annual wholesale loads. (RPU internal distribution losses have averaged almost exactly 5% over the last 10 years.) This constraint is applied by determining a post-hoc, annual adjustment factor (f_R) computed as

$$f_R = \frac{0.95(W)}{(R + C + I + O)} \qquad \text{Eq. 4.1}$$

where *R*, *C*, *I* and *O* represent our forecasted annual Residential, Commercial, Industrial and Other retail loads, and *W* represents or forecasted annual wholesale system load. Our final monthly retail load forecasts are then adjusted by this annual factor (to ensure that the Eq. 4.1 constraint holds). This process is done to force our (somewhat less accurate) retail load forecasts to align with our loss adjusted system load forecasts.

4.1 Monthly residential load model (retail sales)

Our monthly residential load forecasting model is a function of one economic driver (prior month EMP), two current and prior weather effects that quantify the total monthly cooling and extended heating degrees (SumCD and SumXHD), an indicator variable that quantifies an increase in residential load due to late December / early January holiday effects, and four low order Fourier frequencies (Fs(1), Fc(1), Fs(2) and Fc(2)). Mathematically, the model is defined as

$$y_t = \beta_0 + \beta_1[lag(EMP_t)] + \beta_2[(SumCD_t + lag(SumCD_t))/2] + \beta_3[(SumXHD_t + lag(SumXHD_t))/2] + \beta_4[XMas_t] + \beta_5[Fs(1)_t] + \beta_6[Fc(1)_t] + \beta_7[Fs(2)_t] + \beta_8[Fc(2)_t] + \sigma^2 \quad \text{Eq. 4.2}$$

In Eq. 4.2, y_t represents the RPU monthly residential load (GWh) for the calendar ordered monthly observations and forecasts (*t*=1 → Jan 2003, *t*=240 → Dec 2022) and the homogeneous residual errors are assumed to be Normally distributed and temporally uncorrelated. Eq. 4.2 was optimized using ordinary least squares estimation (SAS Reg Procedure).

All input observations that reference historical time periods were assumed to be fixed (i.e., measured without error) during the estimation process. As with our wholesale models, we treated the forecasted economic indices as fixed variables and the forecasted weather indices as random effects. Thus, a first-order Delta method estimate of the forecasting variance can again be calculated in the usual manner (where the second variance term is approximated via simulation, once the parameters associated with the weather effects have been estimated).

It should be noted that Eq. 4.2 was initially defined to include both economic drivers. However, the PCPI parameter estimate was found to be clearly non-significant (see section 4.2) and thus dropped from the final forecasting equation. Likewise, the holiday effect (Xmas) was added to account for an annual residential holiday load increase that is primarily reflected in January billing statements.

4.2 Residential load model statistics and forecasting results

Table 4.1 shows the pertinent model fitting and summary statistics for our residential load forecasting equation. The equation explains 95% of the observed variability associated with the monthly 2003-2010 residential loads and nearly all input parameter estimates are statistically significant below the 0.01 significance level. An analysis of the model residuals confirms that these errors are Normally distributed, devoid of outliers and temporally uncorrelated; implying that our modeling assumptions are reasonable.

The regression parameter estimates shown in the middle of Table 4.1 indicate that monthly residential load increases as either/both weather indices increase (SumCD and SumXHD); an increase in one cooling degree raises the forecasted load about twice as quickly as a one heating degree increase. Note that averages of each current and prior month weather indices are used as input variables in the forecasting equation (to account for the delayed billing effect). RPU residential load is also expected to increase as the area wide employment levels improve over time. However, the residential load data do not show a statistically significant relationship with the PCPI index; i.e., when Eq. 4.2 was initially estimated with both economic indices, the PCPI parameter estimate produced an associated t-test score of 0.42 (p=0.6774).

