Forecasting Uncertainty in Electricity Demand

Tri Kurniawan Wijaya
School of Computer and Communication Sciences
EPFL, Switzerland

Mathieu Sinn and **Bei Chen**
IBM Research – Ireland

Abstract

Generalized Additive Models (GAM) are a widely popular class of regression models to forecast electricity demand, due to their high accuracy, flexibility and interpretability. However, the residuals of the fitted GAM are typically heteroscedastic and leptokurtic caused by the nature of energy data. In this paper we propose a novel approach to estimate the time-varying conditional variance of the GAM residuals, which we call the GAM^2 algorithm. It allows utility companies and network operators to assess the uncertainty of future electricity demand and incorporate it into their planning processes. The basic idea of our algorithm is to apply another GAM to the squared residuals to explain the dependence of uncertainty on exogenous variables. Empirical evidence shows that the residuals rescaled by the estimated conditional variance are approximately normal. We combine our modeling approach with online learning algorithms that adjust for dynamic changes in the distributions of demand. We illustrate our method by a case study on data from Réseau de transport d'électricité, the operator of the French transmission grid.

1 Introduction

Forecasting electricity demand is a key instrument in operational and planning processes of electric utilities. In this paper, we focus on short-term forecasts (with a forecast horizon of 24-48 hour ahead), which are required, e.g., by electricity suppliers to bid generation/load into electricity markets, and by network operators for day-ahead outage planning. Due to an ever growing population and carbon emission, however, the electricity sector has been changing dramatically in the last few years, which made electricity demand forecasting even more challenging. Electric vehicles (EVs), for example, pose a new challenge to electricity grids by drawing a large amount of energy in a very short time. Charging one EV can consume 32 kWh, which is comparable to one household's daily consumption (Ramchurn et al. 2012). Moreover, the integration of renewable energy sources to the grid, as an effort to reduce our dependency on fossil fuel, adds additional complexity to efforts at balancing supply and demand, since their generation is intermittent and unpredictable.

Therefore, it is imperative for utility companies and networks operators to forecast not only the conditional expectation but also the uncertainty of the future demand, compute the risk/benefit associated with it, and incorporate it into their planning process. However, a large body of literature so far has been focused on single-valued, "point" forecast to estimate the conditional mean of the future demand. Compared to the point forecast, forecasting uncertainty is more challenging since it requires us to estimate the entire distribution of the future demand. In this paper, we propose a novel approach to address the problem by modeling the time varying conditional mean and variance of the future demand using Generalized Additive Models (GAM).

Among other forecasting methods in the literature (see Section 2), GAM has been increasingly popular due to its accuracy, flexibility, and interpretability (Wood 2006; Hastie et al. 2009). Those advantages have made GAM attractive also for energy analytics where understandability is a key criteria for a model to be deployed by utility companies. While other (accurate) models are typically opaque and difficult to interpret, GAM consists of transfer functions that are easy to understand.[1] Additionally, the "transparency" of GAM opens possibilities for practitioners to discover new insights about the relationship between some exogenous and response variables, or the other way around, to spot a potential overfitting when a relationship does not conform a well-established physical law.

The contributions of this paper are as follows. We propose a novel GAM^2 (or *GAM squared*) algorithm to forecast uncertainty in electricity demand (Section 3.1). First, we use a GAM to model the conditional expectation of the demand. The residuals of the fitted GAM, however, are typically heteroscedastic and leptokurtic due to the nature of energy data. To this end, we apply a second GAM to the squared residuals to explain the dependence of uncertainty on exogeneous variables and estimate the (time-varying) conditional variance. Under normality assumptions, the estimated conditional mean and variance allow us to construct prediction intervals. Although we showcase our method specifically for

[1] See, e.g., Figure 1 in (Ba et al. 2012), Figure 4 in (Cho et al. 2013).

electricity demand, it can also be applied to other domain, as long as the rescaled residual is (approximately) normal. The online learning mechanism in Section 4.1 is based on (Ba et al. 2012). Then, we propose a novel algorithm that adjust the prediction intervals to the non-stationary nature of electricity demand (Section 4.2). Finally, we illustrate the effectiveness of our approach on real electricity demand data provided by Réseau de transport d'électricité, France (Section 5).

2 Related work

Electricity demand forecasting There exists a rich body of literature on point forecasts, i.e., forecasting the mean demand, typically conditional on a number of exogeneous variables such as time of day, weekday, temperature etc. Various techniques have been considered and applied both in experimental settings and real system designs, e.g., regression (Papalexopoulos and Hesterberg 1990; Hong 2010), Artificial Neural Networks (Khotanzad et al. 1997; Hippert, Pedreira, and Souza 2001), (Seasonal) ARMA (Huang and Shih 2003; Taylor 2010), Boosting (Taieb and Hyndman 2014), Support Vector Regression (Chen, Chang, and Lin 2004; Sapankevych and Sankar 2009), and GAM (Fan and Hyndman 2012; Ba et al. 2012; Cho et al. 2013). With the ongoing roll-out of smart metering in many countries worldwide, there are also other approaches that aim to improve the demand forecast by first segmenting customers into several clusters, forecast each cluster separately, and then aggregate the forecast from the clusters into a single forecast (Misiti et al. 2010; Humeau et al. 2013; Alzate and Sinn 2013). Some other approaches have also used demographic/survey information (Mohamed and Bodger 2005; Kolter and Ferreira 2011; Jarrah Nezhad et al. 2014).

