

A refined parametric model for short term load forecasting

Nathaniel Charlton¹, Colin Singleton¹

CountingLab Ltd, Reading, United Kingdom

Abstract

We present a refined parametric model for forecasting electricity demand, that performed particularly well in the recent Global Energy Forecasting Competition (GEFCom 2012). We begin by motivating and presenting a simple parametric model, treating electricity demand as a function of temperature and day of the year. We then set out a series of refinements to the model, explaining the rationale for each, and using competition scores to demonstrate that each successive refinement step increases the accuracy of the model's predictions. These refinements include combining models from multiple weather stations, removing outliers from the historical data and special treatment of public holidays.

1. Introduction

In this paper we present a refined parametric model for short term load forecasting. Our model performed particularly well in the recent Global Energy Forecasting Competition (GEFCom 2012). In Section 2 we motivate and introduce a simple parametric model, which already performs better than the competition's benchmark model. Section 3 sets out a series of refinements to our model, using Weighted Root Mean Squared Error (WRMSE) scores from the competition to confirm that each refinement improves the results. Section 5 identifies possible avenues for further improvement and Section 6 concludes.

¹URL: www.countinglab.co.uk
Email: {billiejoe,colin}@countinglab.co.uk

New feature	Improvement	Score
Competition benchmark model	-	95,588
Our initial model	-	84,362
Multiple weather stations	901	83,461
Day-of-season terms	4,359	79,102
Four seasons instead of two	2,366	76,736
Local averaging	3,090	73,646
Outlier removal	120	73,526
Public holidays treated specially	2,898	70,628
Smoother temperature forecast	3,541	67,087
Our competition entry score	-	67,214

Table 1: WRMSE scores obtained by our model after each refinement step. (These scores are from the *private* leaderboard, obtained after the competition.)

2. Our basic parametric model

The models we build are based on the well-established idea of multiple linear regression. Existing linear regression models for electrical load include those discussed in Moghram & Rahman (1989) and Ramanathan et al. (1997), as well as the competition’s benchmark model. Our initial model supposes that electricity usage is a function

$$E = \alpha_1 + \alpha_2 d + \alpha_3 T + \alpha_4 Td + \alpha_5 T^2 + \alpha_6 T^2 d \quad (1)$$

where T is the temperature, d is the day number (ranging from 0 to 1649 for the 1650 days of historical data) and $\alpha_1, \dots, \alpha_6$ are coefficients to be determined. This form is suggested by multiplying out the expression

$$(a + bT + cT^2)(rd + k) \quad (2)$$

where a, b, c, r, k are constants. The first factor models the quadratic relationship between temperature and energy use that we observe when exploring the data graphically; the second factor allows the model to reflect any changes over time in the response to temperature. Thus our model takes account of the effect of temperature on energy use (due to heating and air conditioning), long term trends in energy use and interactions between the two.