Figure 4.1 shows the observed (blue points) versus calibrated (green line) residential loads for the 2003-2010 timeframe. Nearly all of the calibrations fall within the calculated 95% confidence envelope (thin black lines); the observed versus calibrated load correlation equals 0.975. Figure 4.2

shows the forecasted monthly system loads for 2011 through 2022, along with the corresponding 95% forecasting envelope. This forecasting envelope encompasses both model and weather uncertainty, while treating the projected economic indices as fixed inputs.

Table 4.2 shows the forecasted monthly RPU residential loads for 2011, along with their forecasted standard deviations. Once again, these standard deviations quantify both model and weather uncertainty. The 2011 forecasts project that our annual residential load should be 700.8 GWh, assuming that the RPU service area experiences typical weather conditions throughout the year.

Table 4.1 Model summary statistics for our monthly residential load forecasting equation.

Residential Sales Model (Jan 2003 - Nov 2010)

Dependent Variable: Residential load (GWh)

Number of Observations Read	240
Number of Observations Used	95
Number of Observations with Missing Values	145

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	21528	2691.02859	204.40	<.0001
Error	86	1132.26036	13.16582		
Corrected Total	94	22660			

Root MSE	3.62847	R-Square	0.9500
Dependent Mean	58.89114		
Coeff Var	6.16132		

Parameter Estimates

Regression Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	20.4563	8.7890	2.33	0.0223	0
lagEmpCC	lag(EMP)	1	1.5273	0.5677	2.69	0.0086	1.07982
sum2CD	[SumCD+lagSumCD]/2	1	0.0969	0.0103	9.34	<.0001	12.26015
sum2HD	[SumXHD+lagSumXHD]/2	1	0.0522	0.0278	1.88	0.0641	6.39824
xmas	XMas Effect	1	9.5376	1.4503	6.58	<.0001	3.21904
s1	Fs(1)	1	-3.7611	1.4296	-2.63	0.0101	7.44135
c1	Fc(1)	1	-4.5433	1.4905	-3.05	0.0031	7.94090
s2	Fs(2)	1	3.9835	0.9151	4.35	<.0001	3.03743
c2	Fc(2)	1	-2.5410	0.7639	-3.33	0.0013	2.09412

