A Weather-Based Forecasting Method for Short-Term Aggregate Power Loads

John D. Hobby, Gabriel H. Tucci and Mustafa K. Doğru

Abstract—This paper introduces a weather-based method for short-term forecasting of aggregate electricity load. We extract the weather- and illumination-dependent load via least-squares fitting for load versus Steadman apparent temperature and full-scale natural illumination. A separate fit is done for each hour of the day, and then Fourier-transform-based spectral analysis handles the resulting residual. The input consists of past load and weather data (temperature, humidity and cloud cover), weather forecasts, and location information. We do extensive performance tests using load data from a mid-size U.S. city as well as for 8 climate zones in the state of Texas. A well-tested exponential smoothing method due to J.W. Taylor serves as a benchmark. The results show that the benchmark outperforms our method up to about four hours ahead, but the new technique is considerably better for longer lead times. The method is highly robust and yields accurate forecasts for the next day or as far ahead as the weather can be forecasted. These characteristics make it a good forecasting tool for supporting decisions in unit commitment, economic dispatching, and electricity purchase in day-ahead markets.

I. INTRODUCTION

Electricity is a perishable commodity since there is no cheap and convenient way of storing overgeneration. Thus, the power industry—regional transmission organizations (RTOs) independent service operators (ISOs), utilities, generators, brokers—is in a constant battle to meet the load in a reliable and cost efficient manner, which makes load forecasting an important input to various routine decisions. In this paper, we focus on short-term load forecasting where the prediction horizon tends to vary from a few minutes to a week.

Short-term load forecasts are vital inputs to many planning and operational decisions that must be made frequently [1], [6], [11]. RTOs and ISOs (or vertically integrated utilities in some countries) solve unit commitment and economic dispatch problems that use fine grain load forecasts from utilities as inputs. Deregulation of electricity markets in the U.S., Europe, and elsewhere has enhanced the importance of short-term load forecasting, and created new decision processes that require load forecasts. Electricity purchases from day-ahead and spot markets by utilities are all based on hourly (or more frequent) load forecasts. Accurate day-ahead load forecasts help utilities to develop and adjust their bidding strategies in the electricity markets. Any demand response program relies heavily on load forecasts. Finally, short-term load forecasting is important for the reliability of the grid since accurate forecasts help prevent overloading and reduce the occurrence of blackouts [11].

In this paper, we develop and test a complete method for short-term load forecasting. Our approach decomposes the load into two components: weather and illumination dependent load, and residual load. For every day type (e.g., working and non-working days) and hour of the day, we analyze the past load and fit a surface as a function of apparent temperature and natural illumination. The fitted surfaces are used to extract the weather- and illumination-dependent load by an approach introduced in our earlier work in [7] and [8]. Subtracting the weather- and illumination-dependent load yields the residual load, which we study using fast Fourier transform based spectral analysis. Our forecasting method uses the following inputs: (i) past load, weather (temperature, humidity and cloud cover) and day-type data, (ii) weather forecast and future day-type data, (iii) latitude and longitude, and the time zone and daylight savings/summer time policy for the region of interest.

We conduct a thorough empirical study to test our forecasting method by using two load data sets from the U.S. The first data set, which is also used in [7] and [8], consists of hourly load observations in 2007-2009 from a mid-size U.S. city. The second data set contains 8 hourly load series for 2009-2011 from The Electric Reliability Council of Texas (ERCOT) [4]. We used the technique, which is based on Holt-Winters exponential smoothing and referred to as HWT, developed in [12] and [13] as the benchmark. HWT is a univariate forecasting method that only uses the past load series, and shown to outperform other univariate methods such as intraday cycle exponential smoothing, autoregressive integrated moving average and neural networks by an empirical study of load series from 10 European countries [15]. In a recent study, J.W. Taylor demonstrates the good performance of HWT using load data from Britain and France [14].

Considering static and rolling horizon versions of our method, we analyze the accuracy of our algorithms and HWT as a function of the forecast lead time. Then motivated by day-ahead markets in the U.S., we study the forecast errors for the next day as a function of the hour of the present day over a whole year. Next, we look at the forecast accuracy for the peak and valley hour of the next day as a function of the hour of the present day. These analysis assume perfect weather forecasts, so we also report the performance of our algorithm in the presence of forecast errors, which we obtained from National Oceanic and Atmospheric Administration (NOAA). Since the
NOAA data gives monthly errors, we develop a methodology for estimating the hourly weather forecast errors to use in the experiments.

The rest of the paper is organized as follows: Section II introduces the data sets used in the empirical study. Section III explains the weather- and illumination-dependent load is calculated. Sections IV and V discuss the residual load analysis and the spectral analysis, and Section VI presents the empirical results together with a discussion on how to tune the parameters of our algorithms. Finally, Section VII provides a discussion and conclusion.

