Electric load forecasting with recency effect: A big data approach

Pu Wang[1]
Bidong Liu[2]
Tao Hong[2]

[1] SAS - R&D, Cary, NC, USA
[2] Energy Production and Infrastructure Center, University of North Carolina at Charlotte, USA

Electric Load Forecasting with Recency Effect: a Big Data Approach[1]

Pu Wang, Bidong Liu and Tao Hong

Abstract: Temperature plays a key role in driving electricity demand. We adopt "recency effect", a term originated from psychology, to denote the fact that electricity demand is affected by the temperatures of preceding hours. In the load forecasting literature, the temperature variables are often constructed in the form of lagged hourly temperatures and moving average temperatures. Over the past decades, computing power has been limiting the amount of temperature variables that can be used in a load forecasting model. In this paper, we present a comprehensive study on modeling recency effect through a big data approach. We take advantage of the modern computing power to answer a fundamental question: *how many lagged hourly temperatures and/or moving average temperatures are needed in a regression model to fully capture recency effect without compromising the forecasting accuracy?* Using the case study based on data from the load forecasting track of the Global Energy Forecasting Competition 2012, we first demonstrate that a model with recency effect outperforms its counterpart (a.k.a., Tao's Vanilla Benchmark Model) in forecasting the load series at the top (aggregated) level by 18% to 21%. We then apply recency effect modeling to customize load forecasting models at low level of a geographic hierarchy, again showing the superiority over the benchmark model by 12% to 15% on average. Finally, we discuss four different implementations of the recency effect modeling by hour of a day.

1. Introduction

At the inception of electric power systems, lighting was the only end use of electricity. As a result, electricity demand was primarily driven by calendar variables. As more and more electricity-powered appliances were invented, the end use became diversified. Increasing penetration of electrical air conditioning systems made the role of weather more and more important in driving electricity demand. Back to 1940s, people realized that electric load was highly dependent on weather (Dryar, 1944).

In the pre-PC (personal computer) era, utility planners and operators created lookup tables and charts, based on historical data and past experience, to capture the relationship between load and weather variables such as temperature and humidity. They then used these charts and tables together with rulers and intuitions to forecast load (Hong, 2014).

When people started using computers for load forecasting in 1980s, the computing power was very limited. Quite often, the model building procedures had to be conducted offline to select variables and estimate parameters, so that the computer can perform online calculation of the load forecast based on the new data and previously calculated variables and parameters (Gross & Galiana, 1987). The offline model building scheme means that the model(s) wouldn't be updated in real time to reflect the most recent status of the power system. Consequently, the forecasting accuracy was more or less compromised.

The technological advancement through late 1990s quickly eliminated the needs of doing offline computation for many load forecasting techniques. Also due to the benefits of the increasing computing capability, people started to apply some of the computation intensive techniques to load forecasting such as Artificial Neural Networks (Hippert, Pedreira, & Souza, 2001) and autoregressive integrated moving average models (Weron 2006). Meanwhile, people started using large amount of variables in load forecasting models (Hippert et al., 2001).

Recency effect in psychology refers to the fact that human beings tend to remember the most recent items. The power grid has a similar character - its demand tends to be significantly affected by the recent temperatures. (Hong, 2010) first adopted this term to illustrate part of a systematic load forecasting methodology, which took lagged temperatures to enhance load forecasting accuracy of a benchmark model. Since then, this term "recency effect" has been widely accepted in the US utility industry, and is part of a commercial software package (SAS® Energy Forecasting) which is currently being used by many power companies worldwide. Note that many papers in the load forecasting literature reported the usage of lagged temperatures. (Papalexopoulos & Hesterberg, 1990) used lagged temperature to calculate lagged heating and cooling degree days for regression models. A Global Energy Forecasting Competition 2012 (GEFCom2012) winning team (Ben Taieb & Hyndman, 2014) used lagged hourly temperature and average daily temperature variables in the competition. Another GEFCom2012 winning team (Nedellec, Cugliari, & Goude, 2014) used exponentially smoothed temperature variables in GEFCom2012. Nevertheless, there has never been an in-depth study stretching the model size to investigate whether large number of lagged and average temperature variables can help improving forecast accuracy.

Despite the big relief on the computation side, model building can still take significant amount of time if one wants to test many variables. Sometimes people still have to juggle the tradeoff between the frequency of model updates and sufficiency of the variables. As discussed in (Hong, 2010), for example, lagged temperature variables were limited to the past three hours of temperatures due to the consideration

of computational constraints. The limitation of computing power constrains were also an issue in the earlier decades as discussed in (Gross & Galiana, 1987).

The methodology presented in this paper is a continuation and extension of (Hong, 2010). We are attempting to take advantage of the modern computing power to answer a fundamental question: *How many lagged hourly temperatures and/or daily moving average temperatures are needed in a regression model to fully capture recency effect without compromising the forecasting accuracy*? Note that we are not exploring the exponentially weighted temperature variables in this paper. The primary reason is to avoid discussing algorithms for fine-tuning the exponential weights. For instance, one heuristic method to select base for the exponential weights was discussed in (Hong, 2010). Other notable discussions about exponential smoothing for electric load forecasting can be found in (Weron, 2006) and (Taylor & McSharry, 2007). The lagged temperatures and moving average temperatures covered in this paper can be regarded as a typical representation of "recency", in the sense that we assign the weight of one to each observation in the moving window and zero weight to the observations outside the moving window. Nevertheless, the proposed framework does not exclude the usage of exponentially weighted temperatures.

Following (Hong, Wang, & White, 2015), we develop the case study based on the load forecasting data of GEFCom2012 published in (Hong, Pinson, & Fan, 2014). There are two big data aspects in this paper: 1) we customize the model for each zone of a geographic hierarchy and each hour of the day; and 2) we leverage modern computing power to develop large load forecasting models with thousands of variables. More discussions about big data in load forecasting will be presented in Section 4.2.

This paper has the following significant contributions: 1) this is the first comprehensive study on modeling recency effect without computational constraints; 2) this is the first time the recency effect is being applied to hierarchical load forecasting, in the contents of both geographical and temporal hierarchies, where recency effect is being modeled in a customized fashion for each zone and each hour of a day; and 3) publicly available data is used to conduct the case study, so that future researchers can reproduce our results.

2. Background

2.1 Data description

One of the objectives of GEFCom2012 was to establish a benchmarking data pool for researchers in the energy forecasting community to compare models. In this paper, we use the data from the hierarchical load forecasting track of GEFCom2012, including 4.5 years of hourly load and temperature across 21 zones (Z_i, i = 1, 2, ...) of a U.S. Utility, of which Z_{21} was the sum of the first 20 zones (Hong, Pinson, & Fan, 2014).

Same as the case study in (Hong, Wang, & White, 2015), we are conducting out-of-sample tests instead of using the methods based on in-sample fit to select models. Here we slice the data to three pieces, the first two years (2004-2005) for training (or in-sample fit, for parameter estimation), the next year (2006) for validation (or post-sample fit, for model selection), and the last full calendar year (2007) for testing (or out-of-sample test, for summary of error statistics). The four years of Z_{21} load data is shown in Figure 1.

We also use the weather station selection and load forecasting results of (Hong, Wang, & White, 2015) as the benchmark for the methodology proposed in this paper. In other words, actual temperature observations are used through the forecast horizon, which is widely used in the industry to develop load forecasting models as discussed in (Nedellec et al., 2014). A more in-depth discussion about weather variables in ex ante vs. ex post load forecasting was presented in (Hong, Wang, & White, 2015).
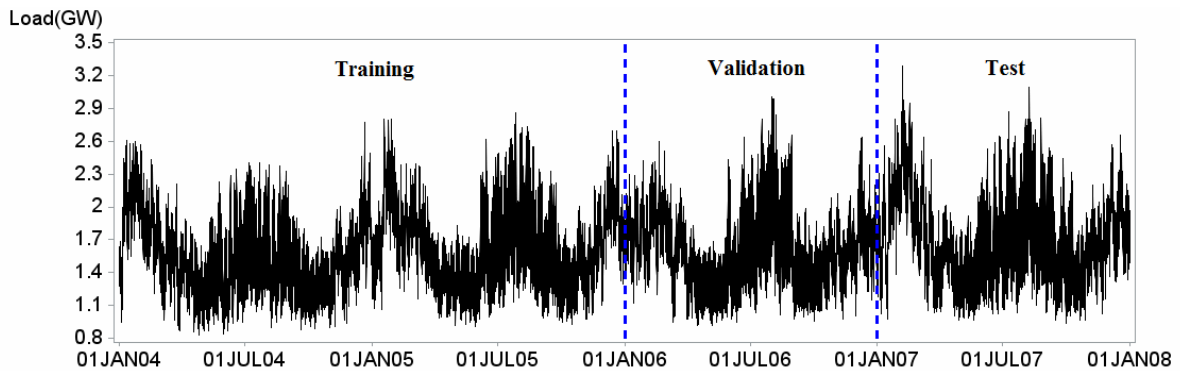

Figure 1. Data from GEFCom2012 load forecasting track.

