Propagating Uncertainty in Tree-Based Load Forecasts

Abdolrahman Khoshrou, Eric J. Pauwels

Centrum Wiskunde & Informatica, The Netherlands
a.khoshrou@cwi.nl, eric.pauwels@cwi.nl

Abstract

This paper discusses the use of ensembles of regression trees as a straightforward but versatile methodology to generate short term (day-ahead) load forecasts for real data from the Global Energy Forecasting Competition 2014. Since temperature is a strong predictor of load, we investigate how forecast uncertainty in temperature can affect the performance of the prediction model. To this end, a singular value decomposition (SVD) based approach is harnessed to simulate noisy but realistic temperature profiles. Our results show that as long as uncertainty is not exceedingly large, it is worthwhile to include temperature forecasts as predictors.

1. Introduction and Related Work

Due to the increasing integration of intermittent renewable energy sources and the lack of affordable large scale storage, balancing electricity supply and demand is becoming more challenging. Consequently, in today's competitive and dynamic environment, an accurate load forecast is highly desirable both from a technical and an economic point of view. Indeed, without accurate forecasts, issues like increasing rates, brownouts or even black-outs are inevitable [1].

Depending on the time horizons, the prediction of the power (load) distribution is classified into short, medium and long term forecasting. The context of more than a couple of months to years in load prediction is studied in long term load forecasting (LTLF). It mainly assists in planning on setting up new power plants. Medium term load forecasting (MTLF) is associated with forecasts targeting few weeks to few months ahead, whereas short term load forecasting (STLF) deals with load estimations for the next few hours to few days. The former is usually done for balance sheet calculations, risk management, purchasing energy and pricing plans. The latter plays a key role in unit commitment and load dispatching. In this paper in addition to proposing a novel STLF method, the effect of uncertainty in the predictors on the performance of the proposed model is investigated.

In the recent literature, there are numerous methods to forecast electricity load over different time horizons. Artificial intelligence (AI) based methods such as artificial neural network [2], support vector machines [3] and fuzzy methods [4] are popular, mostly due to their robustness and ability to tackling the non-linearity between predictors and the target variables. Statistical methods, on the other hand, are mostly popular in the econometric studies, due to the interpretability of their results. Members of this category include regression models, autoregressive models, heteroskedastic models and so on. A bi-level prediction strategy for STLF of micro grids using evolutionary algorithm and neural networks is proposed in [5]. The reported work has the advantage of using an enhanced differential evolution algorithm in upper level to optimize the performance of the forecaster in lower level. In [6], a self organizing map (SOM) approach is introduced to group the load profiles in an unsupervised manner. Each identified cluster is then fed to individual support vector regression (SVR) models to predict the daily profile. Recently, modeling and forecasting the trend-seasonal components and probabilistic (beyond point) forecasts have attracted a lot of attention [7].

Problem statement and contribution As briefly explained, numerous prediction techniques have been applied to STLF problem. In the present work, we opt for a relatively simple yet versatile technique: ensembles of regression trees, as they are better suited to address the heterogeneity of the data (see e.g., [8]). Due to the characteristics of the available data, it is particularly advantageous to include temperature as a predictor. It is therefore of great interest to investigate how sensitive the results are to the noise level in this input variable. To model this uncertainty, we expand the available temperature time series in data-driven orthogonal components for which the simulation becomes straightforward.