Probabilistic forecasting Gneiting and Katzfuss (2014) provide a good overview on probabilistic forecasting. It has been an increasingly important direction to model uncertainty in various fields, such as healthcare (Jones and Spiegelhalter 2012), politics (Montgomery, Hollenbach, and Ward 2012), weather (Palmer et al. 2005; Warner 2011), and finance (Groen, Paap, and Ravazzolo 2013). Probabilistic forecasting has also drawn more and more interest in the energy domain, e.g., for long-term (Hyndman and Fan 2010; Hong, Wilson, and Xie 2014) and short-term electricity demand forecasting (Fan and Hyndman 2012), solar power (Bacher, Madsen, and Nielsen 2009), and wind power forecasting (Wytock and Kolter 2013). While this paper was being written, even a public competition has been dedicated to probabilistic energy forecasts.[2] Probabilistic forecasts typically rely on computationally expensive approaches such as simulation, bootstrapping, or ensemble forecasting, which all require running multiple forecast to generate prediction intervals. In contrast, our approach constructs prediction intervals directly using the estimated conditional mean and variance. Thus, it does not require multiple runs of forecasting simulations/scenarios.

[2]See GEFCom2014, `http://www.drhongtao.com/gefcom`.

3 Modeling uncertainty

3.1 The GAM^2 algorithm

We use the following general regression model for electricity demand:

$$Y_t = \mu(x_t) + \sigma(x_t)\epsilon_t \tag{1}$$

where y_t is the demand at time t, x_t is a vector of covariates (e.g., weekday, time of day, temperature, etc) and ϵ_t is the error, which we assume to have zero mean and unit variance. Note that under this assumption, the conditional mean and variance of Y_t (given x_t) are obtained by

$$\begin{aligned} \mathbb{E}[Y_t] &= \mu(x_t), \\ \mathrm{Var}(Y_t) &= \sigma^2(x_t). \end{aligned}$$

In practice, the functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are unknown and need to be estimated from empirical data. A wide range of methods has been explored in the literature for modeling the conditional mean demand function $\mu(\cdot)$, including Artifical Neural Networks, Support Vector Regression, seasonal time series models, and semi-parametric regression. In this paper, we use GAM and express the conditional mean function as

$$F(\mu(x_t)) = \sum_{i=1}^{I} f_i(x_t),$$

where $F(\cdot)$ is a link function (e.g., the identity or logarithm function). The $f_i(\cdot)$ are called *transfer functions* and have the following form:

$$f_i(x_t) = \mathbf{1}_{(x_t \in A_i)} \beta_i^T b_i(x_t). \tag{2}$$

Here $\mathbf{1}_{(\star)}$ denotes the indicator function which returns 1 if the expression $\star$ is evaluated to true and 0, otherwise, and $b_i(\cdot)$ is a vector of *basis functions* (we mostly use cubic b-splines), typically depending on one or two continuous variables in the vector of covariates x_t. The indicator function allows for "switching on/off" transfer functions, e.g., to model the effect of the time of day depending on the weekday. See the experimental sections for examples.

We learn the GAM model using the `mgcv` package in `R`. This gives us an empirical fit $\widehat{\mu}(\cdot)$ of the conditional mean function $\mu(\cdot)$. In order to forecast the conditional mean at time t, given the vector of covariates x_t, we calculate

$$\widehat{y}_t = \widehat{\mu}(x_t). \tag{3}$$

The forecasting accuracy can be evaluated by considering the errors $\widehat{y}_t - y_t$ (or statistics thereof) on a held-out part of the training set.

Next, we focus on the uncertainty term $z_t := \sigma(x_t)\epsilon_t$ in Eq. 1. For now, our only assumption on the random variable ϵ_t is that it has zero mean and unit variance. We will discuss further assumptions (and their empirical justifications) below. For modeling the conditional variance function $\sigma^2(\cdot)$, we use another GAM model,

$$\sigma^2(x_t) = \sum_{j=1}^{J} g_j(x_t).$$

where the transfer functions g_j have the same form as the f_i in Eq. 2. Note that the functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ may depend on different subsets of the covariates x_t. We will see examples in the experimental sections.

In order to fit a function $\widehat{\sigma}^2(\cdot)$ to empirical data, we proceed as follows:

1. We fit a function $\widehat{\mu}(\cdot)$ for the conditional mean demand, as explained above.
2. We calculate empirical residuals $\widehat{z}_t = \widehat{y}_t - y_t$.
3. We fit a GAM model to the squared empirical residuals $\widehat{z}_t^2$, again using the `mgcv` package in `R`.