2.2 Tao's Vanilla Benchmark model

Tao's Vanilla Benchmark model was first proposed in (Hong, 2010) and then used to produce the benchmark score for GEFCom2012 (Hong, Pinson, et al., 2014). In GEFCom2012, this model was ranked at top 25% among over 100 teams. It has also been implemented as a base model in a commercial software package SAS Energy Forecasting. The same model was used in the weather station selection framework proposed in (Hong, Wang, & White, 2015). The model is a multiple linear regression model:

$$y_t = \beta_0 + \beta_1 Trend + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t), \tag{1}$$

where y_t is the load forecast for time (hour) t; β_i are the coefficients estimated using the ordinary least square method; M_t, W_t and H_t are the month-of-the-year, day-of-the-week, and hour-of-the-day classification variables corresponding to time t, respectivey; T_t is the temperature at time t; and

$$f(T_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t, \tag{2}$$

We use temperature and calendar variables only, but not past loads, in this model and the models to be discussed in later sections, mainly because in many jurisdictions, past loads are not allowed in long term load forecasting to avoid reducing the interpretability of the model. Due to the simplicity and wide adoption of regression analysis in load forecasting, we used multiple linear regression models to convey the ideas of recency effect modeling. Nevertheless, recency effect can also be modeled through other techniques, such as neural networks and support vector regression models.

2.3 Missing salient features

Using two years of data (2005 – 2006) to estimate the parameters of the benchmark model, we can perform one-year ahead ex post forecasting for the 8760 hours in year 2007. Figure 2 shows the actual load and forecasted load from Tao's Vanilla Benchmark model, overlaid with temperature, for a summer week and a winter week. We can observe the following discrepancies between the forecast and actual load:

1) The benchmark model over-forecasts the summer peaks for consecutive days (6/22-24) and under-forecasts the winter peaks for consecutive days (11/15-20);
2) The benchmark model under-forecasts the troughs for consecutive days (11/16-20);
3) The forecast from the benchmark model is leading or lagging the actual load for consecutive hours in several days (6/20, 6/23, 6/24, 11/16-19).

In addition, we can also observe the lag between temperature and load profiles in the summer, which indicates that the lack of temperatures of preceding hours in the load forecasting model may be a reason causing some of the discrepancies mentioned above. This further motivates the study of recency effect. The same two weeks will be plotted later in Figure 4 to show improvement via modeling recency effect.
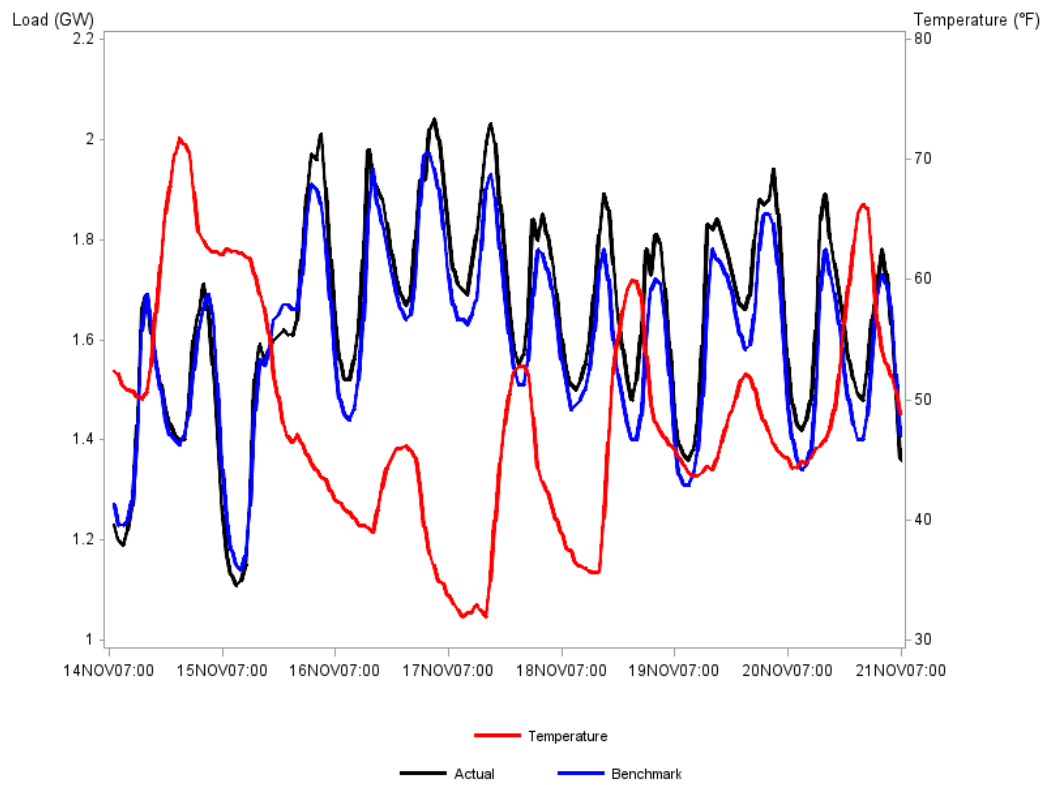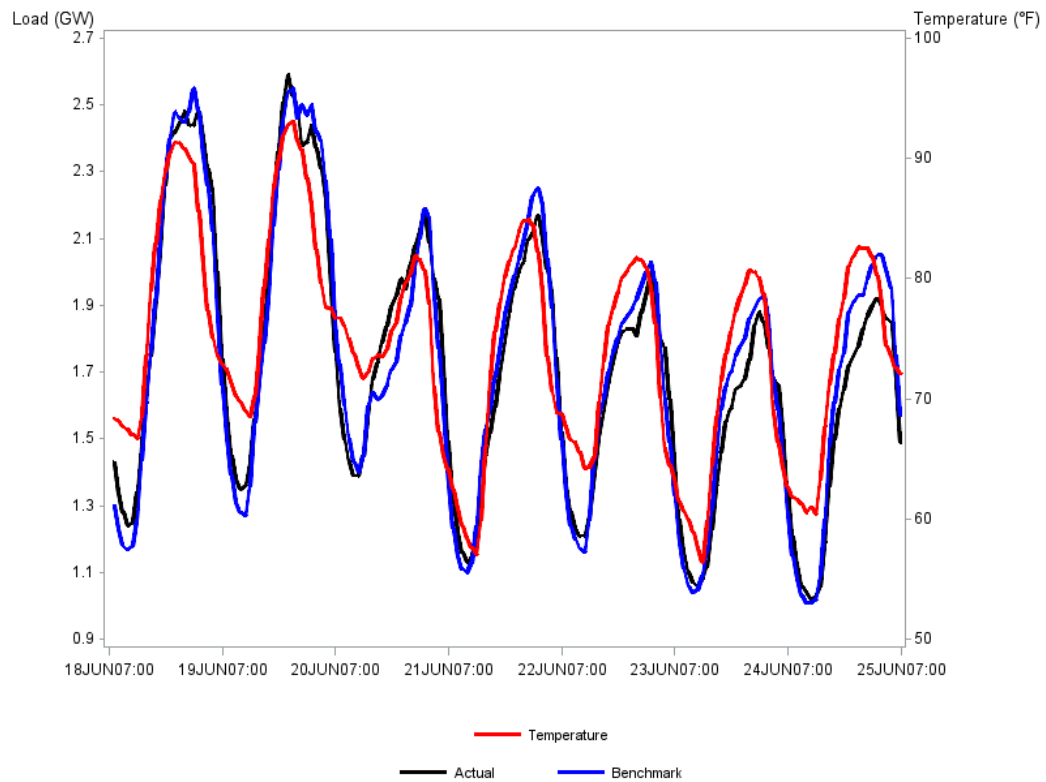
Figure 2. Actual load, forecasted load and temperature plots for a summer week (upper) and a winter week (lower).