2. Data

To demonstrate the robustness of the proposed method and to make the results reproducible, a case study was constructed based on publicly available data (average temperature and aggregated load) from the Global Energy Forecasting Competition 2014 (GEFCom2014) [7]. The challenge in this competition was to construct a probabilistic forecast of the (aggregated) load. Our focus nonetheless is on developing a point

This work was supported in part by the Dutch STW under project grant *Smart Energy Management and Services in Buildings and Grids (SES-BE).*

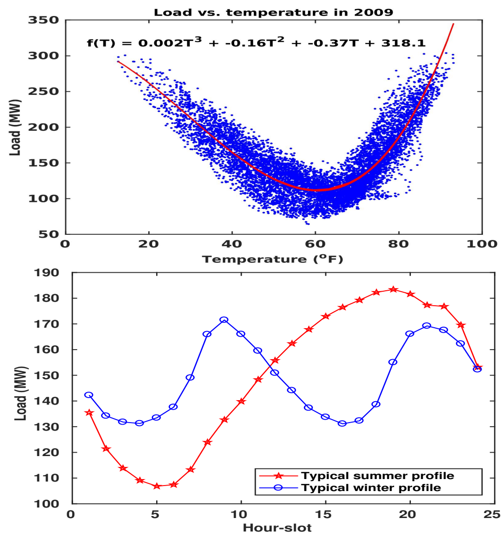

Figure 1. Top: scatter plot of the load vs. temperature (in 2009). An ad-hoc customized polynomial ($f(T) = 0.002T^3 - 0.16T^2 - 0.37T + 318.1$) is superimposed on that. Bottom: comparison of the typical daily load profiles in winter and summer times.

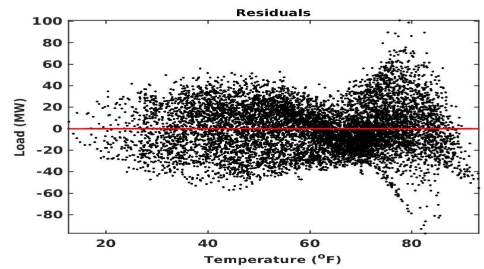

Figure 2. Residuals of the fitted custom polynomial in the top panel of Fig. 1.

forecasting method. Obviously, to design a forecasting model, load history is needed. A large portion of the electricity is being used by the heating and cooling systems in the winter and summer times. Consequently, load is considerably lower in spring and autumn as the temperature is more moderate. This aspect of the data is clearly visible in the top panel of Fig. 1. The dependence on temperature is also responsible for the substantial change in the load profiles over the year as illustrated in the bottom panel of Fig. 1. Furthermore, the consumption patterns also change based on the work schedule, hours of the day, days of the week, and so on. Therefore, calendar information, holidays, and special event information are also of great value. Although temperature is an important (perhaps, the most important) predictor of load there are a number of reasons why, in and of itself, it is not sufficient:

1. As can be seen from the data analysis, load patterns during week and weekend days are substantially different. This reflection of human activity is of course absent in the weather patterns. One can hence expect that the same temperature will result in different load patterns depending on whether we are looking at a week or weekend day.
2. Path dependence: Similar temperatures might also result in different behaviour depending on the immediately preceding situation. For instance, relatively high temperatures in early spring might not result in massive AC activation since people might welcome the change in weather after a cold winter. This would be very different in the summer or autumn.

As explained before, the overall relationship between load and temperature is clear: high and low temperatures require more cooling or heating respectively, and therefore result in higher loads (top panel of Fig. 1). However, it is plain to see that the polynomial fits do not sensibly capture the shape of the observed relationship (Fig. 2) and a more complex model is needed. Before explaining the methodology in Section 4, a brief introduction to singular value decomposition (SVD) technique is brought in Section 3. Its importance lies in the fact that, SVD was used to generate new temperature profiles for uncertainty quantification purposes (Section 4.3). We conclude our work in Section 5.