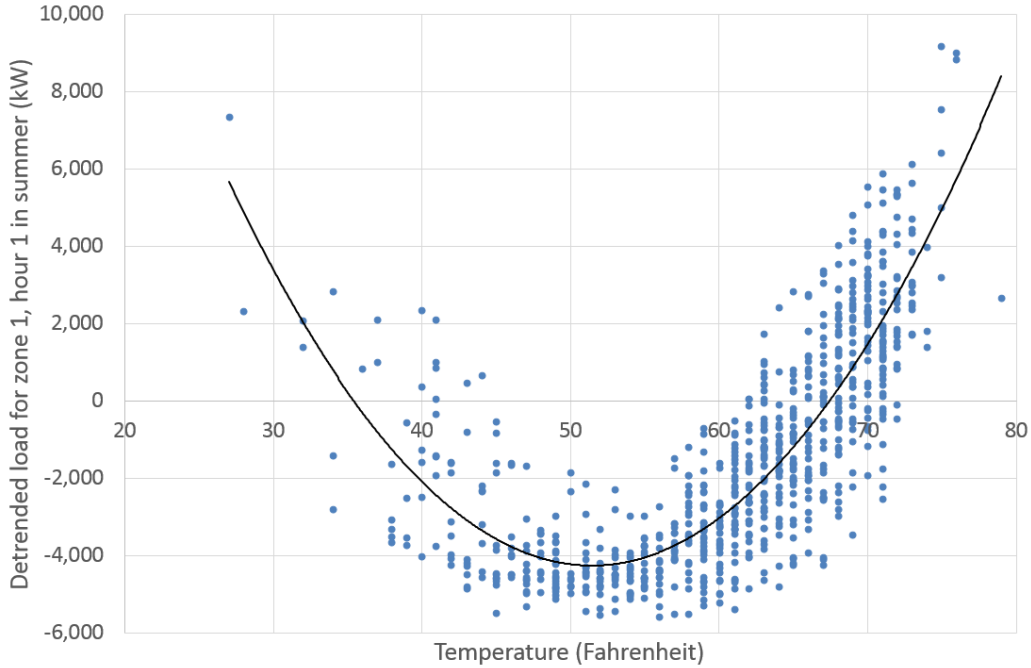


Figure 1: Using a quadratic function $16.642T^2 - 1711.5T + 39751$ to model the relationship between temperature and load for zone 1, hour 1 during the summer, achieving an \bar{R}^2 value (i.e. an adjusted R^2 value) of 0.70779. The load has been detrended by fitting and subtracting a long-term linear model.

Figure 1 shows the relationship between temperature and load for zone 1 at hour 1 during the summer, where we have detrended the load by fitting and subtracting a long-term linear model. Fitting the relationship with a quadratic function, as shown in the figure, achieves an \bar{R}^2 value (i.e. an adjusted R^2 value) of 0.70779. On the other hand, using a cubic function gives a *lower* \bar{R}^2 value of 0.70743, suggesting that a cubic function, despite having one more parameter, cannot really fit the data any better. On the basis of observations such as this we decide not to include T^3 terms in our model (1).

We apply (1) separately for each of the 20 zones and each of the 24 hours of the day. Additionally we divide the year into two seasons (taking summer as April to September inclusive and winter as the rest of the year), and divide the days into two types: weekdays and weekend days. Thus we split the historical load data into $20 \times 24 \times 2 \times 2$ groups and analyse each group of data

separately. This is because we hypothesise that the relationships between energy use, temperature T and day d may be different for each of these $20 \times 24 \times 2 \times 2$ groups. Householders’ reactions to temperature, say, might be different at weekends when they are more likely to be at home. Similarly reaction to temperature may vary with time of day (e.g. at night people are asleep), season (changing use of heating or air conditioning) and zone (due to cultural, demographic or climate-related differences). In-sample testing suggests that this is indeed the case, and that more accurate predictions are made with this model when we split the data into groups than when considering all the data together.

In order to apply (1), however, we need to have a value for the temperature T for each zone. Our initial approach is to find, for each zone, the weather station which “fits the best” with the energy usage of that zone. Consider the group $g = (z, h, s, t)$ for a given zone z , hour of the day h , season s and day type t (weekday or weekend). For each weather station $i = 1, \dots, 11$ we build an energy model

$$E_i = \alpha_1 + \alpha_2 d + \alpha_3 T_i + \alpha_4 T_i d + \alpha_5 T_i^2 + \alpha_6 T_i^2 d \quad (3)$$

by selecting the coefficients $\alpha_1, \dots, \alpha_6$ that minimise the sum of in-sample squared errors for the energy use in group g . We then choose the best weather station, individually for each group g , again minimising the sum of in-sample squared errors. The computation of the coefficients is done using singular value decomposition (SVD) as argued for in (Press et al., 1992, §15.4).

Note that the weather station for each group is chosen independently. This allows, for instance, different weather stations to be used for a particular zone during summer and winter. Knowing nothing about the geography of the zones and weather stations, we find it *a priori* plausible that, if a zone is close to two weather stations, one may be more suitable during summer and the other more suitable during winter, for example because of seasonal wind patterns.