Durbin-Watson D	2.477
Number of Observations	95
1st Order Autocorrelation	-0.239

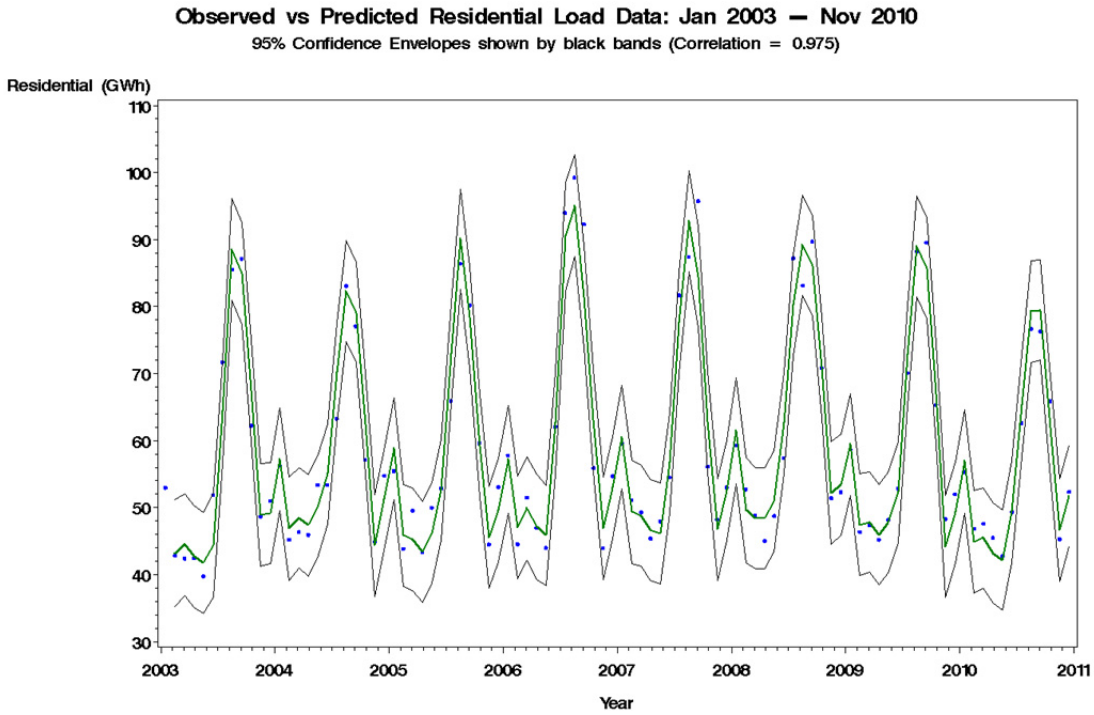

Figure 4.1. Observed and predicted residential load data (2003-2010), after adjusting for known weather conditions.

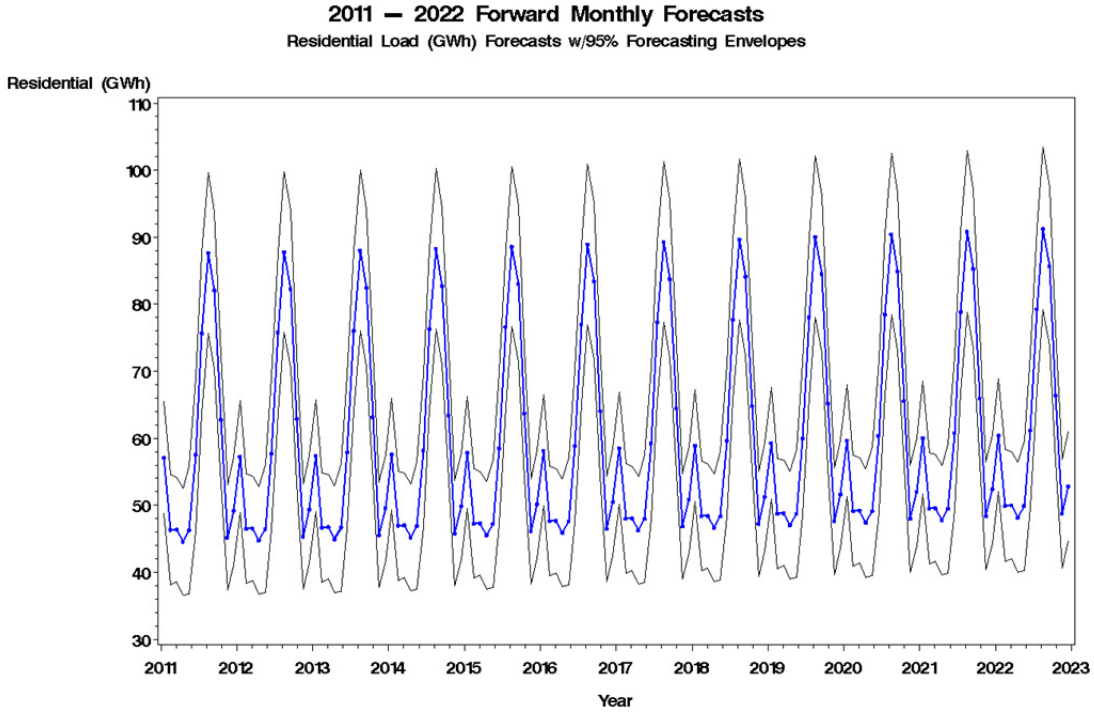

Figure 4.2. Forecasted monthly residential loads for 2011-2022; 95% forecasting envelopes encompass both model and weather uncertainty.

Table 4.2. 2011 monthly residential load forecasts for RPU; forecast standard deviations include both model and weather uncertainty.