Given the covariates x_t, we use this model to compute the estimate $\widehat{\sigma}^2(x_t)$ of the conditional variance. Note that a similar two-stage estimation procedure, based on non-parametric regression, was proposed in (Fan and Yao 1998). Our modeling approach takes into account the dependency of uncertainty on exogenous variable. In particular, the variance of electricity demand is typically higher during peak periods such as in the evening or on days with extreme temperatures. In statistical terms, such time-varying variance effects are known as heteroscedasticity. Normalizing $\widehat{z}_t$ by $\widehat{\sigma}(x_t)$ can be thought of a way to scale the empirical residuals such that the rescaled version has zero mean and one standard deviation.

Distributional assumptions So far we have been only assuming zero mean and unit variance of ϵ_t. If we assume normality, the conditional mean and variance uniquely parameterize the entire conditional distribution of y_t given x_t. As we show in the following paragraphs, this allows us to construct one- and two-sided prediction intervals. Without assuming normality, the conditional variance provides some information about the dispersion (and hence some quantification of uncertainty), which can be used, e.g., by applying Chebyshev's inequality to make probabilistic statements about deviations of the actual demand from its mean value. An empirical analysis in the experimental section shows statistical evidence that the normality assumption is indeed justified.

Another comment: in this paper, we only focus on marginal distributions, i.e., characterizing the uncertainty at one particular time point t. In order to make statements about the uncertainty of demand in a time interval $T = \{t_1, ..., t_{|T|}\}$, one would have to make assumptions about the dependency structure of the process $\{\epsilon_t\}$, however, this goes beyond the scope of this work.

Prediction intervals Let $p \in [0, 1]$ and $q(p)$ be the quantile function of the standard normal distribution, i.e., if X has a standard normal distribution, then $Pr(X \leq q(p)) = p$. We construct the *one-sided prediction interval* at the $p \cdot 100$ *percentile* for Y_t as follows:

$$\phi_t^1(x_t, p) = \widehat{\mu}(x_t) + q(p) \cdot \widehat{\sigma}(x_t). \tag{4}$$

If the estimates $\widehat{\mu}(\cdot)$ and $\widehat{\sigma}^2(\cdot)$ are statistically consistent and ϵ_t follows a standard normal distribution, then we have $Pr(Y_t \leq \phi_t^1(x_t, p)) = p$. In the following, whenever the context is clear, we omit x_t and write $\phi_t^1(p)$ instead.

Similarly, the *two-sided prediction interval* at the $p \cdot 100$ *percentile* is constructed as:

$$\phi_t^2(x_t, p) = \widehat{\mu}(x_t) \pm |\, q((1-p)/2)\, | \cdot \widehat{\sigma}(X_t'). \tag{5}$$

Whenever the context is clear, we omit x_t and refer to the right endpoint of the interval as $\phi_t^{2\triangle}(p)$ and to the left endpoint as $\phi_t^{2\triangledown}(p)$, respectively.

3.2 Evaluation metrics

Practitioners, e.g., trading power demand/supply in day-ahead electricity markets, require reliable prediction intervals. A key performance indicator is: *Does the prediction interval for the 95 percentile, in the long run, cover the demand indeed 95% of the time?*[3] Additionally, smaller interval sizes are preferred over too conservative uncertainty estimates, since they generally lead to more efficient operational decisions. The pinball loss function is widely used in the literature to assess the accuracy of probabilistic/quantile forecast (Koenker and Bassett 1978; Takeuchi et al. 2006). However, it does not express how reliable or how wide the prediction intervals are, and thus provides only limited insights into the quality of uncertainty forecasts. To this end, we propose evaluation metrics that measure explicitly the quantities of interest above, i.e., the coverage accuracy and the width of the intervals.

Empirical coverage We define the empirical one-sided coverage of our estimated intervals for the $p \cdot 100$ percentile during a time period $T = \{t_1, \ldots, t_{|T|}\}$ as:

$$c_1(p, T) = \frac{1}{|T|} \sum_{t \in T} \mathbf{1}_{(y_t \leq \phi_t^1(p))}. \tag{6}$$

This can be seen as an estimate of $Pr(Y_t \leq \phi_t^1(p))$ over a time period T. The closer the coverage is to p, the better. Similarly, the empirical two-sided coverage is defined as:

$$c_2(p, T) = \frac{1}{|T|} \sum_{t \in T} \mathbf{1}_{(\phi_t^{2\triangledown}(p) \leq y_t \leq \phi_t^{2\triangle}(p))}. \tag{7}$$

Similar to the one-sided case, we can see this is an estimate of $Pr(\phi_t^{2\triangledown}(p) \leq Y_t \leq \phi_t^{2\triangle}(p))$ over a time period T, and the closer this coverage is to p, the better.