3. Recency effect

3.1 At aggregated level

To capture the missing salient features discussed above, we are going to include the temperatures of preceding hours in the model. In this paper, we choose the two most commonly used forms of temperature variables, lagged hourly temperatures (T_{t-h}, $h = 1, 2, \ldots 72$) and 24-hour (or daily) moving average temperatures, where the daily moving average temperature of the dth day can be written as:

$$\tilde{T}_{t,d} = \frac{1}{24} \sum_{h=24d-23}^{24d} T_{t-h} , d = 1, 2, \ldots 7, \tag{3}$$

Then Tao's Vanilla Benchmark model can be extended to:

$$y_t = \beta_0 + \beta_1 Trend + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) + \underbrace{\sum_d f(\tilde{T}) + \sum_h f(T_{t-h})}_{recency\ effect\ terms} \tag{4}$$

We take a trial-and-error method to model the recency effect. We vary the number of days from 0 to 7, and the number of lags from 0 to 72. In total, there are 584 (73 by 8) possible "average-lag" (or d-h) pairs.

In this paper, we again use Mean Absolution Percentage Error (MAPE) of hourly loads as the error measure to evaluate load forecasts:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{5}$$

where A_t is the actual value and F_t is the forecasted value at time t.

Enumerating all 584 models based on different d-h pairs for Z_{21}, we can obtain the heat map corresponding to the MAPE values on the validation data (the year of 2006) as shown in Figure 3, which shows the first 48 lags to avoid verbose presentation. The lowest MAPE is 3.54%, or more precisely, 3.542%, which occurs on the model with 2 daily moving average temperatures and 12 lagged hourly temperatures.

h \ d	0	1	2	3	4	5	6	7
0	4.89	4.10	4.09	4.16	4.23	4.13	4.12	4.25
1	4.55	3.92	3.91	3.97	4.05	3.95	3.96	4.08
2	4.34	3.81	3.80	3.87	3.94	3.85	3.85	3.97
3	4.20	3.76	3.74	3.81	3.88	3.79	3.79	3.91
4	4.09	3.71	3.70	3.76	3.84	3.75	3.74	3.86
5	4.00	3.68	3.67	3.73	3.81	3.72	3.71	3.83
6	3.93	3.65	3.64	3.71	3.78	3.70	3.69	3.81
7	3.86	3.63	3.62	3.69	3.76	3.68	3.67	3.80
8	3.81	3.60	3.60	3.67	3.75	3.67	3.66	3.78
9	3.77	3.59	3.58	3.65	3.73	3.66	3.65	3.78
10	3.74	3.58	3.57	3.64	3.72	3.65	3.64	3.77
11	3.73	3.57	3.55	3.63	3.71	3.64	3.63	3.76
12	3.71	3.56	**3.54**	3.62	3.69	3.63	3.62	3.75
13	3.69	3.56	3.54	3.62	3.68	3.62	3.62	3.74
14	3.67	3.57	3.55	3.63	3.69	3.63	3.63	3.76
15	3.66	3.58	3.57	3.64	3.70	3.64	3.64	3.77
16	3.67	3.60	3.58	3.66	3.71	3.66	3.66	3.79
17	3.67	3.62	3.61	3.68	3.73	3.68	3.68	3.81
18	3.67	3.64	3.63	3.71	3.75	3.70	3.70	3.83
19	3.68	3.67	3.65	3.73	3.77	3.72	3.71	3.86
20	3.68	3.69	3.68	3.75	3.79	3.74	3.73	3.88
21	3.69	3.71	3.70	3.77	3.80	3.76	3.76	3.90
22	3.70	3.73	3.72	3.78	3.81	3.78	3.77	3.92
23	3.72	3.74	3.73	3.78	3.82	3.79	3.79	3.94
24	3.73	3.76	3.75	3.79	3.83	3.80	3.80	3.95
25	3.74	3.76	3.77	3.80	3.84	3.81	3.81	3.96
26	3.75	3.78	3.79	3.81	3.85	3.83	3.82	3.98
27	3.76	3.80	3.81	3.83	3.86	3.84	3.84	4.00
28	3.78	3.82	3.83	3.85	3.88	3.86	3.86	4.03
29	3.79	3.84	3.86	3.88	3.90	3.88	3.89	4.06
30	3.81	3.86	3.89	3.90	3.93	3.91	3.91	4.09
31	3.83	3.87	3.90	3.92	3.94	3.93	3.93	4.12
32	3.85	3.89	3.93	3.94	3.97	3.96	3.96	4.14
33	3.87	3.91	3.95	3.97	3.99	3.99	3.99	4.18
34	3.90	3.94	3.98	3.99	4.02	4.02	4.02	4.21
35	3.93	3.97	4.00	4.01	4.03	4.04	4.04	4.24
36	3.94	3.98	4.02	4.03	4.05	4.06	4.06	4.27
37	3.96	3.99	4.05	4.05	4.08	4.08	4.09	4.29
38	3.97	4.00	4.07	4.08	4.11	4.12	4.13	4.33
39	3.99	4.02	4.10	4.10	4.13	4.15	4.16	4.36
40	4.01	4.03	4.12	4.12	4.16	4.17	4.18	4.39
41	4.04	4.05	4.14	4.14	4.18	4.19	4.20	4.42
42	4.07	4.07	4.16	4.15	4.20	4.22	4.23	4.45
43	4.09	4.09	4.17	4.17	4.23	4.24	4.26	4.48
44	4.11	4.11	4.18	4.18	4.25	4.26	4.28	4.51
45	4.13	4.12	4.19	4.20	4.26	4.27	4.30	4.53
46	4.14	4.13	4.21	4.21	4.28	4.29	4.33	4.55
47	4.15	4.14	4.22	4.22	4.30	4.31	4.35	4.58
48	4.16	4.16	4.24	4.24	4.33	4.33	4.38	4.60

Figure 3. Heat map of MAPE values (in %) for recency effect modeling based on validation data (year of 2006) of Z_{21}.

Using this model $(d = 2; h = 4)$ to perform one-year ahead ex post forecasting on the test data (2007) with the previous two years (2005-2006) for parameter estimation, we get the MAPE value of 4.27%. Comparing to the MAPE of Tao's Vanilla Benchmark model on the test data presented in (Hong, Wang, & White, 2015), which is 5.22%, the recency effect model reduces the MAPE value by 18%.