Month	Load (GWh)	Std.Dev (GWh)
JAN	57.12	4.19
FEB	46.32	4.10
MAR	46.39	3.90
APR	44.58	4.02
MAY	46.31	4.77
JUN	57.57	5.70
JUL	75.66	5.97
AUG	87.64	6.00
SEP	82.09	5.84
OCT	62.77	4.81
NOV	45.18	3.94
DEC	49.21	4.02
Annual TOTAL	700.83	

4.3 Monthly commercial + industrial load model (retail sales)

Our composite monthly commercial + industrial load forecasting model is a function of two economic drivers (prior month PCPI and EMP), two current and prior weather effects that quantify the total monthly cooling and extended heating degrees (SumCD and SumXHD), and two low order Fourier frequencies (Fs(1) and Fc(1)). Mathematically, the model is defined as

$$\begin{aligned} y_t = {} & \beta_0 + \beta_1[lag(EMP_t)] + \beta_2[lag(PCPI_t)] + \beta_3[(SumCD_t + lag(SumCD_t))/2] \\ & + \beta_4[(SumXHD_t + lag(SumXHD_t))/2] + \beta_5[Fs(1)_t] + \beta_6[Fc(1)_t] + \sigma^2 \end{aligned}$$

Eq. 4.3

In Eq. 4.3, y_t represents the RPU combined monthly commercial + industrial load (GWh) for the calendar ordered monthly observations and forecasts (*t*=1 → Jan 2003, *t*=240 → Dec 2022) and the homogeneous residual errors are assumed to be Normally distributed and temporally uncorrelated. Eq. 4.3 was optimized using ordinary least squares estimation (SAS Reg Procedure).

Once again, all input observations that reference historical time periods were assumed to be fixed during the estimation process. Likewise, the forecasted economic indices are treated as fixed variables and the forecasted weather indices are again treated as random effects. As before, a first-order Delta method estimate of the forecasting variance can be calculated in the usual manner (where the second variance term is approximated via simulation, once the parameters associated with the weather effects have been estimated).

In order to produce individual commercial and industrial load forecasts, it is necessary to split each monthly load prediction into two components. Upon examining the ratio of the monthly commercial (C) over the commercial + industrial (C+I) loads (i.e., C/[C+I]) since Januray 2007, we found that that this ratio has only varied from 0.295 to 0.338 (mean = 0.312, standard deviation = 0.012). Thus, we have assumed that 31% of each future load forecast represents commercial load, while the remaining 69% of each forecast represents industrial load. This simple post-hoc calculation facilitates the prediction of separate commercial and industrial retail load metrics, respectively.

4.4 Commercial + Industrial load model statistics and forecasting results

Table 4.3 shows the pertinent model fitting and summary statistics for our commercial + industrial load forecasting equation. The equation explains approximately 88.7% of the observed variability associated with the monthly 2003-2010 C+I loads. Note that although the heating degree effect is non-significant (t = 1.24, p=0.2199), we've elected to retain this weather variable in the equation. (Intuitively, a positive heating degree effect is both reasonable and expected.) Note also that an analysis of the model residuals confirms that these errors are Normally distributed, devoid of outliers and temporally uncorrelated.

The regression parameter estimates shown in the middle of Table 4.3 indicate that monthly residential load increases as either/both weather indices increase (SumCD and SumXHD); once again however, the heating degree effect cannot be judged to be statistically significant. As in the residential model, averages of each current and prior month weather indices are used as input variables in the forecasting equation (to account for the delayed billing effect). RPU C+I loads are also expected to increase as either/both the area wide PCPI and/or employment levels improve over time. Additionally, the impact of these estimated economic driver effects appear to be much more pronounced in this C+I equation, as opposed to the residential equation.

Figure 4.3 shows the observed (blue points) versus calibrated (green line) C+I loads for the 2003-2010 timeframe. Once again, nearly all of the calibrations fall within the calculated 95% confidence envelope (thin black lines); the observed versus calibrated load correlation equals 0.942. Figure 4.4 shows the forecasted monthly C+I loads for 2011 through 2022, along with the corresponding 95% forecasting envelope. This forecasting envelope encompasses both model and weather uncertainty, while treating the projected economic indices as fixed inputs. Note that our C+I loads are forecasted to grow at nearly a 2% annual rate, once the local economy fully recovers.