Coverage absolute error (CAE) Given the estimates of one- and two-sided coverages for the $p \cdot 100$ percentile during a time period T, we define the CAE of the one-sided interval as $|p - c_1(p, T)|$ and the CAE of the two-sided interval as $|p - c_2(p, T)|$.

Mean percentage width (MPW) We define the MPW of the one-sided interval at the $p \cdot 100$ percentile as: $\frac{1}{|T|} \sum_{t \in T} |\phi_t^1(x_t, p) - \mu(x_t)|$, and the MPW of the two sided interval as: $\frac{1}{2 \cdot |T|} \sum_{t \in T} \left(\phi_t^{2\triangle}(p) - \phi_t^{2\triangledown}(p)\right)$. Note that, we

[3] See also (Wytock and Kolter 2013; Weron 2014). Note that, reliability over other percentiles can be expressed similarly, we use 95 percentile only as an example.

divide the latter by two so that it has the same dimension as the one-sided version.

4 Online learning

4.1 Adaptive forecasting of electricity demand

In order to track trends in the mean demand and keep the forecasting models up to date, we deploy the online learning algorithm for Additive Models introduced in (Ba et al. 2012). The basic protocol we have in mind is the following: every day at midnight (one could also choose a different time), we provide demand forecasts with uncertainty for the next 24 hours. After the 24 hours have elapsed, we use the observed actual demand values to update our model.

The online learning algorithm by (Ba et al. 2012) works as follows: write the empirical function for the conditional mean as linear combination of model weights and basis functions:

$$\widehat{\mu}_t(x_t) \quad = \quad \widehat{\boldsymbol{\beta}}_t^T \boldsymbol{b}(x_t)$$

Note that here the model weights $\widehat{\boldsymbol{\beta}}_t$ have a time index t. Given the vector of covariates x_t and the actual demand value y_t, the model weights are updated using the formula

$$\widehat{\boldsymbol{\beta}}_{t+1} \quad = \quad \widehat{\boldsymbol{\beta}}_t + \boldsymbol{g}_t(y_t - \widehat{\mu}_t(x_t))$$

where $\boldsymbol{g}_t$ denotes the *Kalman gain*

$$\boldsymbol{g}_t \quad = \quad \frac{\boldsymbol{P}_t \boldsymbol{b}_t}{\omega + \boldsymbol{b}_t^T \boldsymbol{P}_t \boldsymbol{b}_t},$$

$\omega \in (0, 1]$ is a parameter of the algorithm called the *forgetting factor*, and $\boldsymbol{P}_t$ is the *precision matrix* which is updated using the iterative formula

$$\boldsymbol{P}_{t+1} \quad = \quad \omega^{-1}\big[\boldsymbol{P}_t - \boldsymbol{g}_t \boldsymbol{b}_t^T \boldsymbol{P}_t\big].$$

The vector $\widehat{\boldsymbol{\beta}}_0$ and the matrix $\boldsymbol{P}_0$ can be either set to default values, or calibrated on a batch of historical data by fitting a GAM and taking the model parameters and sample precision matrix as initial values.

The forgetting factor ω allows for discounting of past observations: If $\omega = 1$, then all observations are weighted equally, but the smaller ω, the more weight will be given to recent observations. In our experiments, we choose ω using the rule of thumb

$$\text{number of observations per year} \quad \approx \quad \frac{1}{1-\omega},$$

e.g., if we have hourly data and hence $365 \cdot 24$ observations per year, we choose $\omega = 0.9998858$.

4.2 Adaptive construction of prediction intervals

Next, we outline our approach to adapt the prediction intervals to the non-stationary nature of the electricity demand. That is, when the prediction interval achieves the desired coverage, then the adaptive construction algorithm practically has no effect. On the other hand, when the prediction interval starts to cover lower or higher percentile than expected, the algorithm adjust the interval by gradually widening or narrowing it.

Let us denote the empirical coverage as described in Eq. 6 and 7 as c. Then, the coverage $\widehat{c}$ that we should target for the next time periods (or construction horizon) h to cover the desired p portion of the data (or the $p \cdot 100$ percentile) is given by:

$$\begin{aligned} \alpha\, c + (1-\alpha)\widehat{c} \quad &= \quad p \\ \widehat{c} \quad &= \quad \frac{p - \alpha\, c}{1-\alpha}, \end{aligned} \tag{8}$$

where we call $\alpha \in [0, 1)$ the *aggressivity* parameter. The higher the aggressivity parameter α, the more eager we are in adjusting $\widehat{c}$ to cover the $p \cdot 100$ percentile. The construction horizon h should be equal to or greater than the forecasting horizon. Additionally, since Eq. 8 might yield values smaller than 0 or greater than 1, we set the minimum value of $\widehat{c}$ to ε, and the maximum value to $1 - \varepsilon$, where ε is a positive number close to zero.[4]

Let q be the quantile function of the standard normal distribution. Then, for the next time periods h, we construct the (adaptive) intervals as follows.