3. SVD-Based Representation of the Data

Time series data in smart energy systems often show two (or more) distinct time scales. The data exhibit strong diurnal patterns reflecting the daily (or weekly) rhythms of human activities. Apart from that, these relatively fast diurnal patterns are superimposed on slower seasonal variations that have a significant impact on the overall structure of the data. Recasting such a time series as a matrix, in which each column represents the data for a single day, can be helpful in gaining more insight into the data. The advantage of this recasting is two-fold. First, the resulting matrix can be displayed as an image, allowing one to scrutinize subtle or faint features. Second, one can draw on well-established matrix decomposition methods to elucidate underlying data structure. To illustrate the latter point using the data at hand, let's take one year's worth of aggregated hourly load values (L say). We can then recast L as a matrix in which each column represent the data for a single day (i.e. $L \in \mathbb{R}^{24\times 365}$). As a consequence, the matrix L can be satisfactorily represented by a low-rank approximation. Singular value decomposition (SVD) provides us with an efficient algorithm to compute such low-rank approximations [9]. More specifically, given an arbitrary $h \times d$ matrix $A \in \mathbb{R}^{h\times d}$, there exists orthogonal matrices $U \in \mathbb{R}^{h\times h}$ and $V \in \mathbb{R}^{d\times d}$ (both with orthonormal columns) such that:

$$A = USV^T = \sum_{k=1}^{r} \sigma_k U_k V_k^T \qquad (1)$$

where S has the same size as A, and its non-zero elements (singular values: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$) are

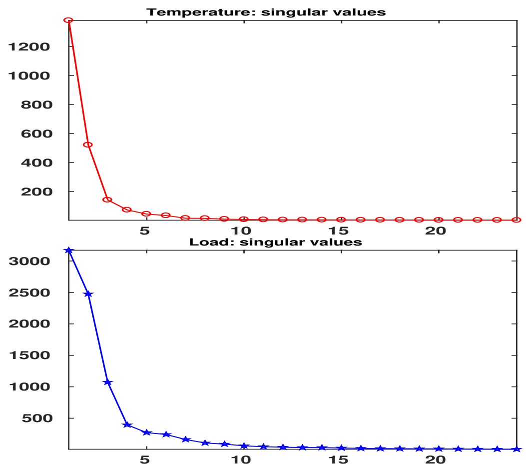

Figure 3. Singular values obtained for load and temperature daily profiles in 2009 (zero annual mean values). Temperature values (top) suggest that a reconstruction of rank-3 approximation would suffice, indicating that temperature is quite regular. Load (bottom) on the other hand, requires a rank-4 or 5 approximation.

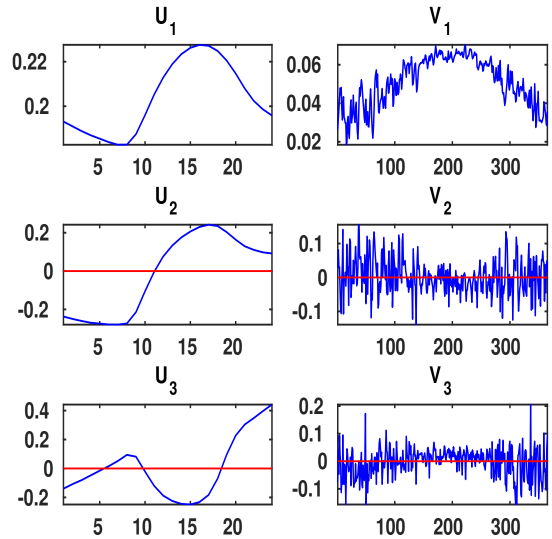

Figure 4. SVD-based decomposition of hourly temperature data for 2009. Left column are the first three columns U_k whereas the right column shows the corresponding V_k's.

uniquely positioned on the main diagonal, in descending order. Furthermore, U_k and V_k denote the k^{th} column of U and V, respectively, and $r = \min\{h, d\}$ (see e.g. [9]). If there is only a small number of dominant singular values (as is the case for the load and temperature data in Fig. 3), then the expansion in (1) can be truncated after the first few terms to yield an adequate approximation A_p of (lower) rank p:

$$A_p = \sum_{k=1}^{p} \sigma_k U_k V_k^T \quad \text{where} \quad p < r. \tag{2}$$