Finally, in order to produce forecasts from our model, we need to predict temperatures for the forecast week, 1st July 2008 to 7th July 2008. Because we are not meteorologists, and wish to concentrate our efforts on understanding load behaviour rather than weather systems, our temperature estimates take a rather simple form. For each weather station $i = 1, \dots, 11$, each day D of the forecast week and each hour $h = 1, \dots, 24$, we estimate a mean temperature $M_{i,D,h}$ using the historical data. Specifically, we look at the

corresponding day of the year in each of the four previous years, and ten days either side of these; we take the mean temperature at hour h over all these days. This makes $M_{i,D,h}$ a mean of $4 \times (10 + 1 + 10) = 84$ data points. We then simply use $M_{i,D,h}$ as our forecast for the temperature at station i at hour h of day D .

As Table 1 shows, our initial model scores 84,362 which is already significantly better than the competition benchmark model. The main differences between our initial model and the benchmark model are as follows. The benchmark model contains T^3 terms in addition to T and T^2 terms, but does not contain Td or T^2d terms. The benchmark model treats each day of the week separately, rather than grouping them as weekdays and weekend days. The benchmark model divides the year into twelve months, rather than into two seasons. Finally as explained above we use different temperature estimates. Note that although *overall* our initial model performs significantly better than the benchmark, we cannot say that *all* of the differences between our initial model and the benchmark model contribute to greater forecasting accuracy.

3. Refinements to our model

We now introduce a series of refinements to the initial model, explaining the rationale for each and noting the improvement in competition score that each provides. Proposed refinements were first tested to see whether they improved the model’s in-sample RMSE. In general, refinements which did so were then used in a competition submission, and kept if they also improved the public score. With this strategy we aimed to make the most of the limited number of submissions available.

3.1. Using combinations of multiple weather stations

Assigning one weather station to each zone is quite simplistic. What if a zone is broadly equidistant from two or more weather stations? What if a weather station is unreliable or faulty? More importantly, when it comes to forecasting electrical load we do not want our forecasts to be wholly dependent on the temperature forecast for a single weather station.

To address these issues, we allow weighted combinations of up to five weather stations to be used for each group $g = (z, h, s, t)$ of data (where z is the zone, h the hour of the day, s the season and t the day type). Our final

model for the group is then a linear combination

$$\beta_1 E_{i_1} + \beta_2 E_{i_2} + \beta_3 E_{i_3} + \beta_4 E_{i_4} + \beta_5 E_{i_5} \quad (4)$$

of the models from the five best fitting weather stations i_1, \dots, i_5 (measured by sum of in-sample squared errors). Again SVD is used to find the coefficients β_1, \dots, β_5 , with the exception that if any station receives a negative coefficient, it is eliminated and the β values recomputed, until all remaining stations have coefficients ≥ 0 . The linear combination (4) is then used to produce the backcasts and forecasts for the group. This modification reduces our score by 901, giving a score of 83,461.

In the context of gas load forecasting, models such as in (4), where there are separate terms for the various weather stations, have been termed *Multiple Weather Station* (MWS) models in Pang (2012). An alternative to MWS is to first average the temperatures from a number of weather stations to make a single “virtual” or “substitute” weather station, and then apply the single-station model as in (1). The virtual weather station approach is used for electrical load forecasting in for example Paravan et al. (2002), Dordonnat et al. (2008) and Pierrot & Goude (2011). Another alternative to MWS is used in Fan & Hyndman (2012). As in Pang (2012), we find that MWS produces better forecasts than the virtual weather station approach.

3.2. Adding day-of-season terms

Examining graphically the fitting errors over the course of the yearly cycle, we find that our model as described so far tends to overestimate around March and April, and underestimate around September and October. A plausible explanation is that householders’ reactions to temperature may change gradually over the course of the yearly cycle. With this in mind, we now add a new variable d' , the number of the day within the season (ranging from 0 to 182). We replace (1) by

$$\begin{aligned} E = & \alpha_1 + \alpha_2 d + \alpha_3 T + \alpha_4 Td + \alpha_5 T^2 + \alpha_6 T^2 d \\ & + \alpha_7 d' + \alpha_8 Td' + \alpha_9 T^2 d' \end{aligned} \quad (5)$$

which is suggested by multiplying out the expression

$$(a + bT + cT^2)(rd + sd' + k) \quad (6)$$

The introduction of these new terms reduces our score by 4,359, giving a score of 79,102.