Adaptive one-sided interval We define the adaptive one-sided interval as

$$\phi_t^1(x_t, p) = \widehat{\mu}(x_t) + q(\widehat{c}) \cdot \widehat{\sigma}(x_t), \tag{9}$$

i.e., we modify the one-sided interval construction in Eq. 4 by using $q(\widehat{c})$ instead of $q(p)$.

Adaptive two-sided interval Similarly, we define the adaptive two-sided interval to cover p portion of the data as

$$\phi_t^2(x_t, p) = \widehat{\mu}(X_t) \pm |\, q((1-\widehat{c})/2)\,| \cdot \widehat{\sigma}(x_t), \tag{10}$$

i.e., we modify the two-sided interval construction in Eq. 5 by using $q((1-\widehat{c})/2)$ instead of $q((1-p)/2)$.

5 Experimental results

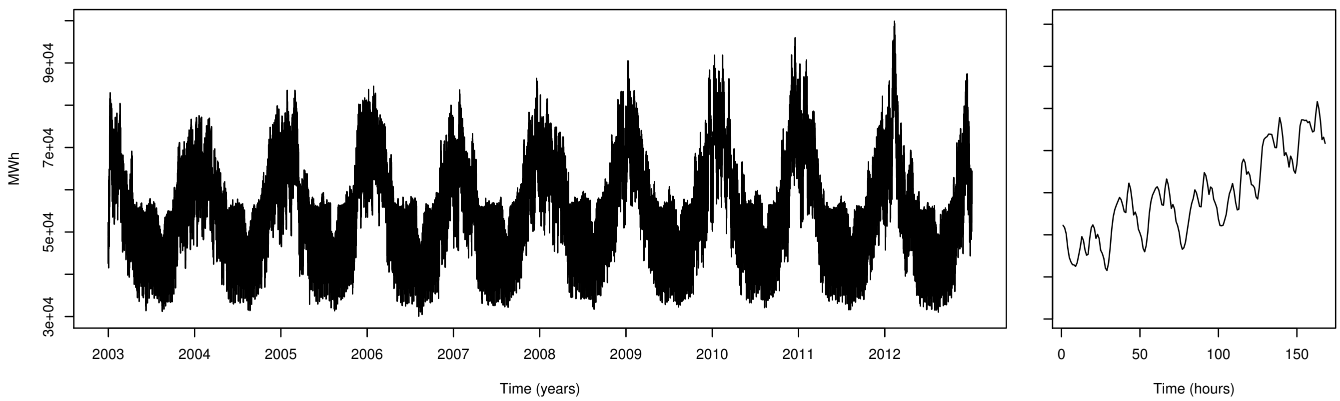
We focus on day-ahead forecasts, where the demand forecast and its prediction interval for each hour for the next 24 hours are delivered once per day at a specific time of day (e.g., midnight).

5.1 Dataset

We use the publicly available dataset provided by the Réseau de transport d'électricité (or simply RTE dataset)[5] which contains the national electricity demand of France from January 2003 to December 2012. While the original data is recorded every 30 minutes, we aggregated the measurement into hourly measurements. The resulting time series is shown in Figure 1: in (a) one can see clear seasonal cycles over the 10 years, while (b) shows typical daily patterns of electricity demand. Notice the overall increasing trend in demand and the unusually high peak in winter 2011/2012 that

[4] In our experiments, we use $\varepsilon = 0.0001$.

[5] See `http://clients.rte-france.com/lang/an/visiteurs/vie/vie_stats_conso_inst.jsp` (accessed: 2014-02-28). See also our supplementary materials at `https://github.com/tritritri/uncertainty`.



(a) (b)

Figure 1: Hourly electricity demand in France; (a) the complete view of the dataset (from January 2003 to December 2012), and (b) the first week of the dataset.

Table 1: Our out-of-sample, non-overlapping, rolling window test period.

Eval.	Train period	Test period
#1	Jan 2003 - Dec 2010	Jan 2011 - Dec 2011
#2	Jan 2004 - Dec 2011	Jan 2012 - Dec 2012

are interesting challenges for evaluating the effectiveness of our online learning mechanisms.

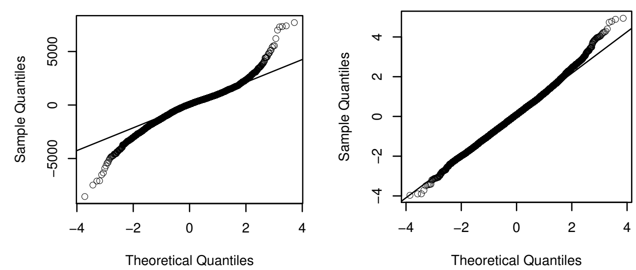
5.2 Forecasting uncertainty

We evaluate the accuracy of our model using out-of-sample, non-overlapping, rolling window test periods. There are two evaluation windows (see Table 1). We use the first eight years as the train period and the subsequent year as the test period.