Table 4.4 shows the post-hoc forecasted monthly commercial and industrial loads for 2011, along with their forecasted standard deviations. Once again, these standard deviations quantify both model and weather uncertainty. The 2011 forecasts project that our annual commercial and industrial loads should be 408.9 and 910.0 GWh, respectively, assuming that the RPU service area experiences typical weather conditions throughout the year and that the 31%/69% commercial/industrial load pattern continues to hold.

Table 4.3 Model summary statistics for our monthly commercial + industrial load forecasting equation.

Commercial / Industrial Sales Model (Jan 2003 - Nov 2010)

Dependent Variable: Comm+Indst (GWh)

Number of Observations Read	240
Number of Observations Used	95
Number of Observations with Missing Values	145

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	15550	2591.67867	115.65	<.0001
Error	88	1972.10683	22.41030		
Corrected Total	94	17522			

Root MSE	4.73395	R-Square	0.8875
Dependent Mean	108.81482		
Coeff Var	4.35047		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	-22.8447	11.7324	-1.95	0.0547	0
lagPCPI	lag(PCPI)	1	3.1678	0.3809	8.32	<.0001	1.52735
lagEmpCC	lag(EMP)	1	2.0739	0.8827	2.35	0.0210	1.53348
sum2CD	[SumCD+lagSumCD]/2	1	0.0569	0.0087	6.53	<.0001	5.08757
sum2HD	[SumXHD+lagSumXHD]/2	1	0.0326	0.0264	1.24	0.2199	3.37087
s1	Fs(1)	1	-5.4814	1.4705	-3.73	0.0003	4.62509
c1	Fc(1)	1	-4.6928	1.5659	-3.00	0.0035	5.14902