II. DATA SETS

The empirical study in Section VI uses two sets of electricity load data from the U.S. to assess the forecast accuracy of our approach. Data Set 1 consists of hourly load observations in 2007, 2008 and 2009 from a mid-size U.S. city with a population over 150,000. This data set was also used in [7]. After some data cleaning, the data is summarized in Figure 1 in monthly buckets. The linear trend lines fitted for minimum, maximum and mean load show a declining electricity consumption. Moreover, this city uses electricity for heating, which explains the high electricity consumption in January. Figure 2 shows how average load changes for each day of the week over the course of a day in 2008; we observe a similar behavior in the other years. Note that major holidays are treated separately and represented by series “Hol”. Observe that weekdays, holidays and weekends show different daily load profiles.

![Fig. 1. Monthly range, and median and mean load for a mid-size U.S. city in 2007-2009. (Data Set 1)](image1)

The source of Data Set 2 is ERCOT, which supplies power to approximately 23 million consumers – equivalent to 85% of the Texas load. ERCOT divides its coverage area into 8 weather zones for planning purposes: (1) Coast (Houston, Victoria, Galveston), (2) East (College Station, Tyler, Lufkin), (3) Far West (Midland & Odessa, Fort Stockton), (4) North (Wichita Falls, Paris), (5) North Central (Dallas, Forth Worth, Waco), (6) South Central (Austin, San Antonio), (7) Southern (Corpus Christi, Laredo, Brownsville), (8) West (Abilene, San Angelo). Figure 3 shows these zones on a Texas map. ERCOT data set consists of 8 time series for hourly electric load of each weather zone for 2009, 2010 and 2011.

![Fig. 3. ERCOT weather zones. (Source: [5])](image2)

Figures 4 and 5 show the monthly range, median and mean for the total ERCOT load in 2009-2011, and average load profile for each day of the week in 2009, respectively. Compare Figure 1 and 4, and Figure 2 and 5. Both data sets possess seasonality where consumption increases in summer and winter months. However, the deviation between the summer and winter consumption levels are quite distinct. Unlike in Figure 1, the minimum, maximum and mean load shows an increasing trend in Figure 4. This may be explained by the global recession in 2008-2009 [1]; Figure 1 spans the time period when the U.S. started falling into recession, and Figure 4 covers the rebound. Also, the average load for weekends and holidays is lower than the weekdays, but the holidays are more similar to the weekend pattern in the ERCOT data set. As expected,
both data sets have intra-day variation, which is displayed in Figures 2 and 5.

![Graph showing load [MW] for each hour of the day from 01/09 to 11/11](image)

Fig. 4. Monthly range, and median and mean load for the entire ERCOT area in 2009-2011. (Data Set 2)

![Graph showing average load for each day of the week and the holidays in 2009. (ERCOT total load)](image)

Fig. 5. Average load for each day of the week and the holidays in 2009. (ERCOT total load)

Data Sets 1 and 2 exhibit unique load characteristics that are different from some of the load series reported in the literature. The daily load profiles in Figures 2 and 5 do not show a clear decrease in the afternoon hours as in [10] and [14] for Britain, and in [2] for Iran. Moreover, the electricity consumption is much higher during the summer than in the winter in our data sets unlike in Britain [14] and France [3] where the opposite holds.

III. WEATHER AND ILLUMINATION RELATED LOAD

The weather- and illumination-dependent load component can be analyzed separately for each hour of the day by fitting a surface as a function of apparent temperature and natural illumination (log Lux) as explained in [7]. Moreover, it would be possible to have separate surface–fitting problems for non-working days and for various seasons. We should also mention that the main reason we use hours is because our data is on an hourly basis, but other time granularities can be used.

Since both the weather and illumination components are very dependent on the hour of the day, it is better to fit each hour separately as was explained in [7]. Thus the dependence of electricity load on apparent temperature and illumination can be modeled by fitting a surface composed of B-spline-based bi-cubic patches. Depending on the data’s time granularity, it may be best to combine a few time steps worth of data and let the surface control points be linear functions of time.

In any case, cubic B-splines give $C^2$ continuity, and the control points that define the surface come from a linear least squares problem with the following characteristics:

- A $6 \times 3$ array of surface patches since temperature is more important than illumination,
- Regularization terms to encourage near zero second derivatives at the boundaries of the $6 \times 3$ array,
- When finding the minimum of the resulting surface, the domain is restricted to the convex hull of the data points in (temperature, illumination) space.

The latitude and longitude are needed because natural illumination requires sun and moon positions, and the time zone is needed in order to correlate the result with the load and weather data.

We observe that the fitted surface behaves convexly in the temperature for a fixed illumination, and it is increasing in the illumination component (as it gets darker) for a fixed temperature. Except for possible greenhouse effect at high illumination levels, this is the expected behavior of the load as a function of these two variables. We then shift this surface down so that its minimum is zero because there should be a temperature and illumination value where no load is necessary for either purpose. This is supported by the literature, e.g., [9]. The shifted surface will be our function model for the weather-related and illumination-related electricity load as a function of the apparent temperature and illumination. Subtracting this function from the total load gives us the non-weather related and non-illumination related electricity load (residual load) for that particular hour. We repeat this procedure for each hour of the day to get a complete model. Thus for each hour of the day we have a slightly different surface.