To elaborate more, Fig. 4 illustrates a plot of the first three columns of U_k (left) and V_k (right) for the 2009 hourly temperature data. The columns U_k can be interpreted as daily profiles and successive increments, while the coefficients of V_k represent the corresponding scaling factors. Put differently, the original time series is represented as a linear combination of the (data-driven) profiles specified by the columns of U while the V columns provide the corresponding coefficients. For instance, looking at figure, we clearly recognize in U_1 (top left panel) an averaged daily temperature profile, whereas the corresponding coefficients in V_1 outline daily temperature evolution over the year (top right panel). The middle panels display the most dominant corrective incremental profile U_2 (left) and the corresponding coefficients V_2 (right) which needs to be added to the first profile to get a better approximation. Similarly for the third profile (bottom). Looking at the right column one gets the distinct impression that temperatures are less variable during the summer (middle parts). In Section 4.3 we will use this decomposition to investigate how uncertainty affects the forecasting results.

4. Prediction using Tree Ensembles

After the data exploration above, this section focuses on our approach to the point (single-valued) forecasting problem. We have opted to use an ensemble of regression trees to predict the daily load prognoses given the daily temperature profiles, date and time. The rationale underpinning this choice is that trees are well suited to handle the heterogeneous input data which comprise both continuous and discrete variables. In addition, tree models are modular in that new predictors can easily be added. Furthermore, it is well-known that ensemble models are less prone to overfitting. So, tree ensemble models promise to strike a good balance between flexibility and generalizability.

4.1. Basic methodology

The individual (weakly trained) regression trees in the ensemble are constructed using least-squares boosting method [10]. We have tried three different models (see below), which all have been trained on years 2005 through 2009, and tested on the data from 2010. In all of the following cases, the aim is to predict the 24 hourly load values for the next day d ($L(d)$). The common inputs in all three models are discrete values such as month of the year ($M \in [1:12]$), day of the week ($W \in [1:7]$), and hour of the day ($H \in [1:24]$). Furthermore, we assume that the load and temperature profiles for the previous day ($d-1$) are also available.

Model 1: The first case, we only use information which is available on day $d-1$. In particular, we do *not* use the predicted temperature profile for day d. This provides us with a baseline performance gain which we will gauge the next models.
Predictors: $[M, W, H, L(d-1), T(d-1)]$.

Model 2: In this and the next model we do use the actual(!) temperature profile $T(d)$ to predict the load on that day. The rationale for this is that fairly accurate temperature predictions for day d are available on day $d-1$. Furthermore, in Section 4.3 we will try and quantify the amount of additional uncertainty this assumption can introduce in the forecast.
Predictors: $[M, W, H, L(d-1), T(d-1), T(d)]$.

Model 3: This model expands the previous one by adding first and second derivatives of the temperature and load profiles. The thinking behind this expansion of variables is that oftentimes the actual values are not as important as the general underlying trend $((.)'$: first derivative) or trend changes $((.)''$: second derivative).
Predictors: $[M, W, H, L(d-1), L'(d-1), L''(d-1), T(d-1), T'(d-1), T''(d-1), T(d), T'(d), T''(d)]$.

4.2. Experimental Results

The accuracy of the forecasts is evaluated using a conventional method, i.e., the mean average percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \qquad (3)$$

where y_t is the hourly value of the load profile from GEFComp2014 timeseries and $\hat{y}_t$ is the corresponding forecast value (using one of the models specified above).

Table 1. MAPE (in %) results for the three models

	Train (2005-9)	Test (2010)
Model 1 (24 hrs)	9.57	10.35
Model 2 (24 hrs)	4.42	5.06
Model 3 (24 hrs)	3.95	4.81

4.3. Sensitivity to Uncertainty on Temperature Forecasts

As pointed out above, we have used the temperature profiles for the next day as a proxy for the temperature forecasts. For that reason the error rates reported in Table 1 are over-optimistic and we need to attempt to quantify the extra amount of uncertainty that results from using forecasts rather than actual values.