Figure 4 shows the comparison between the forecasts from benchmark model and recency effect model on the same weeks as shown in Figure 2. We can observe that 1) the recency effect model captures peaks and troughs much closer than the benchmark model; and 2) The leading/lagging issue of the benchmark model has been resolved by including the recency effect. Overall, the model with recency effect helps capture more salient features than Tao's Vanilla Benchmark model.
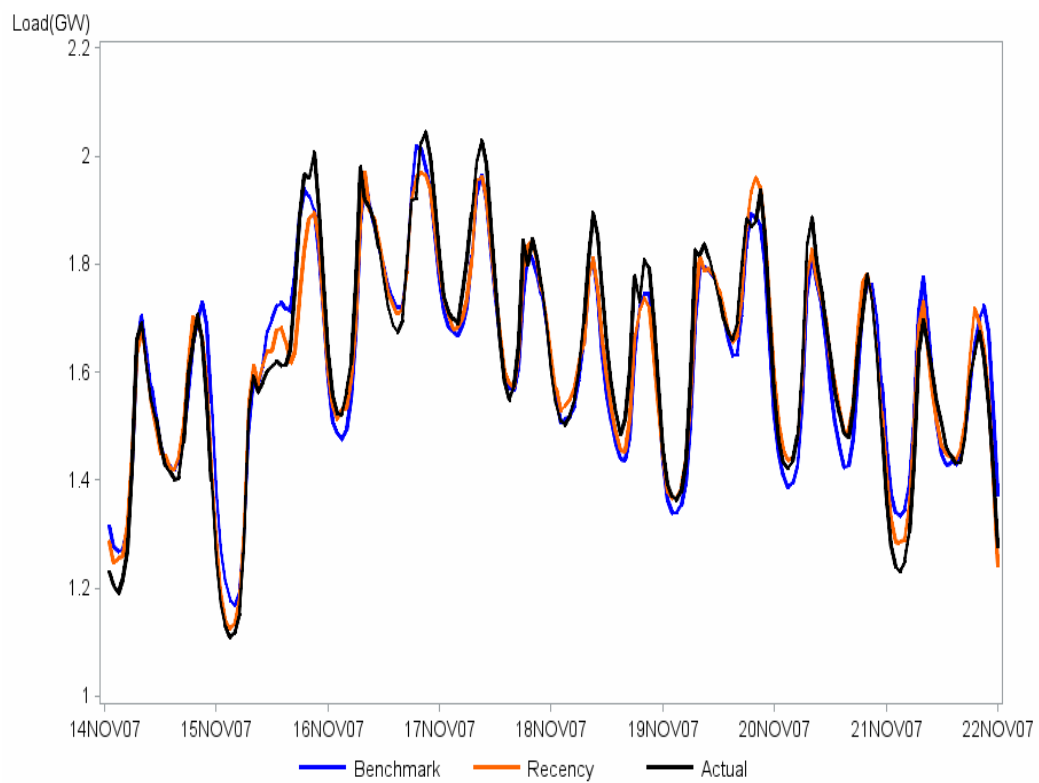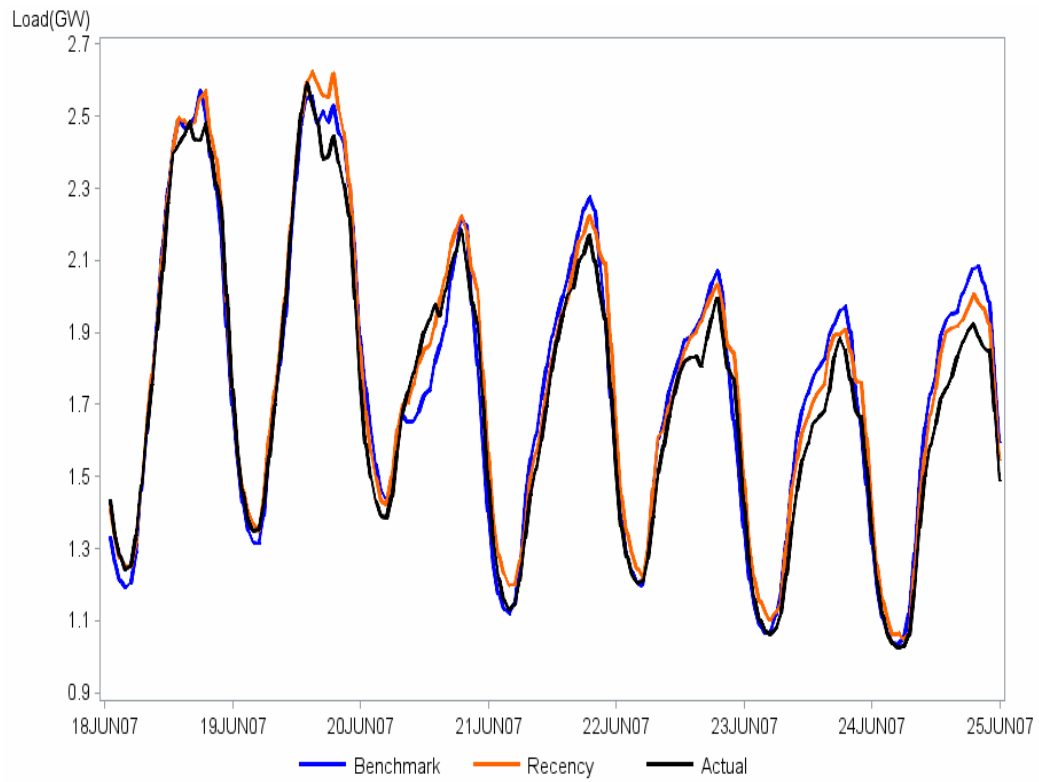
Figure 4. Actual load, forecasted load from benchmark model, and forecasted load from recency effect model for a summer week (upper) and a winter week (lower).

3.2 In a geographic hierarchy

The geographic hierarchy in hierarchical load forecasting refers to the households, small areas, sub-regions, regions, rolling up to the top aggregated level. The GEFCom2012 data itself has 20 zones at low level, with Z_{21} as the sum of those 20 zones. In this section, we apply the recency effect modeling to all 21 load zones of GEFCom2012 data.

GEFCom2012 data did not include the locational information. To have a rough estimate on the locations, we first take the annual average temperature of each virtual weather station. We then sort the average temperatures in ascending order to represent zones from north to south. Figure 5 shows the *d-h* pairs for the 19 zones (18 regular zones plus Z_{21}) sorted by average annual temperature. Overall, for most of the 19 zones, the recency effect appears to be different. This is due to the fact that different zones may have different compositions of residential, commercial and industrial customers, which together may have different responses to the weather patterns.
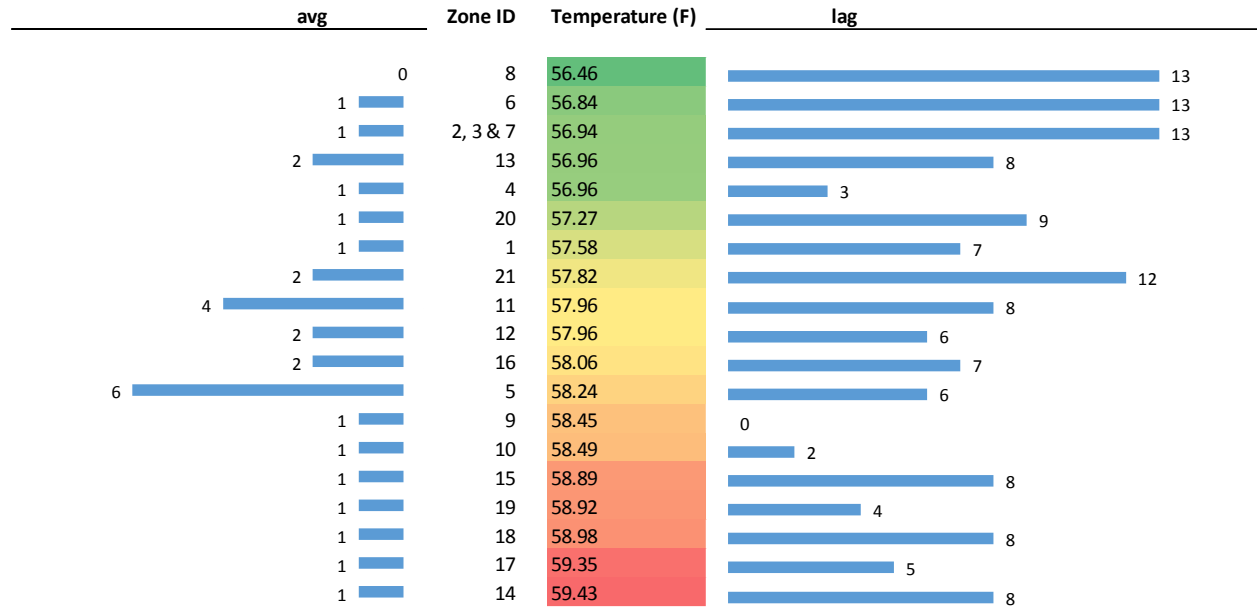
Figure 5. Recency effects (in *d-h* pairs) for the regular zones sorted by annual average temperature.

Table 1 lists the MAPE values for one-year ahead ex post forecast of year 2007 using the previous two years (2005 – 2006) for parameter estimation. As discussed in (Hong et al., 2015), Z_4 experienced a major outage; Z_9 is an industrial customer. Therefore, these two special zones are listed separately from the others. At the aggregated level, modeling recency effect helps reduce the MAPE value of benchmark

model by 18%. At the low level, when modeling the recency effect in a customized fashion, the forecasting error can be improved for the 17 of the 18 regular zones. On average, the recency effect models reduce the MAPE values of the benchmark model by 12%. The improvement of forecast accuracy through customization on each low-level zone is also consistent with the conclusion of (Hong, Wang, & White, 2015), which indicates that different zones may need different sets of weather stations to enhance the forecast accuracy.