IV. ANALYZING THE RESIDUAL LOAD

After finding the weather- and illumination-related load at each hour in the historical load data and subtracting this from the reported loads, we have a sequence of residual load to analyze. The forecasted future values for temperature, humidity, and log scale atmospheric opacity can be linearly interpolated if necessary in order to get weather-related load at each hour in the requested future time period. Hence, the remaining task is to model residual load as a function of time so that it can be extrapolated into the future.

Figure 6 shows the residual load for Data Set 1 for a one-month period. Not surprisingly, it exhibits a strong daily
variation, but the daily average (blue line) also has weekly variations and other details. Zooming out to show a whole year (Figure 7) reveals a seasonal variation.

Fig. 6. Hourly residual load (gray), with daily average (blue). This covers about 1 month of 2007 including Memorial Day (magenta).

Fig. 7. 2007 hourly residual load (gray), with daily average (blue) and vertical lines that delimit holidays (magenta).

The complexity of Figures 6 and 7 suggests a general technique such as FFT-based spectral analysis, but it seems questionable to handle the daily average (blue) the same way. Averaging over whole weeks and applying a smoothing filter isolates the seasonal variation as shown in Figure 8. The smoothing filter is based on simple averaging functions $S_k$ that make $w_0, w_1, \ldots, w_51$ into $\bar{w}_0, \bar{w}_1, \ldots, \bar{w}_51$, where each $\bar{w}_i$ is the average of all $w_j$ where $|i - j| \leq k$ and $0 \leq j < 52$. Since composing such a function with itself 3 times gives approximately Gaussian behavior, we take

$$S_{k_3} (S_{k_2} (S_{k_1} (w_0, w_1, \ldots, w_51))))$$

where $k_1 \approx k_2 \approx k_3$.

We shall refer to this as $S_{k_3}^{seas}$, where $k_3^{seas} = k_1 + k_2 + k_3$ and $[k_3^{seas} / 3] \leq k_1, k_2, k_3 \leq [k_3^{seas} / 3]$.

Is it a good idea to handle dependence on the day of the week via smoothing functions $S_{k_3}^{dow}$ for some $k$? As shown in Figure 9, the raw, unsmoothed data for this variation is quite noisy. A smoothing function $S_{k_3}^{dow}$ can be used if we separate the data for each particular day of the week and do a separate smoothing operation for each as suggested by Figure 10.

Thus we have taken the noisy data shown in Figure 9, created a separate function of time for each day of the week, and smoothed each such function by applying $S_{k_3}^{dow}$. Therefore, we have seven smoothed functions one for each day of the week. Combining these smoothed functions produces a well-behaved day-of-the-week variation that changes gradually over the course of 6 months as shown in Figure 11.
Fig. 11. Daily average residual load scaled by $y/w_i$ for about 6 months of 2007 (pink) and a version smoothed via 7 applications of $S_{k_{low}}^{3k_{low}}$ with $k_{low} = 44$ (red). Vertical lines denote week boundaries and holidays (magenta).

V. SPECTRAL ANALYSIS

Let us denote by $\hat{w}_i$ the smoothed weekly average (i.e. after applying $S_{k_{low}}^{3k_{low}}$ to the weekly averages $w_i$) as in Figure 8. These smoothed weekly averages $\hat{w}_i$ can be combined with the function $d_j$ from Figure 11’s smoothed day-of-week and normalized by the annual average $y$ to give an estimator $\hat{w}_i d_j/y$ for future values of the daily average residual load. For holidays, the most reasonable predictor is just the average residual load for the previous year’s instance of that holiday. The remaining task is to analyze deviations from these daily averages; e.g., the deviations shown in Figure 12. The deviations are constructed by subtracting from the residual the daily averages $w_i$. Therefore, we have a zero mean function which has also zero mean for each week of the year.

1) Generate replacement data for hours where lack of load data or lack of weather data prevents computing how residual load deviates from its daily average. One could try to do this so as to optimize the effect on the frequency spectrum, but we found that it suffices to linearly interpolate the deviation values.

2) Compute the FFT of the $n = 8736$ hourly deviation values $x_0, x_1, \ldots, x_{8735}$. This gives $q_0, q_1, \ldots, q_{8735}$, where $q_j = \sum_k \omega^k_{8736} x_j$ and $\omega_{8736} = \exp(\frac{2\pi i}{8736})$.

3) Among frequency coefficients $q_0, q_1, \ldots, q_{4368}$, find the ones of largest magnitudes and let $|q|$ be the $n_{peak}$-th largest for some $n_{peak}$. Then let $\bar{q}_0, \bar{q}_1, \ldots, \bar{q}_{8735}$, be the result of zeroing all $q_j$ for which $|q_j| < (1 + \epsilon_{peak}) |q|$ for $\epsilon_{peak}$. (Note that each peak occurs twice because $q_j$ and $q_{8736-j}$ are complex conjugates.)