Durbin-Watson D	2.567
Number of Observations	95
1st Order Autocorrelation	-0.284

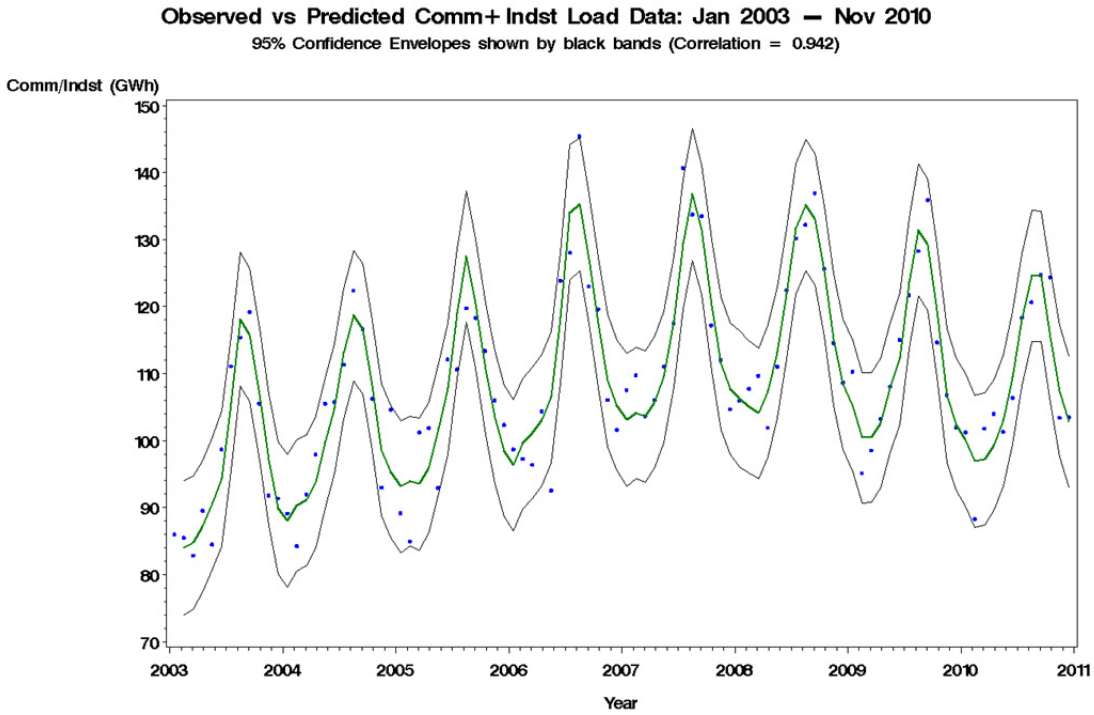

Figure 4.3. Observed and predicted C+I load data (2003-2010), after adjusting for known weather conditions.

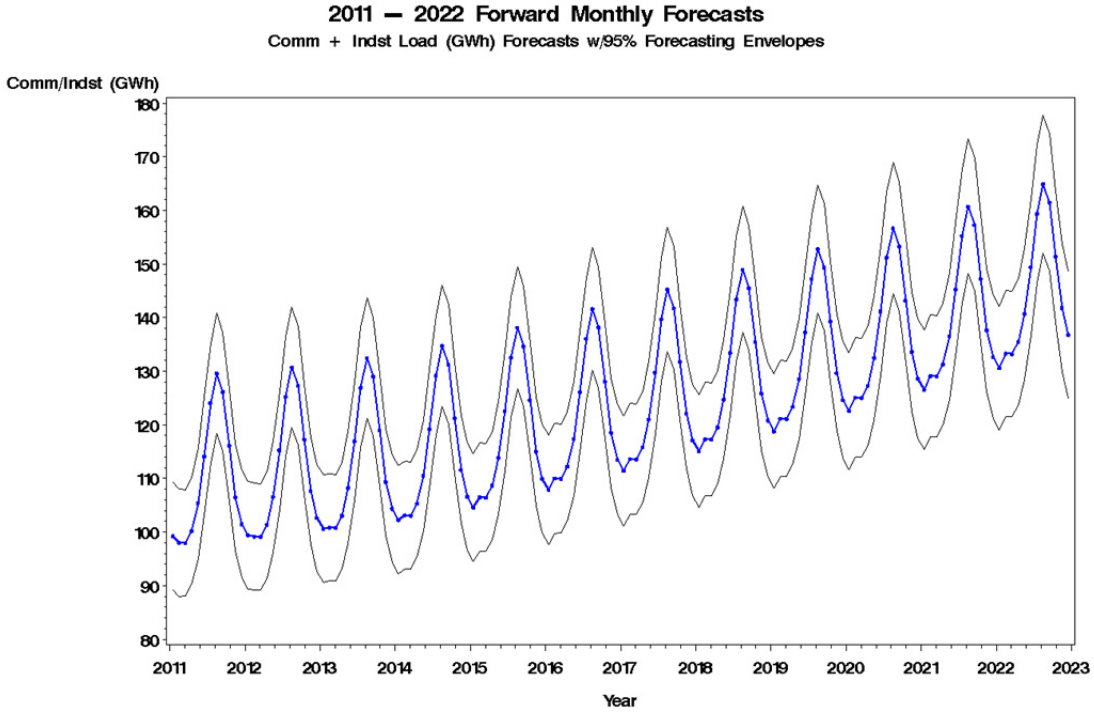

Figure 4.4. Forecasted monthly C+I loads for 2011-2022; 95% forecasting envelopes encompass both model and weather uncertainty.

Table 4.4. 2011 monthly commercial and industrial load forecasts for RPU; forecast standard deviations include both model and weather uncertainty.