Table 1 MAPE values (in %) of the Recency effect vs. benchmark models on 21 zones of the GEFCom2012 load forecasting data (one-year ahead ex post forecasting for 2007). Z_4 experienced a major outage; Z_9 is an industrial customer.

	Zone	Recency	Benchmark
Aggregated Zone	21	**4.27**	5.22
Regular Zones	1	**5.80**	7.01
	2	**4.90**	5.62
	3	**4.90**	5.62
	5	10.71	**9.88**
	6	**4.90**	5.55
	7	**4.90**	5.62
	8	**6.66**	7.50
	10	**5.62**	6.70
	11	**6.53**	7.70
	12	**5.54**	6.78
	13	**6.81**	7.39
	14	**8.13**	9.38
	15	**6.63**	7.44
	16	**6.80**	8.12
	17	**4.32**	5.26
	18	**5.50**	6.72
	19	**6.68**	7.90
	20	**5.02**	5.74
	Average (regular zones)	**6.13**	7.00
Special Zones	4	16.13	**16.08**
	9	140.9	**139.16**

3.3. By hour of a day

Hourly electricity demand series naturally has multiple seasonal patterns, such as hours of a day, days of a week, months of a year, and so forth. These calendar variables automatically form the temporal hierarchy. In the load forecasting literature, some researchers forecasted the load based on the methodology of using all the historical data to derive one model (Hong, 2010)(Hong, Wilson, & Xie, 2014)(Hong & Wang, 2014). Many other researchers sliced the data to 24 pieces, so that the parameter estimation of the load forecasting model is conducted individually for each of the 24 groups (Ramanathan, Engle, Granger,

Vahid-Araghi, & Brace, 1997)(Fan & Hyndman, 2012)(Black & Henson, 2014)(Ben Taieb & Hyndman, 2014). There has not been any formal comparison between the two. In addition, there has not been any formal study to explore the possibilities of modeling each hour with different variable combinations, though (Ramanathan et al., 1997) briefly mentioned that different variable combinations were used to model different hours. In this section, we are testing four options of cross validation and model settings as shown in Table 2:

Table 2 Different implementations of recency effect modeling

	Training	Validation	*d-h* Pair
A	1	1	1
B	24	24	1
C	1	24	24
D	24	24	24

A. This is our base case and has been conducted in Section 3, where we have identified the *d-h* pair through one training dataset and one validation dataset. This implementation results in the same variables (a.k.a. one *d-h* pair) and parameters for all 24 hours.

B. We first slice the data to 24 pieces by hour of the day. Each piece includes a training set and a validation set. We estimate a model for each hour of the day using the training set of the corresponding hour, and then calculate the MAPE on the validation set. We select the *d-h* pair that results in the lowest average MAPE on the 24 validation sets as the pair for the 24 models. This implementation also results in the same variables for all 24 hours, but the parameters are estimated separately for each hour of a day. In total, we have 24 sets of parameters, one for each hour.

C. We slice the validation data to 24 pieces by hour of the day. We then estimate a model using the full (i.e., all hours) training set. The *d-h* pair that results in the lowest MAPE on one validation set is selected as the pair for the corresponding hour of the day. In total, we select 24 *d-h* pairs. In other words, this implementation results in the 24 sets of variables and parameters, one for each hour of a day.

D. The data processing is similar to that of Option B. We first slice the data to 24 pieces by hour of the day. Each piece includes a training set and a validation set. Instead of selecting one common *d-h* pair for all 24 models, we are designating one *d-h* pair for each model. In total, we select 24 *d-h* pairs. This implementation results in the 24 sets of variables and parameters, one for each hour of a day.

Based on the above four implementation options, we can develop four models, of which the parameters (values of d and h) are shown as the radar charts in Figure 6. Since options A and B are to use one d-h pair across all 24 hours of a day, their territories are shown as circles in the radar charts. When releasing this constrain of having the same d-h pair for all hours, different hours of a day may present different recency effects. This observation is similar to an observation made in Table 1, where different zones may present different recency effects. It is again due to the fact that the relationship between electricity consumption and the temperatures varies throughout a day. Nevertheless, we can observe similar or same d-h pairs in some of the adjacent hours. Figure 6 also shows that the territories of options B and D are significantly less than those of options A and C. In other words, Options B and D have much less number of variables than options A and C. This is because the observations used to estimate each model in options B and D is only 1/24 of the observations for estimating a model from options A and C. The number of observations in the training data limits the amount of variables a model can hold.

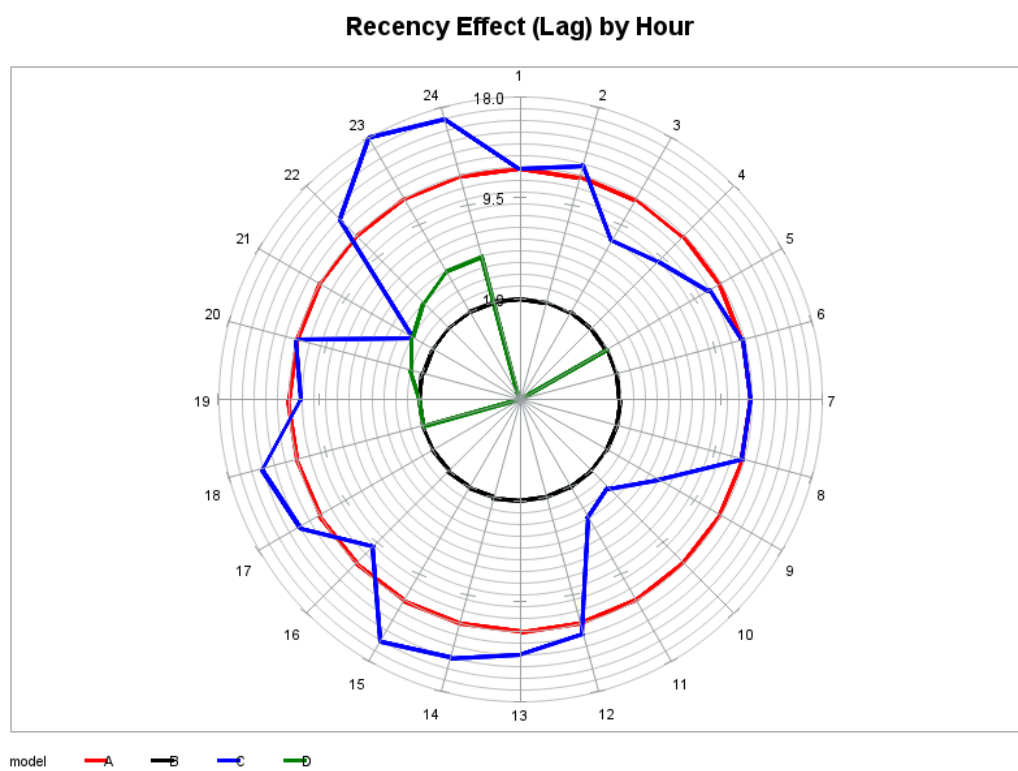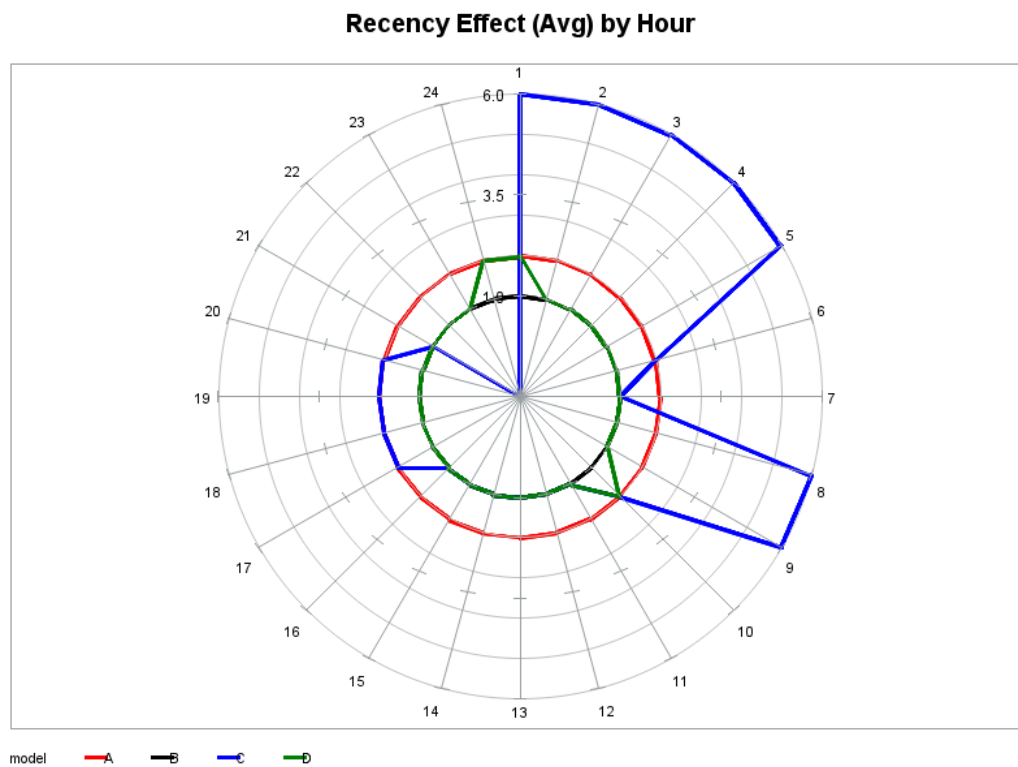# Recency Effect (Avg) by Hour