4) Obtain the smoothed hourly deviations $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{8735}$ by computing the inverse FFT of $\bar{q}_0, \bar{q}_1, \ldots, \bar{q}_{8735}$.

Figure 13 shows part of the function from Figure 12 in more detail together with a sample result from spectral analysis. Before beginning spectral analysis, the data are cleaned by dropping hourly load values that differ from the median by more than a factor of 10. Then we shorten the year to 364 days to obtain a whole number of weeks, and perform the following steps.

1) Generate replacement data for hours where lack of load data or lack of weather data prevents computing how residual load deviates from its daily average. One could try to do this so as to optimize the effect on the frequency spectrum, but we found that it suffices to linearly interpolate the deviation values.

2) Compute the FFT of the $n = 8736$ hourly deviation values $x_0, x_1, \ldots, x_{8735}$. This gives $q_0, q_1, \ldots, q_{8735}$, where $q_j = \sum_k \omega^k_{8736} x_j$ and $\omega_{8736} = \exp(\frac{2\pi i}{8736})$.

3) Among frequency coefficients $q_0, q_1, \ldots, q_{4368}$, find the ones of largest magnitudes and let $|q|$ be the $n_{peak}$-th largest for some $n_{peak}$. Then let $\bar{q}_0, \bar{q}_1, \ldots, \bar{q}_{8735}$, be the result of zeroing all $q_j$ for which $|q_j| < (1 + \epsilon_{peak}) |q|$ for $\epsilon_{peak}$. (Note that each peak occurs twice because $q_j$ and $q_{8736-j}$ are complex conjugates.)

4) Obtain the smoothed hourly deviations $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{8735}$ by computing the inverse FFT of $\bar{q}_0, \bar{q}_1, \ldots, \bar{q}_{8735}$.

Figure 14 shows the low frequency part of the spectrum computed by the FFT step. The main benefit of spectral analysis comes from processing it to eliminate noise as explained in Step 3. Quite a few frequency components are above the noise. Table I gives them in order of decreasing strength with only a few of the last ones being clearly dubious.

A possible change to the above spectral analysis procedure is to take logarithms of the hourly and daily average residual load, apply spectral analysis to the deviation between these
Table I

THE STRONGEST FREQUENCY COMPONENTS FROM FIGURE 14

<table>
<thead>
<tr>
<th>strength</th>
<th>cyc/day</th>
<th>hrs/cyc</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>165987951</td>
<td>1.000</td>
<td>24</td>
<td>1/day</td>
</tr>
<tr>
<td>117027347</td>
<td>2.000</td>
<td>12</td>
<td>2/day</td>
</tr>
<tr>
<td>32738662</td>
<td>0.857</td>
<td>28</td>
<td>6/week</td>
</tr>
<tr>
<td>26277272</td>
<td>2.143</td>
<td>11.2</td>
<td>15/week</td>
</tr>
<tr>
<td>253804743</td>
<td>1.857</td>
<td>12.9231</td>
<td>13/week</td>
</tr>
<tr>
<td>22567128</td>
<td>1.143</td>
<td>21</td>
<td>8/week</td>
</tr>
<tr>
<td>20867798</td>
<td>3.000</td>
<td>8</td>
<td>3/day</td>
</tr>
<tr>
<td>19423234</td>
<td>0.714</td>
<td>33.6</td>
<td>5/week</td>
</tr>
<tr>
<td>19033865</td>
<td>4.000</td>
<td>6</td>
<td>4/day</td>
</tr>
<tr>
<td>18494296</td>
<td>2.857</td>
<td>8.4</td>
<td>20/week</td>
</tr>
<tr>
<td>18333122</td>
<td>1.286</td>
<td>18.6667</td>
<td>9/week</td>
</tr>
<tr>
<td>18238723</td>
<td>5.000</td>
<td>4.8</td>
<td>5/week</td>
</tr>
<tr>
<td>15492288</td>
<td>2.286</td>
<td>10.5</td>
<td>16/week</td>
</tr>
<tr>
<td>14915227</td>
<td>1.714</td>
<td>14</td>
<td>12/week</td>
</tr>
<tr>
<td>13916645</td>
<td>2.714</td>
<td>8.84211</td>
<td>19/week</td>
</tr>
<tr>
<td>13358443</td>
<td>4.143</td>
<td>5.7931</td>
<td>29/week</td>
</tr>
</tbody>
</table>

logarithms, and invert the process when using the smoothed deviation; i.e., exponentiate and multiply the smoothed daily average by the exponentiated smoothed deviation.

VI. EMPIRICAL RESULTS

This section presents the results of various empirical tests that evaluate the accuracy of our method using the data sets introduced in Section II. Our main benchmark is the HWT method of J.W. Taylor [14],[13], and we consider not only the algorithm of Sections III–V (referred to as “Static”), but also a rolling horizon version of it (“Rolling”). Rolling reruns the entire algorithm each time there is another day of past history, while Static extrapolates the old history further into the future using the new weather forecast.