Unfortunately we do not have access to the historic forecast data, and it is therefore difficult to quantify the corresponding amount of uncertainty directly. To investigate how sensitive our results are to additional uncertainty, we therefore add noise to the input temperature profiles and feed these perturbed inputs into the tree ensemble, upon which we can compute the corresponding change in MAPE. However, simply adding independent Gaussian noise to the hourly values of individual temperature curves results in unrealistically jagged profiles.

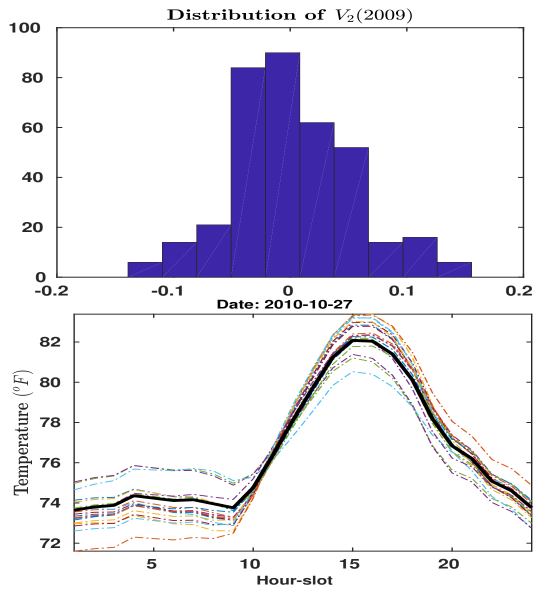

Figure 5. Top: Histogram of the V_2 coefficients for the 2009 temperature data (365 values). Note that the distribution is approximately normal with zero mean and $std(V_2) \approx 0.05$. Bottom: Twenty examples of generated noisy temperature profiles. The solid line is the actual profile. Noise was generated by tweaking the V_2, V_3 and V_4 coefficients to which we added independent $\mathcal{N}(0, \epsilon^2)$ noise (with $\epsilon = 0.01$).

We therefore propose to use the SVD decomposition results to create realistic noise. From the training set (covering 2005 through 2009) we know that the singular values σ_k as well as the daily (incremental) profiles U_k show negligible change over the years. So we can reuse them for the test year 2010. The real variation is in the coefficients V_k which behave much more erratically from day to day (Fig. 4). So in order to create noisy temperature profiles we proceed as follows:

1. For each of the coefficients V_2 through V_4 we estimate the corresponding standard deviation $s_k = std(V_k)$. The histogram for V_2 is shown in the top panel of Fig. 5, which shows that $s_2 \approx 0.05$. In fact, from the data analysis it turns out that all three standard variations s_2 through s_4 have similar values of about 0.05. (Recall however that the contribution to the final profile is scaled up or down by the corresponding singular value for which we know: $\sigma_2 > \sigma_3 > \sigma_4$).

2. Next, for any particular day d in the test year for which we want a forecast, we take the actual temperature profile for that day $T = T(d)$, compute the corresponding SVD coefficients $v_1^0, v_2^0 \dots v_4^0$ and then perturb them by adding zero-mean Gaussian noise: $v_k^n = v_k^0 + \mathcal{N}(0, \epsilon^2)$. These perturbed SVD coefficients are then used to reconstruct a noisy version of the temperature profile. An example (for 20 different noise samples) is shown in the lower panel of Fig. 5.

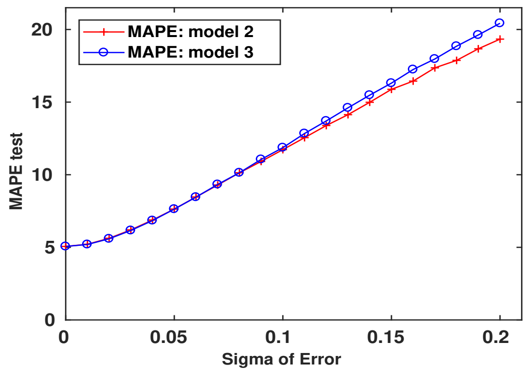

Figure 6. Impact on MAPE (for models 2 and 3) of uncertainty on temperature profile. The *x*-axis displays the standard deviation ϵ of the Gaussian noise applied to the $V_2 : V_4$ coefficients. Notice how model 3 is doing slightly worse than model 2 for large values of ϵ.