3.3. Changing the number of seasons

Also motivated by the observation that our fitting errors were greatest at certain times of year, we now split the year into four seasons instead of two: spring (March, April and May), summer (June, July and August), autumn (September, October and November) and winter (December, January and February). This reduces our score by 2,366, giving a score of 76,736.

3.4. Local averaging

To correct for any local systematic over- or under-estimation in our models we perform local averaging. Consider Monday 9AM–10AM for the first day of one of the backcast weeks. Once we have generated an estimate for this hour, using (linear combinations of) (5), we multiply it by the following factor:

$$\frac{A_{-14} + \dots + A_{-1} + A_1 + \dots + A_{14}}{P_{-14} + \dots + P_{-1} + P_1 + \dots + P_{14}} \quad (7)$$

Here P_{-14}, \dots, P_{-1} denote the values predicted by our model for 9AM–10AM on the 14 days prior to the backcast week, and A_{-14}, \dots, A_{-1} denote the actual electricity usage for the same days. Similarly P_1, \dots, P_{14} denote the values predicted by our model for the 14 days after the backcast week, and A_1, \dots, A_{14} are the actual usages for the same days. Thus if our model consistently overestimates by a factor of 1.2 around the time of a particular backcast period, for example, our final backcast will still be correct. Such local systematic errors might be due, for example, to a fraction of the properties in the zone being served by a different zone around the backcast period, or by a factory shutting down or starting up. When forecasting, we have no data for the period following the forecast week, so local averaging is performed using just the 14 days prior to the forecast week. Local averaging reduces our score by 3,090, giving a score of 73,646.

The load history for one of the zones, zone 10, shows an abrupt increase in load at the beginning of 2008, by a factor of about 2.8, which persists until the end of the data. A discussion in the competition forum suggests that this is because the electricity company redefined the zone boundary. We tried adding special adjustments to our model to account for this sudden change but, as long as we used local averaging, no such adjustments produced any improvement. Therefore we believe that local averaging handles the sudden change well.

We experimented with using an additive local correction rather than a multiplicative one, but during in-sample testing this did not perform as well.

3.5. Outlier removal

The historical load data contains some values which are implausibly low, sometimes even zero. We do not know the cause of these, but it is possible that they originate from power outages or measurement failures. For each zone we compute an hourly mean load (across all hours and all days of the history) and then discard any days of data containing hourly loads smaller than 20% of the mean. 158 days of data containing outliers are discarded, 147 of them in a single zone (zone 9). Adding outlier removal reduces our score by 120, giving a score of 73,526.

3.6. Treatment of public holidays

Following what is suggested in Hong (2010), we treat the following special days as weekend days: Memorial Day, Labour Day, Thanksgiving Day and the day after, New Year’s Day, Independence Day and Christmas Day. Other public holidays receive no special treatment. This refinement reduces our score by 2,898, giving a score of 70,628.

3.7. Using a smoother temperature forecast

As explained in Section 2 our initial temperature forecasts were computed by taking time-of-year means, using days within a “window” of 10 days either side of the target day. By changing this window to 25 days either side, the temperature forecasts become smoother and our forecasts get better: we reduce our score by 3,541, giving a score of 67,087.

Experiments with different temperature forecasts confirm as expected that the load forecast, and its score, are very sensitive to the temperature forecast used. Yet at the same time, this aspect of the competition task is somewhat artificial: in reality, utilities wishing to forecast demand can simply purchase commercially available weather forecasts. Of course, such weather forecasts would make use of many variables in addition to past temperatures, such as atmospheric pressure, humidity, cloud cover and so on. Even so, these forecasts are not always right and vary with geography within each zone. One key advantage of our model is that it combines forecasts from a number of weather stations and thus is not reliant on getting every weather station forecast exactly right.