We use three measures of prediction accuracy: mean absolute percentage error (MAPE), root mean squared percentage error (RMSE), and coefficient of variation of RMSE ($c v_{R M S E}$). The MAPE results are of primary interest since this measure is widely used by utilities. Given time series for load $X_1, X_2, \ldots, X_N$, and the corresponding forecasts $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N$,

$$M A P E = \frac{1}{N} \sum_{i=1}^{N} \frac{|X_i - \hat{X}_i|}{X_i}$$

where $\hat{X} = \frac{1}{N} \sum_{i=1}^{N} X_i / N$ and RMSE=$\sqrt{\frac{\sum_{i=1}^{N} (X_i - \hat{X}_i)^2}{N}}$.

The numerical tests were conducted as follows. For Data Set 1, HWT takes the hourly load of the entire year 2007 (2008) as an input to optimize the smoothing and residual autocorrelation adjustment parameters. This one year worth of data is also used to initialize the necessary variables. Then, the HWT algorithm is applied on an hourly basis for 2008 (2009) load data, and the forecast accuracy is evaluated. Our approach not only uses 2007 (2008) load data, but also the weather data for 2007 (2008). For predicting 2008 (2009) load, the weather data for 2008 (2009) is used as “perfect” weather forecast.

Recall from Section II that there are 8 ERCOT weather zones and Data Set 2 contains the hourly loads for each one. As in Data Set 1, HWT uses 2009 (2010) load data for parameter optimization and initialization for to forecast 2010 (2011) load for each zone and the total ERCOT load. In a similar fashion, our approach uses 2009 (2010) weather and load series, and 2010 (2011) weather data as ‘perfect’ weather forecast to predict 2010 (2011) load for each zone. Summing up the load forecasts gives the forecast for ERCOT total load. The weather data used for each zone (e.g., Southern) was created calculating the weighted average of the weather observations obtained from NOAA’s National Weather Services for the major 2–3 cities mentioned in Section II for that particular zone (e.g., Corpus Christi, Laredo, Brownsville for the Southern zone). The weights were determined based on the relative populations of the cities. This tends to yield better results than taking the weather data for the most populated city in the zone. For example, in the Southern zone for 2010, MAPE of the Static algorithm for hourly load, peak demand and valley demand drop from 8, 8.7 and 8.4 down to 7.2, 8.1 and 7.5 respectively when weighted average weather data is used instead of the weather data for Corpus Christi. Nevertheless, the extent of the improvement depends on the region and may be negligible.

A. Tuning the Parameters for Best Prediction Accuracy

Section V used various parameters whose values need to be fixed based on the data, but not in a manner that uses this same training data for evaluating the algorithm. Hence the parameter tuning is based only on Data Set 1, using 2007 data to predict 2008 load as accurately as possible. Results for Data Set 1 will be quoted for 2008 data predicting 2009 load, and there will be no separate tuning for Data Set 2. Although forecast errors are considered below, the training procedure uses actual 2008 weather data as if it were a perfect weather forecast.

The following parameters need to be tuned:

- $k_{\text{seas}}$ the number of weeks for the width of the $S_k^{\text{seas}}$ filter that smooths the weekly average residual consumption values that describe the seasonal variation.
- $k_{\text{dow}}$ the number of weeks for the width of the $S_k^{\text{dow}}$ filter that smooths the annual variation in average consumption for each day of the week.
- $n_{\text{peak}}$ the maximum number of frequency spectrum peaks to consider in Step 3 when looking for nonnoise.
- $\epsilon_{\text{peak}}$ a fudge factor by which to multiply the smoothed deviation function from spectral analysis before applying it to the smoothed daily average residual consumption. Ideally, no such factor should be needed so we expect $\epsilon$ to be near 1.0.
- $Q^{\log}$ a boolean flag for whether or not to take logarithms before spectral analysis as suggested near the end of the previous section.
\(Q^{unkft}\) is a boolean flag for whether or not to use an FFT-based procedure for adjusting previously unknown residual deviation. The previous section did not describe this in detail since \(Q^{unkft} = 0\) turns out to be best.

Each of these parameters can be given to the prediction engine as a command line option, so the parameter values can be optimized by manually trying values and observing the resulting RMSE for 2008 load predictions. Testing the resulting parameters by using 2008 data (and actual 2009 weather) to predict 2009 load gives the top line of Table II.

It is instructive to see how far from optimal the chosen parameters are when predicting 2009 load. The rest of Table II summarizes this reoptimization process. The main problems are that the optimal \(k^{lead}\) value is quite different and 2009 benefits from nontrivial values of the dubious rescaling factor \(r\). Horizontal lines make it easier to see what happens when particular parameters are varied. Although the bottom line of the table gives the best result for 2009 predictions, we shall use the top-line parameters that were tuned based on 2008 predictions.