Month	Comm Load (GWh)	Std.Dev (GWh)	Indst Load (GWh)	Std.Dev (GWh)
JAN	30.78	1.56	68.50	2.72
FEB	30.38	1.55	67.63	2.62
MAR	30.36	1.53	67.57	2.53
APR	31.06	1.53	69.14	2.47
MAY	32.67	1.60	72.73	2.48
JUN	35.37	1.70	78.74	2.57
JUL	38.47	1.73	85.62	2.73
AUG	40.18	1.74	89.42	2.87
SEP	39.11	1.72	87.05	2.94
OCT	35.99	1.61	80.11	2.93
NOV	33.02	1.54	73.49	2.87
DEC	31.46	1.54	70.02	2.81
Annual TOTAL	408.85		910.02	

4.5 Modeling and forecasting results for the Other customer class

All remaining RPU customers not classified into one of our three primary customer classes (residential, commercial and industrial) have historically been grouped into an "Other" class. The loads associated with this class currently account for about 1.5% of our total retail load; note that this class is primary comprised of city accounts, street lighting and miscellaneous agricultural customers.

Since January 2007, the monthly loads associated with the Other customer class have exhibited a fairly stable, seasonal pattern that appears to be independent of changing economic conditions. However, this pattern does show a marginal relationship with the observed monthly cooling degrees (SumCD), outside of one obvious outlier month (January 2009). As such, our load forecasting model for this customer class was defined to be a function of the current and prior month cooling degrees, two low order Fourier frequencies (Fs(1) and Fc(1)), and one indicator variable to account for the January 2009 outlier. The corresponding model estimation results (derived using ordinary least squares) are shown in Table 4.5; note that this equation describes about 70% of the observed load variation.

Table 4.6 shows the monthly load forecasts for 2011 along with their forecasted standard deviations. As with all previous forecasts, these standard deviations quantify both model and weather uncertainty. However, the weather uncertainty in these forecasts is minimal, since the estimated weather effect is quite trivial. Also, these forecasts do not grow over time, since the forecasting equation for this latter customer class includes no economic driver variables.

Table 4.5 Model summary statistics for our monthly “other” load forecasting equation.

Other (Non-RCI) Sales Forecasts (Jan 2007 - Dec 2010)

Dependent Variable: All Other Class (GWh)

Number of Observations Read	192
Number of Observations Used	48
Number of Observations with Missing Values	144

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1.74244	0.43561	24.74	<.0001
Error	43	0.75699	0.01760		
Corrected Total	47	2.49942			

Root MSE	0.13268	R-Square	0.6971
Dependent Mean	2.72214		
Coeff Var	4.87415		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	2.6413	0.0464	56.88	<.0001	0
sum2CD	[SumCD+lagSumCD]/2	1	0.0006	0.0004	1.64	0.1092	5.66498
s1	Fs(1)	1	-0.1555	0.0534	-2.91	0.0057	3.89845
c1	Fc(1)	1	0.1249	0.0457	2.73	0.0091	2.84933
outlier	{Jan 2009}	1	0.5495	0.1372	4.00	0.0002	1.04783

Durbin-Watson D	1.698
Number of Observations	48
1st Order Autocorrelation	0.145

Table 4.6. 2011 monthly other customer class load forecasts for RPU; forecast standard deviations include both model and weather uncertainty.

Month	Load (GWh)	Std.Dev (GWh)
JAN	2.72	0.14
FEB	2.62	0.14
MAR	2.53	0.14
APR	2.47	0.14
MAY	2.48	0.14
JUN	2.57	0.14
JUL	2.73	0.15
AUG	2.87	0.15
SEP	2.94	0.15
OCT	2.93	0.14
NOV	2.87	0.14
DEC	2.81	0.14
Annual TOTAL	32.54	

4.6 Final post-hoc forecasting alignment

As described earlier at the beginning of section 4, a post-hoc correction factor was applied to all retail forecasts. This correction factor (calculated via Eq. 4.1.) was used to constrain the annual sums of our retail load forecasts to equal our (loss adjusted) system load forecasts. These annual adjustment factors ranged from 1.000 (2011) to 0.981 (2022), respectively.