Recency Effect (Lag) by Hour

Figure 6. The average-lag parameters for the four options A through D.

We then apply these four models for one-year ahead ex post forecasting of the loads of 2007. The error statistics of these four options are listed in Table 3. On average, option A leads to the lowest MAPE values. In other words, identifying recency effect modeling using one training data and one validation data outperforms the other three alternative implementations in this case study.

Note that the common thoughts in 1990s and 2000s are to use one set of variables for all hours of a day, which makes the computation easy. (Ramanathan et al., 1997) is one of the earliest papers that reported customized models for different hours of a day. Here in this paper we further advance the state-of-the-art by leveraging the modern computing power to release the limit of the number of variables being included in a regression model and showing superior accuracy.

Furthermore, this part of the case study also offers an empirical evidence showing that slicing the data to 24 pieces to develop one model for each hour, as what many papers in the load forecasting in the literature have done, is not necessarily better than the other three alternatives as listed in Table 2. On the other hand, although option A has the lowest MAPE value in this case study, it may or may not be the best one in other case studies with different datasets. When slicing the data to multiple pieces or modeling each hour individually, we are reducing the noise for each group. Meanwhile, we are also reducing the signal that could have been leveraged from other groups. In other words, we have to find out the balance between signal and noise, which is unknown prior to trial-and-error. And the answer can be different on different data sets.

Table 3 MAPE values (in %) of recency effect models by hour of a day (one-year ahead ex post forecasting for 2007)

Hour	A	B	C	D	Benchmark
1	**4.37**	4.50	4.77	4.68	5.82
2	4.56	4.59	4.97	**4.32**	5.84
3	4.59	4.28	5.24	**4.27**	5.82
4	4.69	4.47	5.22	**4.23**	5.71
5	4.61	4.41	5.07	**4.41**	5.75
6	4.66	**4.50**	4.66	4.52	5.70
7	4.86	4.93	**4.78**	4.81	5.56
8	4.28	4.21	4.70	**4.19**	4.85
9	3.71	**3.64**	4.00	3.70	4.32
10	**3.39**	3.61	3.46	3.75	4.16
11	3.53	3.98	**3.48**	3.81	4.25
12	**3.75**	4.18	3.79	4.07	4.51
13	**4.00**	4.45	4.02	4.32	4.52
14	**4.20**	4.54	4.22	4.57	4.68
15	**4.50**	4.62	4.54	4.80	4.82
16	**4.61**	4.73	4.67	4.82	5.02
17	4.65	4.70	**4.64**	4.81	5.12
18	4.70	4.72	4.68	**4.72**	5.39
19	4.33	4.27	4.31	**4.27**	5.52
20	**4.22**	4.34	4.22	4.23	5.75
21	4.14	4.19	**4.07**	4.20	5.36
22	3.90	4.27	**3.81**	4.02	5.56
23	4.00	4.48	**3.82**	4.46	5.53
24	4.30	4.49	**4.06**	4.95	5.73
Average	**4.27**	4.38	4.38	4.37	5.22

4. Discussions

4.1 More comparisons

So far in this paper, all recency effect models have been compared with Tao's Vanilla Benchmark model using the weather stations selected in (Hong, Wang, & White, 2015). The forecast horizon has been set to be one year. In this section, we will expand the comparative study to other models in the setting of 24-hour ahead ex post forecasting for year 2007. The forecasts of hours 1 through 24 are issued on rolling basis at each midnight using a 730 days moving window for parameter estimation.

Here we introduce the four models for further comparisons:

1) An advanced model (Adv_1) modified from the recency effect models proposed in (Hong, 2010), where we let $d = 1$ and $h = 3$ for all hours. This is to see whether stretching the number of average temperatures and lagged temperatures can help enhance the accuracy comparing with the models developed with computational constrains.

2) An advanced model (Adv_2) simplified from the models in (Ben Taieb & Hyndman, 2014). Here we plug the same temperature variables identified from (Ben Taieb & Hyndman, 2014) to eq(2) with each hour modeled as a separate group. Note that this is not a direct reproduction of the models in (Ben Taieb & Hyndman, 2014), because we are not including the lagged loads in this paper as mentioned earlier in Section 2.2. Moreover, the model structure we are using in this paper is different from the models in (Ben Taieb & Hyndman, 2014). Following the analogy of integration discussed in (Hong & Fan, 2016), a subset of the variables selected by (Ben Taieb & Hyndman, 2014) for each hour may not be the optimal selection in the model structure here. The purpose of this comparison is to have a sanity check whether the performance of the recency effect models proposed in this paper is worse than a simplified version of the models from a top 5 team of GEFCom2012.

3) A daily seasonal naïve benchmark (SNM-D). It uses the load of hour t-24 as the forecast for load of hour t.

4) A weekly seasonal naïve benchmark (SNM-W). It uses the load of hour t-168 as the forecast for load of hour t.

Table 4 lists the 24-hour ahead ex post forecasting MAPE values for all 21 zones from the recency effect model (option A), two advanced models (Adv_1 and Adv_2), the benchmark model, and two seasonal naïve models. At the aggregated level (Z_{21}), modeling recency effect helps reduce the MAPE value of

benchmark model by 21% for 24-hour ahead rolling forecasting. At the low level, the recency effect models (option A) win 12 of the 18 regular zones, reducing the MAPE values of the benchmark model by 15% on average.

Both seasonal naïve models present very poor performance comparing with the other four models. Seasonal naïve models are commonly used for benchmarking purposes in other industries, such as retail and manufacturing industries. In load forecasting, the two applications where seasonal naïve models are most useful are: 1) benchmarking forecast accuracy on very unpredictable loads, such as household level loads; and 2) comparisons with univariate models. In most other applications, however, the seasonal naïve models and other similar naïve models are not quite meaningful due to the lack of accuracy.

Table 4 MAPE values (in %) of the Recency effect vs. benchmark models on 21 zones of the GEFCom2012 load forecasting data (24-hour ahead ex post forecasting for 2007). Z_4 experienced a major outage; Z_9 is an industrial customer.