### TABLE II
**PARAMETER OPTIMIZATION, COMPARING HOW WELL 2009 PREDICTIONS MATCH ACTUAL VALUES**

<table>
<thead>
<tr>
<th>(k^{lead})</th>
<th>(k_{low})</th>
<th>(w_{peak})</th>
<th>(w_{low})</th>
<th>(Q^{unkft})</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>44</td>
<td>24</td>
<td>.08</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>44</td>
<td>24</td>
<td>.04</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.9</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.85</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.75</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.85</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.8</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.75</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.7</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>44</td>
<td>27</td>
<td>.04</td>
<td>.7</td>
<td>1</td>
</tr>
</tbody>
</table>

### B. Forecast Lead Time Analysis

We calculate MAPE up to 48 hours in increments of one hour for Data Set 1 for 2008 and 2009, and Data Set 2 for 2010 and 2011. Figure 15 plots the results for the three methods under consideration for Data Set 1 in 2009, and the entire ERCOT in 2010. HWT outperforms our Static (Rolling) method up to 5 (4) hours for Data Set 1, but its performance deteriorates considerably as the forecast lead time extends. Similarly, HWT performs better than both versions of our algorithm up to 3 hours for the ERCOT total load. Beyond 7 hours, MAPE figures are above 10% for HWT in Figure 15. Moreover, HWT shows a clear cyclical behavior where the accuracy improves in the last quarter of one-day (hours 18-24) and two-day (hours 42-48) lead times. Static has a level MAPE while the error for Rolling increases slightly as the lead time lengthens. (From 4.85 to 5.03 for Data Set 1 and from 3.64 to 3.78 for ERCOT total as the lead time increases from 1 hour to 48 hours.)

Recall that we run our method for all 8 weather zones and sum the forecasts for to generate the forecast for the ERCOT total load. The behavior observed in Figure 15 is valid for all weather zones except two: North (2.3% of Texas population) and East (4.9% of Texas population). The results for these weather zones are given in Figure 16. First, note that the MAPE curve for HWT still shows a daily seasonality, but unlike the behavior in the other zones and in Figure 15, there is a saddle decrease in North and no decrease in East. (This behavior of HWT is similar to the one in Figure 9 of [14].) Second, the performance of our approach is not as good as in the other zones that Rolling catches HWT at hour 7 and 8 for East and North, respectively.

Note that the MAPE values in Figure 15 are much higher than Taylor’s [14]. It appears that both Data Sets 1 and 2 are less predictable than Taylor’s U.K. data. We verified this by running our HWT implementation on the U.K. data—lack of climate zone breakdowns prevents running our algorithms.

### C. Day-Ahead Forecast Accuracy

This analysis is motivated by the day-ahead markets in the U.S. Each utility in an ISO/RTO submits its hourly load for the next day (24 hours) at a certain hour each day (e.g., 12:00 noon for New England ISO). These load forecasts are used to schedule power generation and to calculate the wholesale unit price for each hour of the next day. Figure 17 plots the results for Data Set 1 in 2009, and the ERCOT total load in 2010. Static and Rolling achieve constant MAPEs and both of them outperform HWT by a large margin at every hour of the day. The performance difference between our method and HWT is particularly high for the ERCOT total load. The similar behavior is also observed in all weather zones but the North zone, where HWT outperforms the static method by a small margin between hours 20 and 24. Nevertheless, Rolling performs better than HWT at all hours in the North zone.
MAPE

Fig. 16. Forecast accuracy as a function of the forecast lead time for the North and East weather zones in 2010.

MAPE

Fig. 17. Forecast accuracy as a function of the hour of the day for Data Set 1 (SC) load in 2009 and the ERCOT total load in 2010. (At every hour of the present day, MAPE for the next day is calculated and averaged over all days.)

D. Peak & Valley Hour Load

The forecasts for the peak hour (the highest load hour) and the valley hour (the lowest load hour) load in the next day are important inputs to the unit commitment and economic dispatch problems solved by ISOs/RTOs on a daily basis, and affect the unit cost of electricity at different time periods in the day-ahead markets. This analysis is designed to test how good our method can forecast the peak and valley hour load for the next day as a function of the hour of the current day. We calculated the MAPE between the forecast and the actual load for the peak and valley hours of each day, which are given in Figure 18.

Our method shows a high level of forecast accuracy with a robust performance. Both Static and Rolling outperform HWT at almost all hours in both of the data sets. Even though the forecast accuracy of HWT improves as the hour approaches the end of the day—especially for the valley hour load—the MAPE values are quite high. All MAPE for the peak hour load is above 9% level in both load series. The behavior in Figure 18 is observed in almost all weather zones in the ERCOT data with the exceptions of (i) the East zone valley hour load where HWT performs better than Static, but worse than Rolling, and (ii) the North zone peak hour load where HWT outperforms both versions of our method.