Modeling the conditional mean In this experiment, we model the conditional mean of the electricity demand using the following GAM:

$$\begin{aligned}\log(\mathbb{E}[Y_t]) &= \beta_0 + \beta_1 \texttt{tmrwDayType}_t + \\ &\textstyle\sum_{j=1}^{8} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{1,j}(\texttt{HourOfDay}_t) + \\ &f_2(\texttt{TimeOfYear}_t) + f_3(\texttt{TempC}_t, \texttt{Humidity}_t) + \\ &f_4(\texttt{TempC}_t, \texttt{HourOfDay}_t) + f_5(\texttt{LagNSameDTLoad}_t) + \\ &\textstyle\sum_{j=1}^{8} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{6,j}(\texttt{LagLoad}_t),\end{aligned} \tag{11}$$

The covariate DayType is a categorical variable with 8 values that represent the seven days of the week and public holidays, tmrwDayType models transitions between different day types, LagLoad is the load for the same hour of the previous day, and LagNSameDTLoad is the average load for the same hour of the previous two days with the same day types (i.e., weekday/weekend). The covariate TempC (and Humidity) are the weighted average of the temperature (and the humidity) of the 6 biggest cities in France, proportional to their annual energy consumption. This model achieves a Mean Absolute Percentage Error (MAPE) of 1.65 (averaged over the two test periods).



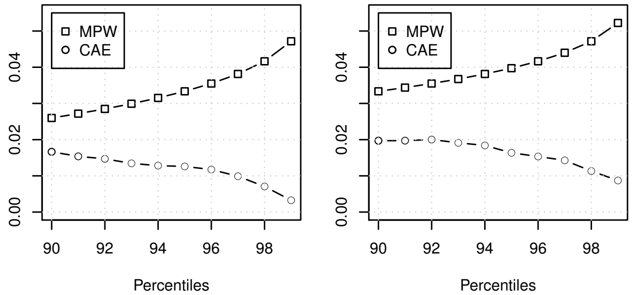
(a) (b)

Figure 2: Normal Q-Q plot of (a) the empirical residuals $\widehat{z}_t$, (b) the rescaled residuals $\widehat{\epsilon}_t$.

Modeling the conditional variance As described in Section 3.1, we obtain the estimated conditional variance function $\widehat{\sigma}^2(\cdot)$ by fitting a second GAM to the squared empirical residuals $\widehat{z}_t^2$. We model it as follows:

$$\begin{aligned}\log(\mathbb{E}[\widehat{z}_t^2]) &= \beta_0 + \beta_1 \texttt{DayType}_t + \\ &\textstyle\sum_{j=1}^{5} \mathbf{1}_{(\texttt{DayType}_t = j)} f_{1,j}(\texttt{HourOfDay}_t) + \\ &f_2(\texttt{TimeOfYear}_t) + f_3(\texttt{TempC}_t, \texttt{Humidity}_t) + \\ &f_4(\texttt{TempC}_t, \texttt{HourOfDay}_t).\end{aligned} \tag{12}$$

To justify our normality assumption and evaluate the effectiveness of our model above, we show the normal Q-Q plot of the empirical residuals $\widehat{z}_t$ (Figure 2a) and the rescaled empirical residuals $\widehat{\epsilon}_t$ (Figure 2b). The rescaled residuals are indeed much more normal than the original residuals $\widehat{z}_t$. Next, to evaluate the accuracy and the width of the prediction intervals we compute their CAE and MPW. Figure 3 shows that the coverage error (CAE) is around 1%-2% (approximately at the same level as the forecasting error for the conditional mean) with lower error and higher interval width (MPW) in the tail. Moreover, note that our MPW are fairly small (mostly around 3%-4%). For instance, if the MPW at the 95 percentile is around 4% and the average demand is around 55000 MWh, then on average, the two-sided prediction interval for the 95 percentile is around $[52800, 57200]$



(a) one-sided intervals (b) two-sided intervals

Figure 3: The mean percentage width (MPW) and the coverage absolute error (CAE) of the estimated (a) one- and (b) two sided prediction intervals, averaged over the two test periods.

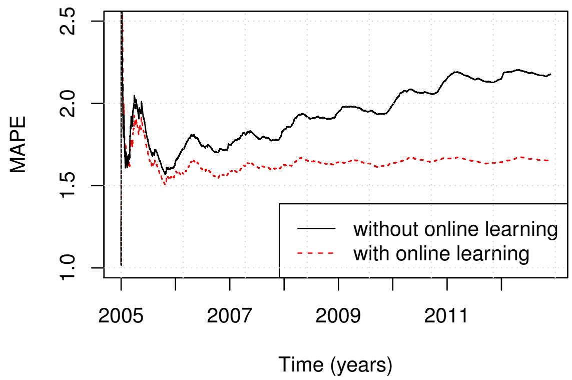


Figure 4: The forecasting error (MAPE) over time for the two models, i.e., with and without online learning. The online learning mechanism succeeds to keep the forecasting error low over time.