	Zone	Recency (A)	Adv_1	Adv_2	Benchmark	SNM-D	SNM-W
Aggregated Zone	21	**3.86**	3.97	4.45	4.88	8.40	14.30
Regular Zones	1	**5.62**	5.82	6.83	7.00	11.28	19.01
	2	**4.59**	4.65	5.36	5.25	8.55	12.57
	3	**4.59**	4.65	5.36	5.25	8.55	12.57
	5	7.69	**7.28**	8.62	8.32	12.59	20.47
	6	**4.57**	4.65	5.29	5.27	8.55	12.79
	7	**4.59**	4.65	5.36	5.25	8.55	12.57
	8	**6.02**	6.26	7.24	6.85	9.40	15.96
	10	**4.97**	**4.97**	6.17	5.85	12.23	15.80
	11	**5.71**	5.81	6.38	6.88	10.77	18.13
	12	5.11	**5.09**	5.83	6.41	11.94	20.31
	13	**6.31**	6.36	7.60	7.13	9.01	15.23
	14	7.60	**7.56**	9.12	9.10	15.66	25.14
	15	6.29	**6.26**	7.34	7.19	10.56	17.69
	16	6.37	**6.31**	7.83	7.69	14.12	23.75
	17	**4.13**	4.14	4.94	5.09	10.58	15.65
	18	**5.25**	5.29	6.26	6.50	11.58	19.51
	19	**6.14**	6.16	7.50	7.60	13.12	21.77
	20	**4.54**	4.62	5.38	5.20	8.15	12.72
	Average	**5.56**	5.59	6.58	6.55	10.84	17.31
Special Zones	4	**14.51**	**14.51**	15.97	14.74	15.52	20.73
	9	109.84	107.21	115.18	113.79	47.53	120.06

Table 5 lists the MAPE values of the recency effect model (option A), two advanced models mentioned above, and the benchmark model. This time we zoom into the error statistics by hour. We only show the results from Z_{21} to avoid verbose presentation, because results from the other zones do not add additional

information. Overall, the recency effect model wins 18 of the 24 hours of a day, while the Adv_1 model wins the other 6. On average, the recency effect model is the most accurate one among the four, followed by the Adv_1 model, and then Adv_2 model.

Table 5 MAPE value (in %) comparison of recency model (option A) and two advanced models and the benchmark by hour of a day (24-hour ahead ex post forecasting for 2007)

Hour	Recency (A)	Adv-1	Adv-2	Benchmark
1	**3.73**	4.26	4.57	5.41
2	**3.83**	4.22	4.67	5.33
3	**3.88**	4.24	4.42	5.33
4	**3.98**	4.18	4.81	5.28
5	**3.95**	4.11	4.24	5.33
6	**4.13**	4.14	4.25	5.24
7	4.53	**4.33**	5.04	5.18
8	4.08	**3.86**	4.51	4.58
9	3.51	**3.36**	3.61	3.97
10	3.16	**3.11**	3.20	3.74
11	**3.30**	3.33	3.55	3.87
12	**3.51**	3.56	4.07	4.20
13	**3.59**	3.69	4.49	4.33
14	**3.76**	4.00	4.68	4.52
15	**4.00**	4.22	4.99	4.62
16	**4.12**	4.30	5.63	4.73
17	**4.23**	4.26	5.09	4.72
18	**4.39**	4.51	4.95	5.05
19	**4.06**	4.10	4.31	5.37
20	3.93	**3.87**	4.50	5.37
21	3.66	**3.63**	4.22	4.91
22	**3.57**	3.68	3.86	5.15
23	**3.73**	4.11	4.42	5.32
24	**3.97**	4.28	4.64	5.47
Average	**3.86**	3.97	4.45	4.88

4.2 Big data in electric load forecasting

The utilities have been working with big data since it was called just "data". We have witnessed the growth of data to big data in this smart grid era. Nowadays, the term "big data" is no longer new to the

utility industry. In fact, big data is gradually becoming a buzzword in the industry and academia. With the term "big data" in the title of this paper, we would like to expand the discussion a bit to the concept of big data in load forecasting, or energy data analysis in general.

Fundamentally, the "big" in big data should be relative to what happened in the past. Therefore, we define the big data criteria in relative terms:

1) The data size is larger than what typical data analysis tools (including the IT infrastructure around the tools) can handle. For instance, if MS Excel 2003 is the tool to conduct data analysis, then a data file with 5 million rows is big. If the data should come in at 1GB/sec, but the internet downloading speed is 10MB/sec, then the data stream is big.

2) The computing time is longer than the analysis time or lead time for the business applications. If it takes a few hours to think of the design of an algorithm, but testing the algorithm takes a few weeks, then it is big data problem. If it takes a few hours to calibrate an hour-ahead forecasting model, which should be re-calibrated every hour, then it is a big data problem.

3) The problem requires analysis at a higher level of granularity than usual. If the typical load forecasting process relies on monthly data, moving to daily or hourly data may bring you the big data challenge.

Although these three criteria do not have to be met at the same time to qualify a big data problem, they are indeed connected with each other. Analyzing high resolution data with the customized models pushed down to low levels often requires advanced data analysis tools and significant computing time.

This paper has big data in its title, because it covers the latter two criteria. The regression models we developed in this paper contain up to thousands of variables. This requires significant amount of time for parameter estimation, much longer than our thought process. Moreover, we have customized the models based on each zone of a geographic hierarchy and each node (hour) of the temporal hierarchy. The forecasting error reduction through our proposed approach also suggests that adopting high performance computing facilities in load forecasting may make sense in some circumstance.

The fact that the recency effect model beats Adv_1 model confirms the effectiveness of taking a big data approach to forecast accuracy improvement. Nevertheless, we can also observe the diminishing improvement as the customized modeling effort is being pushed down to the geographic hierarchy, which indicates that there may be a limit for such approach. In other words, we may not want to apply these large models down to the very low level of the power systems. This also makes sense, because at very low level, i.e., household level, the load is much more volatile and unpredictable than the load at

aggregated levels. Additional research is needed to identify the appropriate level to conduct customized modeling efforts.

In terms of forecasting techniques, we have applied multiple linear regression with ordinary least square estimator in this paper mainly for its simplicity and practicality. Nevertheless, other techniques, such as ridge regression and Bayesian regression or dynamic factor models can also be plugged into our solution framework. The key takeaway from this paper is how to model the recency effect in customized fashion. Regression analysis is just one of the many techniques that can be used to convey the same message.

4.3 Future research directions

This paper not only answers the fundamental question of how many average and lagged temperatures are needed in the model, but also opens a wide variety of research topics for future work:

1) Cross validation methods towards optimal recency effect. We enumerate the same 584 d-h pairs in Figure 2 to draw another heat map for the test data (the year of 2007) of Z_{21} using data of 2005 and 2006 to estimate the parameters. As shown in Figure 7, the lowest MAPE value in hindsight is 4.17%, which occurs with $d = 0$ and $h = 16$. Since the future is stochastic by nature, the d-h pair selected previously from the validation data (the year of 2006), in this case $d = 2$ and $h = 12$, may not be the optimal d-h pair in hindsight. The similar phenomenon was observed and discussed in (Hong et al., 2015) as optimal weather station selection. In this paper, we used a simple out-of-sample test to select models. A future research direction can be examining various cross validation methods such as the ones discussed in (Arlot & Celisse, 2010) for model selection.

2) Enhancement of computational efficiency. In this paper, we did not apply any advanced search algorithms to identify the d-lag pair that results in the lowest validation error. Based on Figure 3, it is clear that searching the whole space, i.e., 8 by 73, may not be necessary. We can try or design some heuristic search algorithms to reduce the search space without compromising the forecasting accuracy.

3) Dimension reduction. A model with thousands of parameters to estimate may be larger than necessary. We can investigate whether dimension reduction techniques can help make less verbose model while reducing the forecast error. For instance, we can try to combine some hours of a day to reduce the number of parameters to be estimated. A similar analogy was taken in (Hong, 2010) to combine some weekdays together. The similar ideas have also been applied to electricity price forecasting, as discussed in Weron (2014).

4) Forecast combination. We have generated many models with similar structure and accuracy in this paper. Rather than selecting one model via cross validation, we can also combine the various forecasts. In (Liu, Nowotarski, Hong, & Weron, 2015), these similar models are named as "sister models", with their forecasts named as "sister forecasts". The forecast combination output can be a probabilistic forecast as done in (Liu et al., 2015) or another point forecast.