3. To quantify the effect of temperature uncertainty on the forecast results we generate for each actual day profile $T(d)$ one hundred perturbed profiles according to the scheme outlined above. All of these profiles are fed into the prediction model and the forecasts are duly compared to the actually observed values. This allows us to compute the corresponding MAPEs.
4. For completeness' sake, we point out that we did not perturb V_1 as this is a proxy of the average temperature on a particular day, for which uncertainty is negligible. Similarly, there is little to be gained from perturbing higher order coefficients (V_5 etc) as their impact on the profile is slight.

The results of these experiments (for models 2 and 3) are shown in Fig. 6, where we plot MAPE as a function of the standard deviation ϵ of the Gaussian noise imposed on the SVD coefficients V_2 through V_4. When there is no uncertainty (i.e. $\epsilon = 0$) the results for both models are the ones reported in Table 1 (i.e. $MAPE \approx 5$). Increasing ϵ to about 0.05 (the standard deviation seen in the training data) inflates the MAPE of both models to about 7.5. Furthermore, the MAPEs attain a value of 10 for $\epsilon \approx 0.1$ which means that for that amount of noise it is no longer advantageous to include a temperature forecast as a predictor. Put differently, models 2 and 3 are then doing worse than the baseline model 1. Finally, for even larger values of ϵ model 3 does slightly worse than model 2. This is probably due to the inclusion of derivatives in model 3, which are well-known to be more sensitive to noise.

5. Conclusion and Future work

In this paper we discussed the use of ensembles of regression trees as a straightforward but versatile methodology to generate short term (day-ahead) load forecast for real data from the Global Energy Forecasting Competition 2014 . Since load strongly depends on temperature (heating and cooling), performance of the prediction models is significantly boosted if temperature is included as a predictor. However, for real-time day-ahead prediction, actual temperatures are not available, only forecasts. We therefore investigated how uncertainty on the temperature forecasts affects the performance of the prediction model. To this end, we introduced an SVD-based noise model and showed that as long as uncertainty is below a certain threshold, it is worthwhile to include temperature forecasts as predictors. We intend to extend this methodology towards a larger class of probabilistic load forecasting tasks.

6. References

[1] T. Hong *et al.*, "Energy forecasting: Past, present, and future," *Foresight: The International Journal of Applied Forecasting*, no. 32, pp. 43–48, 2014.

[2] K. A. Keitsch and T. Bruckner, "Modular electrical demand forecasting framework-a novel hybrid model approach," in *Systems, Signals & Devices (SSD), 2016 13th International Multi-Conference on*. IEEE, 2016, pp. 454–458.

[3] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356–4364, 2013.

[4] A. Badri, Z. Ameli, and A. M. Birjandi, "Application of artificial neural networks and fuzzy logic methods for short term load forecasting," *Energy Procedia*, vol. 14, pp. 1883–1888, 2012.

[5] N. Amjady, F. Keynia, and H. Zareipour, "Short-term load forecast of microgrids by a new bilevel prediction strategy," *IEEE Transactions on smart grid*, vol. 1, no. 3, pp. 286–294, 2010.

[6] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 392–401, 2006.

[7] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[8] T. Mehenni and A. Moussaoui, "Data mining from multiple heterogeneous relational databases using decision tree classification," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1768–1775, 2012.

[9] G. Strang, *Introduction to linear algebra*. Wellesley-Cambridge Press Wellesley, MA, 1993, vol. 3.

[10] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.