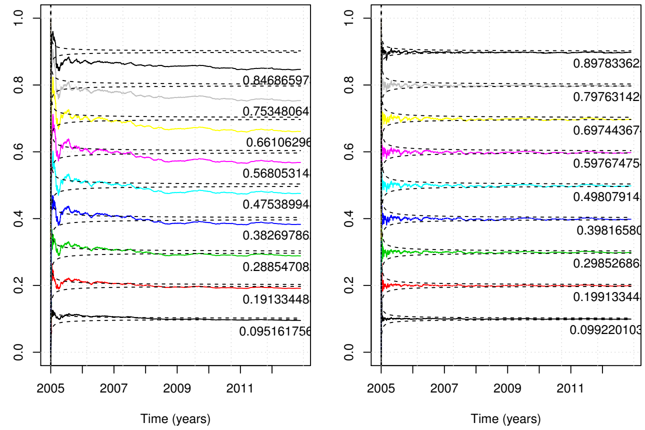
MWh.

5.3 Online learning

Although we have already relatively small prediction errors (1%-2%), this high accuracy might not be sustained over a long period time. In a real-world implementation, a drop in model performance could be caused by trends and dynamic changes in the way people consume electricity, e.g., related to new electricity tariffs, the emergence of new appliances, or changes in macroeconomic conditions. In this section, we show the effectiveness of the online learning mechanism in addressing these challenges and maintaining high forecasting accuracy. For this experiment, we use the first two years of the RTE data set as the train period and the last eight years as the test period.

Figure 4 shows the forecasting error (MAPE) of both models, without and with online learning. Although at the beginning both models have similar performance, the MAPE of the model without online learning continues to increase over time, whereas with online learning we are able to keep the error low. Intuitively, using the online learning mechanism, the model is able to adapt to the dynamic changes of the electricity demand, and thus prevent the MAPE from increasing.

Next, to evaluate the effectiveness of the adaptive construction of the prediction intervals, we compute the cov-



(a) (b)

Figure 5: Empirical coverage (a) without and (b) with adaptive interval construction over time. The solid lines represent the two-sided coverages for the (from top to bottom) 90, 80, ..., 20, and 10 percentiles. The numbers on the right hand side show the empirical coverages at the end of the test period. The dotted lines are the binomial (95%) confidence interval for each percentile, which serve as a guide on approximate ranges of acceptable coverages.

erage of the intervals without and with the adaptive construction for various percentiles, i.e., 90, 80, ..., 20, 10 percentiles. For the adaptive construction, we use the parameters $\alpha = 0.95$ and $h = 24$ hours (equal to the forecasting horizon). Figure 5 shows that the adaptive construction successfully maintains the coverage accuracy for all the percentiles shown, whereas without it, the accuracy tends to decrease over time (with lower accuracy for the higher percentiles).

6 Conclusion and future work

In this paper, we proposed a novel GAM^2 algorithm to forecast uncertainty in electricity demand by modeling the time-varying conditional mean and variance. Our method is efficient since it does not require multiple runs of simulations or bootstrapping. We assessed the coverage accuracy as well as the width of the prediction intervals. Furthermore, we incorporated online learning mechanisms to adapt the forecasted mean and prediction intervals to the dynamic changes in the distribution of the demand. Although we showcase our method specifically for electricity demand, we are confident that it is applicable to other domain as well.

In the future, we plan to investigate spatio-temporal dependencies and correlations between uncertainty of demand and renewable energy sources. Another important direction is to study the trade-off between the coverage accuracy and the interval width emerging from the adaptive construction algorithm.

Acknowledgment

The authors would like to thank Yannig Goude and Alhussein Fawzi for their helpful insights and discussions. Tri Kurniawan Wijaya is supported in part by an IBM PhD Fellowship award.

References

Alzate, C., and Sinn, M. 2013. Improved Electricity Load Forecasting via Kernel Spectral Clustering of Smart Meters. In *The 13th IEEE International Conference on Data Mining (ICDM)*, 943–948.

Ba, A.; Sinn, M.; Goude, Y.; and Pompey, P. 2012. Adaptive Learning of Smoothing Functions: Application to Electricity Load Forecasting. In Bartlett, P. L.; Pereira, F. C. N.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *NIPS*, 2519–2527.

Bacher, P.; Madsen, H.; and Nielsen, H. A. 2009. Online short-term solar power forecasting. *Solar Energy* 83(10):1772 – 1783.

Chen, B.-J.; Chang, M.-W.; and Lin, C.-J. 2004. Load Forecasting using Support Vector Machines: A Study on EUNITE Competition 2001. *IEEE Transactions on Power Systems* 19(4):1821–1830.

Cho, H.; Goude, Y.; Brossat, X.; and Yao, Q. 2013. Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association* 108(501):7–21.