E. Weather Forecast Accuracy

We have analyzed the performance and accuracy of our algorithm for forecasting the electricity load under perfect weather forecast. However, perfect weather forecasts are an unrealistic assumption and the weather prediction is subject to uncertainty. For these reasons, we study the performance of our algorithm under weather forecast uncertainty. We obtain data from NOAA that gives us the algebraic, mean and root square mean error for the maximum and minimum daily temperature averaged for each month. We also obtained several years of data. Our objective is to model from these data the probability distribution of the errors so we can add synthetic noise to our data. Our objective is to model from these data the probability distributions and correlations. For each month \( p \) we have the daily error observations \( \{x_1^{(p)}, x_2^{(p)}, \ldots, x_{N_p}^{(p)}\} \) and \( \{y_1^{(p)}, y_2^{(p)}, \ldots, y_{N_p}^{(p)}\} \) where \( N_p \) is the number of days in month \( p \). We can reasonably assume that the random variables \( x_i^{(p)} \) and \( y_j^{(q)} \) are independent if \( i \neq j \) or \( p \neq q \). Analogously for the random variables \( y \). This would mean that errors are independent for different days. Of course there could be correlations between \( x_i^{(p)} \) and \( y_j^{(p)} \).

The data we have from NOAA are the daily average and mean errors for each month. More precisely, we have for each month \( p \) the observations

\[
\bar{a}_p = \frac{1}{N_p} \sum_{i=1}^{N_p} x_i^{(p)} \quad \text{and} \quad \text{mse}_p = \frac{1}{N_p} \sum_{i=1}^{N_p} |x_i^{(p)}|^2
\]
\[ ay_p = \frac{1}{N_p} \sum_{i=1}^{N_p} y_i^{(p)} \quad \text{and} \quad \text{mse}_y = \frac{1}{N_p} \sum_{i=1}^{N_p} |y_i^{(p)}|^2. \]

It is relatively easy to compute the mean, \( m_x \) and \( m_y \), of the random variables \( x \) and \( y \) by just computing the mean of the random variables \( ax \) and \( ay \) (we do have several months of observations). Analogously, we can estimate their variances reasonably well. More precisely, it is not hard to see that

\[ \sigma_x^2 = N_p \sigma_{ax}^2 \quad \text{and} \quad \sigma_y^2 = N_p \sigma_{ay}^2 \]

If we knew beforehand that these random variables were Gaussian then the problem would be easy and we would only need to compute the standard deviations, \( \sigma_{ax} \) and \( \sigma_{ay} \), of the variables \( ax \) and \( ay \) and from them we would get the variance of the random variables \( x \) and \( y \). Since Gaussian random variables are determined by their mean and variance we would be done. However, we do not know if this is the case.

The problem is that the probability distribution of the variables \( ax \) and \( ay \) is, up to a constant factor, the \( N_p \)-fold convolution of the probability distribution of \( x \) and \( y \) respectively. For this reason, we compute the pdf of the variables \( ax \) and \( ay \) and then modify it to have the right variance by stretching the scale.

Applying the previous analysis we would have a proxy for the individual probability distributions of \( x \) and \( y \). Now estimating their correlations is easier. More specifically, we want to estimate \( \sigma_{xy}^2 := \mathbb{E}[(x - m_x)(y - m_y)] \), which can be found by

\[ \frac{1}{P} \sum_{p=1}^{P} (ax_p - m_x)(ay_p - m_y) \]

where \( P \) is the total number of months available.

The next problem is how to sample pairs of random variables \((x, y)\) with the right statistics. We first sample the random variables \( \tilde{x} \) and \( \tilde{y} \) according to the statistics of \( x \) and \( y \) respectively and in an independent fashion (hence \( \sigma_{xy}^2 = 0 \)). Recall that we do have good estimates for \( m_x, \sigma_x, m_y, \sigma_y \) and \( \sigma_{xy} \). Let \( \lambda \) be such that \( \frac{2\lambda}{1 + \lambda^2} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \) and define

\[ \tau = \frac{\sigma_x}{\sqrt{1 + \lambda^2}} \left( \frac{\tilde{x} - m_x}{\sigma_x} + \lambda \frac{\tilde{y} - m_y}{\sigma_y} \right) + m_x, \]

\[ \gamma = \frac{\sigma_y}{\sqrt{1 + \lambda^2}} \left( \lambda \frac{\tilde{x} - m_x}{\sigma_x} + \frac{\tilde{y} - m_y}{\sigma_y} \right) + m_y. \]

It is easy to check that these random variables have the right variances and covariances. Moreover, under the assumptions that the true probability distributions of \( x \) and \( y \) (apart from their mean and variance) are not too dissimilar then the pair \((\tau, \gamma)\) should have the right statistics.

**Numerical Experiments:** We implemented the weather forecast error model described above and generated random weather forecasts to test our Static algorithm. The NOAA data contain weather forecast errors for 12 and 24 hour forecast lead times. For each forecast lead time, we created 20 random weather series for year 2009, fed each series to the static algorithm together with the actual load and weather data for 2008, and calculated the forecast accuracy with a 95% confidence interval around the mean. Table III shows the MAPE values for the static algorithm with and without weather forecast errors (Static-wfe and Static-pw) for Forecast Lead Time (Section VI-B), Day-Ahead Load Forecasting (Section VI-C) and Peak & Valley Hour Load (Section VI-D) Analysis. As expected, incorporating weather forecast errors yields higher MAPE figures. In addition, 12-hour weather forecast lead time results in a better load forecast accuracy in comparison to 24-hour lead time. Although weather forecast errors have a negative effect on the accuracy of our method, the performance comparison between our method and HWT is not affected much. While the static method with perfect weather outperforms HWT on and beyond a forecast lead time of 6 hours, this threshold increases to 7 hours in the presence of weather forecast errors. For the day-ahead prediction and the peak hour load analysis, our static algorithm outperforms HWT at every hour of the day even in the presence of weather forecast errors. The results for the valley hour load are mixed. The static algorithm under the weather forecast errors is better than HWT until hour 22 beyond which HWT outperforms by a small margin.