Lag/d	0	1	2	3	4	5	6	7
0	5.22	4.51	4.57	4.60	4.64	4.80	4.88	4.98
1	4.94	4.40	4.48	4.52	4.57	4.75	4.86	4.93
2	4.75	4.33	4.41	4.45	4.50	4.68	4.79	4.86
3	4.60	4.29	4.37	4.40	4.46	4.64	4.75	4.82
4	4.49	4.25	4.34	4.36	4.43	4.60	4.71	4.78
5	4.40	4.24	4.32	4.34	4.40	4.57	4.68	4.74
6	4.34	4.23	4.31	4.32	4.39	4.55	4.66	4.72
7	4.28	4.24	4.30	4.32	4.39	4.54	4.65	4.71
8	4.25	4.25	4.30	4.32	4.39	4.54	4.64	4.71
9	4.23	4.25	4.30	4.31	4.38	4.53	4.63	4.69
10	4.21	4.25	4.29	4.31	4.38	4.51	4.61	4.68
11	4.20	4.26	4.28	4.31	4.38	4.51	4.59	4.67
12	4.20	4.26	4.27	4.31	4.38	4.50	4.58	4.66
13	4.19	4.27	4.27	4.31	4.38	4.50	4.57	4.66
14	4.18	4.26	4.27	4.31	4.38	4.49	4.57	4.66
15	4.18	4.26	4.26	4.31	4.38	4.50	4.57	4.66
16	**4.17**	4.25	4.26	4.31	4.39	4.50	4.58	4.67
17	4.18	4.26	4.26	4.32	4.40	4.51	4.59	4.67
18	4.18	4.26	4.27	4.33	4.40	4.51	4.60	4.68
19	4.19	4.28	4.29	4.34	4.41	4.53	4.61	4.69
20	4.20	4.28	4.30	4.35	4.42	4.54	4.63	4.70
21	4.21	4.29	4.31	4.36	4.42	4.55	4.64	4.71
22	4.22	4.29	4.32	4.37	4.43	4.56	4.65	4.72
23	4.23	4.30	4.34	4.38	4.43	4.57	4.66	4.73
24	4.24	4.31	4.35	4.39	4.44	4.57	4.67	4.73
25	4.25	4.31	4.36	4.41	4.46	4.59	4.68	4.74
26	4.26	4.32	4.38	4.43	4.48	4.60	4.69	4.75
27	4.28	4.34	4.41	4.45	4.50	4.62	4.71	4.78
28	4.29	4.36	4.43	4.47	4.52	4.65	4.74	4.81
29	4.30	4.37	4.45	4.48	4.54	4.67	4.75	4.83
30	4.31	4.39	4.46	4.50	4.57	4.69	4.77	4.86
31	4.33	4.41	4.48	4.52	4.59	4.71	4.80	4.89
32	4.35	4.43	4.50	4.53	4.60	4.73	4.82	4.92
33	4.36	4.44	4.50	4.54	4.60	4.73	4.82	4.93
34	4.37	4.46	4.50	4.54	4.61	4.74	4.84	4.96
35	4.39	4.47	4.51	4.56	4.62	4.77	4.86	4.99
36	4.40	4.48	4.52	4.57	4.63	4.78	4.88	5.02
37	4.41	4.49	4.55	4.59	4.66	4.81	4.92	5.06
38	4.43	4.49	4.56	4.61	4.68	4.83	4.94	5.09
39	4.44	4.50	4.56	4.62	4.69	4.85	4.96	5.11
40	4.45	4.51	4.57	4.64	4.70	4.86	4.97	5.13
41	4.47	4.51	4.57	4.65	4.71	4.87	4.99	5.15
42	4.49	4.53	4.58	4.67	4.73	4.88	5.00	5.18
43	4.51	4.54	4.59	4.68	4.74	4.89	5.02	5.20
44	4.53	4.55	4.61	4.69	4.76	4.91	5.03	5.21
45	4.54	4.56	4.62	4.71	4.78	4.93	5.06	5.24
46	4.55	4.58	4.63	4.72	4.79	4.95	5.07	5.26
47	4.57	4.59	4.65	4.74	4.82	4.97	5.09	5.28
48	4.58	4.60	4.66	4.74	4.82	4.98	5.10	5.28

Figure 7. Heat map of MAPE values (in %) for recency effect modeling based on test data (2007) of Z_{21}.

5. Conclusion

In conclusion, this paper investigates the recency effect for electric load forecasting, the fundamental relationship between load and temperatures of preceding hours. The idea is to explore how many lagged hourly temperatures and daily moving average temperatures are needed to enhance the load forecasting accuracy without the constraint of computing power. The case study is based on the data published in GEFCom2012. We demonstrate the recency effect from three aspects: 1) aggregated level (Z_{21}) of the geographic hierarchy; 2) the 20 zones at low level of the geographic hierarchy; 3) the 24 hours of a day. Overall, the recency effect model significantly outperforms the benchmark model based on MAPE of the holdout sample (test data, year of 2007 in this paper). For each zone of the geographic hierarchy and each hour of the day, the recency effect appears to be different. The case study has shown that customizing the model for each zone or hour of the day using the proposed approach can help enhance the accuracy of the hierarchical load forecasts.

References

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. doi:10.1214/09-SS054

Ben Taieb, S., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting*, *30*(2), 382–394. doi:10.1016/j.ijforecast.2013.07.005

Black, J. D., & Henson, W. L. W. (2014). Hierarchical load hindcasting using reanalysis weather. *IEEE Transactions on Smart Grid*, *5*(1), 447–455. doi:10.1109/TSG.2013.2278475

Dryar, H. A. (1944). The effect of weather on the system load. *Transactions of the American Institute of Electrical Engineers*, *63*(12), 1006–1013. doi:0.1109/T-AIEE.1944.5058843

Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*. doi:10.1109/TPWRS.2011.2162082

Gross, G., & Galiana, F. D. (1987). Short-term load forecasting. *Proceedings of the IEEE*, *75*, 1558–1573. doi:10.1109/PROC.1987.13927

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*. doi:10.1109/59.910780

Hong, T. (2010). *Short term electric load forecasting*. North Carolina State University. Retrieved from http://www.lib.ncsu.edu/resolver/1840.16/6457

Hong, T. (2014). Energy forecasting : past , present , and future. *Foresight: The International Journal of Forecasting*, (32), 43–49.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, *in press*.

Hong, T., Pinson, P., & Fan, S. (2014). Global Energy Forecasting Competition 2012. *International Journal of Forecasting*, *30*(2), 357–363. doi:10.1016/j.ijforecast.2013.07.001

Hong, T., & Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, *13*(1), 91–103. doi:10.1007/s10700-013-9166-9

Hong, T., Wang, P., & White, L. (2015). Weather station selection for electric load forecasting. *International Journal of Forecasting*.

Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, *5*(1), 456–462.

Liu, B., Nowotarski, J., Hong, T., & Weron, R. (2015). Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Transactions on Smart Grid*, *in press*. Retrieved from http://dx.doi.org/10.1109/TSG.2015.2437877

Nedellec, R., Cugliari, J., & Goude, Y. (2014). GEFCom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, *30*(2), 440–446. doi:10.1016/j.ijforecast.2013.07.004

Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, *5*, 1535–1547. doi:10.1109/59.99410

Ramanathan, R., Engle, R., Granger, C. W. J., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, *13*(2), 161–174. doi:10.1016/S0169-2070(97)00015-0

Taylor, J.W., & McSharry, P.E. (2007). Short-term load forecasting methods: an evaluation based on European data. *IEEE Transactions on Power Systems*, 22(4), 2213–2219.

Weron, R. (2006). Modeling and forecasting electricity loads and prices: a statistical approach. Wiley, Chichester.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future, International Journal of Forecasting. 30(4), 1030-1081.

Pu Wang is a Senior Research Statistician Developer at SAS. She is specialized in operations research and statistical analysis with applications in retail and energy industries. She has been conducting research and development of emerging techniques in market response modeling and large scale demand forecasting. She received her B.Eng. in Industrial Engineering from Tsinghua University, Beijing, and the Master and PhD degrees in Industrial Engineering from North Carolina State University.

Bidong Liu is a PhD student at the Infrastructure and Environmental Systems program of University of North Carolina at Charlotte. He received his B.S. in Resource & Environmental Science from Zhejiang University and his M.S. in Economics from North Carolina State University.

Tao Hong is the Director of BigDEAL (Big Data Energy Analytics Laboratory), NCEMC Faculty Fellow of Energy Analytics, Graduate Program Director, Assistant Professor of Systems Engineering and Engineering Management, and Associate of Energy Production and Infrastructure Center at University of North Carolina at Charlotte. He is the Founding Chair of IEEE Working Group on Energy Forecasting, General Chair of Global Energy Forecasting Competition (gefcom.org), lead author of the online book Electric Load Forecasting: Fundamentals and Best Practices (otexts.org/elf), and author of the blog Energy Forecasting (blog.drhongtao.com). Dr. Hong received his B.Eng. in Automation from Tsinghua University in Beijing and his PhD with co-majors in Operations Research and Electrical Engineering from North Carolina State University.

HSC Research Report Series 2015

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html