Fan, S., and Hyndman, R. 2012. Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. *IEEE Transactions on Power Systems* 27(1):134–141.

Fan, J., and Yao, Q. 1998. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85(3):645 – 660.

Gneiting, T., and Katzfuss, M. 2014. Probabilistic Forecasting. *Annual Review of Statistics and Its Application* 1(1):125–151.

Groen, J. J. J.; Paap, R.; and Ravazzolo, F. 2013. Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics* 31(1):29–44.

Hastie, T.; Tibshirani, R.; Friedman, J.; Hastie, T.; Friedman, J.; and Tibshirani, R. 2009. *The Elements of Statistical Learning*, volume 2. Springer.

Hippert, H. S.; Pedreira, C. E.; and Souza, R. C. 2001. Neural Networks for Short-Term Load Forecasting: A Review and Evaluation. *IEEE Transactions on Power Systems* 16(1):44–55.

Hong, T.; Wilson, J.; and Xie, J. 2014. Long-Term Probabilistic Load Forecasting and Normalization With Hourly Information. *IEEE Transactions on Smart Grid* 5(1):456–462.

Hong, T. 2010. *Short-Term Electric Load Forecasting*. Ph.D. Dissertation, North Carolina State University.

Huang, S.-J., and Shih, K.-R. 2003. Short-Term Load Forecasting via ARMA Model Identification Including Non-Gaussian Process Considerations. *IEEE Transactions on Power Systems* 18(2):673–679.

Humeau, S.; Wijaya, T. K.; Vasirani, M.; and Aberer, K. 2013. Electricity Load Forecasting for Residential Customers: Exploiting Aggregation and Correlation between Households. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, 1–6.

Hyndman, R., and Fan, S. 2010. Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems* 25(2):1142–1153.

Jarrah Nezhad, A.; Wijaya, T. K.; Vasirani, M.; and Aberer, K. 2014. SmartD: Smart Meter Data Analytics Dashboard. In *The 5th ACM International Conference on Future Energy Systems (e-Energy'14)*.

Jones, H. E., and Spiegelhalter, D. J. 2012. Improved Probabilistic Prediction of Healthcare Performance Indicators using Bidirectional Smoothing Models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175(3):729–747.

Khotanzad, A.; Afkhami-Rohani, R.; Lu, T.-L.; Abaye, A.; Davis, M.; and Maratukulam, D. 1997. ANNSTLF-A Neural-Network-based Electric Load Forecasting System. *IEEE Transactions on Neural Networks* 8(4):835–846.

Koenker, R., and Bassett, Gilbert, J. 1978. Regression quantiles. *Econometrica* 46(1):pp. 33–50.

Kolter, J. Z., and Ferreira, J. 2011. A large-scale study on predicting and contextualizing building energy usage. In Burgard, W., and Roth, D., eds., *AAAI*. AAAI Press.

Misiti, M.; Misiti, Y.; Oppenheim, G.; and Poggi, J.-M. 2010. Optimized clusters for disaggregated electricity load forecasting. *REVSTAT* 8:105–124.

Mohamed, Z., and Bodger, P. 2005. Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy* 30(10):1833 – 1843.

Montgomery, J. M.; Hollenbach, F. M.; and Ward, M. D. 2012. Ensemble Predictions of the 2012 US Presidential Election. *PS: Political Science & Politics* 45:651–654.

Palmer, T.; Shutts, G.; Hagedorn, R.; Doblas-Reyes, F.; Jung, T.; and Leutbecher, M. 2005. Representing Model Uncertainty in Weather and Climate Prediction. *Annual Review of Earth and Planetary Sciences* 33(1):163–193.

Papalexopoulos, A., and Hesterberg, T. 1990. A Regression-based Approach to Short-Term System Load Forecasting. *IEEE Transactions on Power Systems* 5(4):1535–1547.

Ramchurn, S. D.; Vytelingum, P.; Rogers, A.; and Jennings, N. R. 2012. Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence. *Commun. ACM* 55(4):86–97.

Sapankevych, N., and Sankar, R. 2009. Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine* 4(2):24 –38.

Taieb, S. B., and Hyndman, R. J. 2014. A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting* 30(2):382 – 394.

Takeuchi, I.; Le, Q. V.; Sears, T. D.; and Smola, A. J. 2006. Nonparametric quantile estimation. *The Journal of Machine Learning Research* 7:1231–1264.

Taylor, J. W. 2010. Triple Seasonal Methods for Short-Term Electricity Demand Forecasting. *European Journal of Operational Research* 204(1):139–152.

Warner, T. T. 2011. *Numerical Weather and Climate Prediction*, volume 89. Cambridge University Press Cambridge, UK.

Weron, R. 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. HSC Research Reports HSC/14/07, Hugo Steinhaus Center, Wroclaw University of Technology.

Wood, S. 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

Wytock, M., and Kolter, J. Z. 2013. Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields. In *IEEE 52nd Annual Conference on Decision and Control (CDC)*, 1019–1024.