We also conducted numerical experiments with some of the weather regions (Coastal, East and South Central) of Data Set 2, and observed similar trends.

**TABLE III**

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Lead Time</th>
<th>Static-wfe</th>
<th>Static-pw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast Lead Time</td>
<td>24-hour</td>
<td>8.04±0.06</td>
<td>6.89</td>
</tr>
<tr>
<td></td>
<td>12-hour</td>
<td>7.78±0.06</td>
<td>6.89</td>
</tr>
<tr>
<td>Day-Ahead</td>
<td>24-hour</td>
<td>6.78±0.08</td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>12-hour</td>
<td>6.41±0.10</td>
<td></td>
</tr>
<tr>
<td>Peak Hour Load</td>
<td>24-hour</td>
<td>9.88±0.14</td>
<td>8.27</td>
</tr>
<tr>
<td></td>
<td>12-hour</td>
<td>9.56±0.11</td>
<td></td>
</tr>
</tbody>
</table>

**VII. DISCUSSION AND CONCLUSION**

In this paper, we present a weather-based short-term load forecasting method and test its performance using hourly load data sets from the U.S. Our method extracts the load component for weather and natural illumination by least-squares fitting for Steadman apparent temperature and log-scale illumination simultaneously for each day type and hour of the day. We use temperature, humidity and atmospheric opacity as weather related input variables as well as the positions of the sun and the moon for estimating the natural illumination. Subtracting this component gives the residual load, which is analyzed by FFT spectral analysis.

We test the performance of our method extensively using one load data set from a mid-size U.S. city, and another from 8 regions of the state of Texas. These data sets provide a wide range of population and weather heterogeneity. For the empirical study, we consider the method described in this paper (referred to as Static) and a rolling horizon version of
it (referred to as Rolling). We use J.W. Taylor’s exponential smoothing based method of HWT as a benchmark.

We conduct four types of empirical analysis. The first one tests the forecast accuracy up to two days ahead. Both Static and Rolling algorithms are robust and yield almost constant mean absolute percentage error (MAPE) while HWT shows a cyclical behavior and better MAPE at short forecast lead times. The results show that HWT outperforms Rolling method up to a few hours ahead (e.g., 3 hours for the total ERCOT load in 2010 and 4 hours for the U.S. city load in 2009), beyond which the accuracy of HWT degrades rapidly. This observation is in line with the result in [14] that HWT outperformed a weather-based forecasting tool used by the British transmission company up to five hours ahead, but beyond that the weather-based method was better. It is clear that HWT is a good forecasting method for very short lead times making it an attractive forecasting tool for supporting electricity purchases in the spot markets. Our method offers better forecasts beyond a forecast lead time threshold, so it would be preferable for any decision process that requires forecasts longer time ahead.

The second analysis, which is motivated by the day-ahead markets, focuses on next day forecast accuracy as a function of the hour of the present day. Static and Rolling shows robust performance with Rolling having better accuracy. HWT is not competitive at any hour of the day resulting in high MAPE. The third analysis looks at how well our method can forecast the load in next day’s peak and valley hour load as a function of the hour of the present day. Rolling outperforms HWT in forecasting peak and valley hour load in both data sets. The performance difference is high especially for the peak hour load. Finally, the last analysis explores the effect of weather forecast errors on the load forecast accuracy of our approach. We first develop a model to estimate the hourly weather forecast errors from the monthly aggregated data that we obtain from NOAA. Then using this model, we create random weather forecasts and apply the Static algorithm to compute the load forecast errors. As expected, errors are higher than the case with perfect weather, but the trends and conclusions deduced from the first three analysis remain the same.

One general observation as a result of this empirical study is that the Rolling algorithm consistently outperforms the Static algorithm suggesting that rerunning the entire algorithm on a frequent basis with the new load data improves the accuracy. The results show the relatively high accuracy of our method, especially in forecasting loads for the next day in fine time granularity, and highlights its robustness. These characteristics make our method a good forecasting tool for supporting decisions like unit commitment, economic dispatching, buying electricity in day-ahead markets, developing bidding strategies, etc.

One direction for future research is to adopt our method to building-level load forecasting, which is underway.

ACKNOWLEDGMENT

The authors would like to thank James W. Taylor for valuable discussions, and Brenton MacAloney of NOAA for providing the weather forecast accuracy data.

REFERENCES