International Journal of Forecasting 30 (2014) 364–368

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

A refined parametric model for short term load forecasting

Nathaniel Charlton [1], Colin Singleton *,[1]

CountingLab Ltd, Reading, United Kingdom

ARTICLE INFO

Keywords:
Electricity
Regression
Forecasting competitions
Combining forecasts
Demand forecasting

ABSTRACT

We present a refined parametric model for forecasting electricity demand which performed particularly well in the recent Global Energy Forecasting Competition (GEFCom 2012). We begin by motivating and presenting a simple parametric model, treating the electricity demand as a function of the temperature and day of the data. We then set out a series of refinements of the model, explaining the rationale for each, and using the 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 treatments 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 then sets out a series of refinements of our model, using Weighted Root Mean Squared Error (WRMSE) scores from the competition to confirm that each refinement does actually improve the results. Section 5 identifies various possible avenues for further improvement, and Section 6 concludes.

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 loads include those discussed by Moghram and Rahman (1989) and Ramanathan, Engle, Granger, Vahid-Araghi, and Brace (1997), as well as the competition's benchmark model. Our initial model supposes that electricity usage is a function

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

where T is the temperature, d is the day number (ranging from 0 to 1649 for the 1650 days of historical data), and $\alpha_1, \ldots, \alpha_6$ 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 and 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 into account the effect of temperature on energy use (due to heating and air conditioning), long term trends in energy use, and interactions between these two.

Fig. 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

* Corresponding author.
E-mail addresses: billiejoe@countinglab.co.uk (N. Charlton), colin@countinglab.co.uk (C. Singleton).

[1] URL: www.countinglab.co.uk.

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.07.003

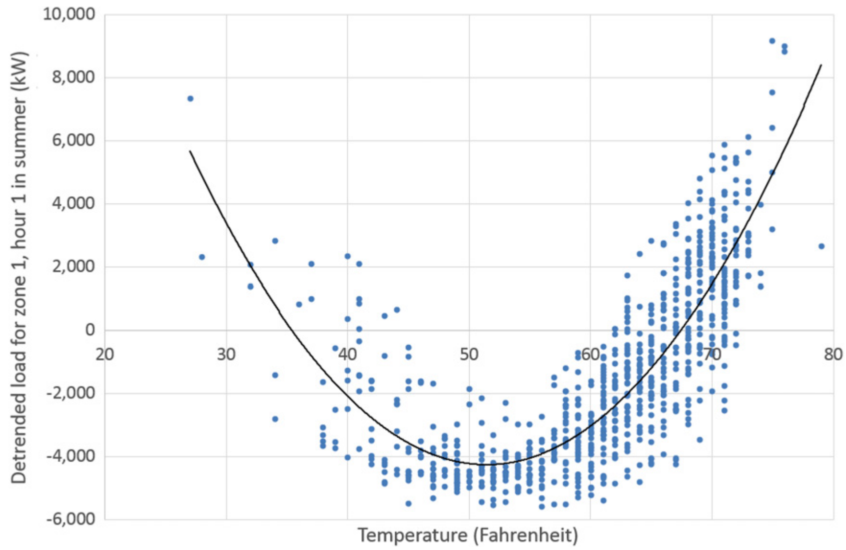

Fig. 1. Using a quadratic function $16.642T^2 - 1711.5T + 39\,751$ to model the relationship between temperature and load for zone 1, hour 1 during the summer, achieving an $\bar{R}^2$ value (i.e., an adjusted R^2 value) of 0.70779. The load has been detrended by fitting and subtracting a long-term linear model.

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 Eq. (1) separately for each of the 20 zones and each of the 24 h of the day. In addition, we also 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. For example, householders' reactions to temperatures might be different on weekends, when they are more likely to be at home. Similarly, the reaction to temperature may vary with the time of day (e.g., at night people are asleep), season (changing use of heating or air conditioning), and zone (due to cultural, demographic and climate-related differences). In-sample testing suggests that this is indeed the case, and that this model produces more accurate predictions when we split the data into groups than when considering all of the data together.

In order to apply Eq. (1), however, we need to have a value for the temperature T for each zone. Initially, we determine, 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 $i = 1, \ldots, 11$, we build an energy model

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

by selecting the coefficients $\alpha_1, \ldots, \alpha_6$ 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 was argued for by Press, Teukolsky, Vetterling, and Flannery (1992, Section 15.4).

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

New feature	Improvement	Score
Competition benchmark model	–	95,588
Our initial model	–	84,362
Multiple weather stations	901	83,461
Day-of-season terms	4359	79,102
Four seasons instead of two	2366	76,736
Local averaging	3090	73,646
Outlier removal	120	73,526
Public holidays treated specially	2898	70,628
Smoother temperature forecast	3541	67,087
Our competition entry score	–	67,214

Note that the weather station for each group is chosen independently. This allows, for instance, two different weather stations to be used for a particular zone in summer and winter. Knowing nothing about the geography of the zones and weather stations, we find it plausible *a priori* that, if a zone is close to two weather stations, one may be more suitable during summer and the other 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 the behaviour of loads rather than of weather systems, our temperature estimates take a rather simple form. For each weather station $i = 1, \ldots, 11$, each day D of the forecast week, and each hour $h = 1, \ldots, 24$, 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 each of these days; we then take the mean temperature at hour h over all of these days. This makes $M_{i,D,h}$ the mean of $4 \times (10 + 1 + 10) = 84$ data points. We then simply use