Our final annual, class-specific adjusted retail forecasts are reported on Demand Form 1.b in our 2011 CEC IEPR submission packet. The monthly 2011-2022 forecasts for our three primary retail customer classes are also shown in Figure 4.5. Note that two general features are apparent. First, our forecasted residential loads exhibit a much more pronounced reaction to summer temperature effects. This pattern reflects the increased load associated with running residential air conditioning units during the June-September summer season in the RPU service territory. Second, the forecasted 10-year load growths associated with our commercial and industrial classes are significantly higher than our forecasted residential load growth. Assuming that the local economy fully recovers, there is a much greater potential for increased commercial and industrial growth in our service territory. The potential for new residential development is far more restricted, given current Riverside City zoning regulations (and City Council adopted slow-growth initiatives).

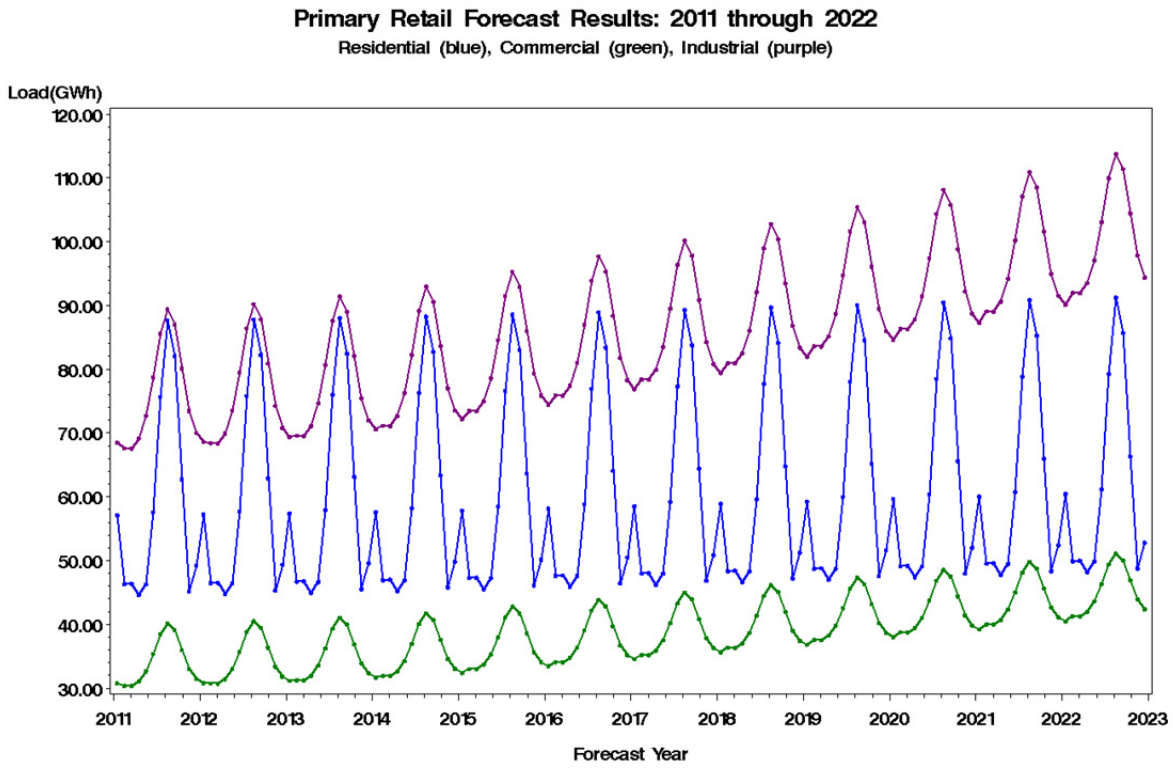

Figure 4.5. RPU monthly retail load forecasts (Jan 2011 - Dec 2022) for the residential, commercial and industrial customer classes.