3.8. Our final competition entry

Our competition entry, selected on the basis of the public scores, scored 67,214, marginally higher than our best score reported in Table 1. This is

because our entry contained two further “refinements” which in-sample testing and the public leaderboard scores suggested would improve the model, but in fact did not; this was revealed in additional analysis performed after the competition. The further modifications to the model were an additive adjustment γ to the temperature forecast and multiplicative adjustments for the seven different days of the week. Given more time to spend on the competition, we would ideally have employed further model validation techniques such as k -fold cross-validation. This would have given a more thorough picture of which changes genuinely improved the model. We would also have explored more ideas including some of those listed in Section 5.

4. Computational efficiency

With no effort at optimisation, our C++ program performs the whole forecasting and backcasting in 356 seconds, running on one core of a 2.2Ghz Intel laptop. This figure includes all activities, such as outlier removal, selection of appropriate weather stations for each zone, temperature forecasting and model fitting. We could easily parallelise the operation, fitting the various models concurrently.

5. Ideas for further improvements

In future we will pursue the following ideas, which we hope will lead to substantial improvements in forecast and backcast accuracy.

- Link the models for the different hours of the day, for example by explicitly including term(s) for the previous hour(s) or by simply constraining some of the coefficients in (5) to take the same values across multiple hours of the day. Currently these are fitted completely independently.
- Alternatively, fit different parametric forms to different hours of the day or different zones.
- Study the effect of varying sunrise and sunset times, which determine when domestic and street lighting is used.
- Perform separate fits for each of the backcast or forecast periods, minimising weighted in-sample squared errors. By giving higher weights to days closer to the backcast or forecast period in question, we may obtain a better fit for that period.

- Try out alternative approaches for modelling zone 9, which is industrial load and is determined mainly by factory schedules rather than temperatures.
- Revisit the issue of outlier removal. Adding outlier removal reduced the WRMSE by only 120, which we found surprisingly low. In particular it is possible that some of the low but non-zero values in the atypical zone 9, the industrial zone, are correct measurements that we wrongly excluded.
- Explore whether fixing any of the coefficients α_j in (5) to zero produces better models.
- Investigate whether the data contains any evidence of Demand Side Management, and develop ways to account for this in our model.
- Add economic factors to our model.

6. Conclusions

We presented a refined parametric model for short term load forecasting. We began with a linear model, as in (1), that is simple and transparent, uses well-understood techniques and is easy and efficient to implement. A series of refinements improved the accuracy of our model; these included combining models from multiple weather stations, removing outliers from the historical data and special treatment of public holidays. The C++ code of our forecasting program is freely available under the GPL license and we encourage readers to experiment further with our methods.

References

- Dordonnat, V., Koopman, S., Ooms, M., Dessertaine, A., & Collet, J. (2008). An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, *24*, 566 – 587.
- Fan, S., & Hyndman, R. (2012). Short-term load forecasting based on a semi-parametric additive model. *Power Systems, IEEE Transactions on*, *27*, 134–141.

- Hong, T. (2010). *Short term electric load forecasting*. Ph.D. thesis North Carolina State University.
- Moghram, I., & Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on Power Systems*, *4*, 1484–1491.
- Pang, B. (2012). *The Impact of Additional Weather Inputs on Gas Load Forecasting*. Master's thesis GasDay Laboratory, Marquette University Wisconsin, USA.
- Paravan, D., Debs, A., Hansen, C., Becker, D., Hirsch, P., & Golob, R. (2002). Influence of temperature on short-term load forecasting using the EPRI-ANNSTLF. Presented at Balkan Power Conference, Belgrade, 2002.
- Pierrot, A., & Goude, Y. (2011). Short-term electricity load forecasting with generalized additive models. In *ISAP power* (pp. 593–600).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C (2nd ed.): The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press.
- Ramanathan, R., Engle, R. F., Granger, C. W. J., Vahid-Araghi, F., & Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, *13*, 161–174.