

Chapter 14

Case Study: Low Voltage Demand Forecasts



This chapter demonstrates the practical implementation of short term (day-ahead) forecasts for the application of residential low voltage networks. It is split into two main parts: An in-depth examination of a short term forecasting case study of residential low voltage networks (Sect. 14.2); and a example python code demonstrating how to implement some of the methods and techniques in practice (Sect. 14.3).

The case studies serve to demonstrate how to:

- identify the main challenges when implementing short term forecasts.
- use the techniques from Chap. 6 to analyse the data, and identify important features.
- use the analysis to choose several forecast models (from those presented in Chaps. 9 and 11). This includes both point and probabilistic models.
- test, compare and evaluate the forecasts.

The chapter begins by a short discussion of how to design a forecast trial which will frame the case study that follows later.

14.1 Designing Forecast Trials

It is worth reiterating some of the core elements which should be considered prior to, and while, developing the forecasts. These elements are important to ensure that the model is designed appropriately with minimal bias in methodology, and to ensure that the results are properly tested. The full forecasting procedure is outlined in Chap. 12 and will be followed implicitly throughout. This chapter will focus on the following main considerations.

1. **Initial Experimental Design:** Before plotting the data it is worth sketching out an initial experimental design and audit the available data used to produce and test the forecasts. What type of data is being considered? Is it expected to have seasonalities? What is the resolution of the data, half hourly, every ten minutes?

Is there sufficient data to produce a informative result? If there is, what is an appropriate split of the data into training, validation and testing sets (see Sect. 8.2)? It is important to think about these questions before analysing the data to prevent introducing bias into the test. Further, once the data has been split, it is advisable to avoid analysing the test set prior to generating forecasts to avoid ‘cheating’ and seeing the true values before submitting the forecast. A final consideration is to decide on what error measure to use (see Chap. 7). An incorrectly chosen error metric can skew the results, and makes it difficult to evaluate and interpret the results.

2. **Visualisation and Data Analysis:** It is essential to try and learn as much as possible about the underlying features and relationships in the data. In Chap. 5 a number of tools were presented showing how to achieve a better understanding of the data. Simple time series plots can highlight large scale behaviours, scatter plots can identify strong relationships between variables, and autocorrelation plots can highlight periodicities and autoregressive behaviours in the data.
3. **Pre-processing:** A necessary component to the data analysis is data cleansing and pre-processing. Poor quality data can make for misleading analysis and meaningless results. To use a common phrase in machine learning: ‘garbage-in garbage out’. Before applying any models, check for anomalous data and missing values as shown in Sect. 6.1.2, and then either replace or remove them from the dataset. The analysis in the previous step can be used to choose the appropriate replacement values.
4. **Model Selection and Training:** As shown in Chaps. 9 and 10 there are a wide range of possible forecast methodologies and choosing the correct models requires utilising the learning from the data analysis, considering the specific requirements for the application, as well as learning from the forecasters own experience. The validation set can be an essential tool for narrowing down the choice of models. It is also vital that appropriate benchmarks (see Sect. 9.1) are selected to help assess the accuracy of the core models. Section 12.2 presents Further criteria which can be considered to help select the initial methods.
5. **Testing and Evaluation:** The trained models must be applied to the unseen testing set. By scoring and comparing the forecast methods with the error measures (See Chap. 7) a better understanding can be forged about what makes some methods more accurate and what are the important (or unimportant) features. This step will allow the forecaster to develop further improvements in future iterations of the models.

Each of the above steps will be illustrated in the following case study.

14.2 Residential Low Voltage Networks

This application considers short term (in this example up to four days ahead) load forecasting for residential low voltage demand on substation feeders and will be used to demonstrate probabilistic (Chap. 11) as well as point forecast methods (Chaps. 9 and 10). The entire section will be based upon the authors research presented in [1].

Here, the term residential low voltage (LV) network demand (or just residential networks) is used to describe the network connected to the secondary substations of the electricity distribution network within a residential area (Sect. 2.1). Although the connected customers will usually be residential they may also consist of small commercial customers such as offices, hairdressers, etc. The demand time series represents the aggregated demand of consumers fed electricity directly from the substation (ignoring any electrical losses in the cables of course). This typically consists of around 40–50 consumers. Furthermore, since these consumers are typically residential, human behaviour tends to be a strong determinant of the demand patterns and hence daily and weekly periodicities are expected. The data considered here will be for 100 residential feeders in the area of Bracknell, a medium sized town in the southeast of England.

At the low voltage, demand is much more volatile than higher voltage due to the low aggregation of consumers (Sect. 2.3) which means probabilistic forecasts can be quite useful for quantifying the uncertainty in the demand.

14.2.1 Initial Experimental Design

The data consists of half-hourly load data for 100 residential low voltage feeders beginning on 20th March 2014 up to the 22nd November 2015 inclusive, a total of 612 days. Typically there are 4–6 feeders which come from a residential low voltage substation and on average there are around 45 consumers per feeder with the largest having 109 residential consumers. A further seven had no available connectivity information due to missing information in the database, so it is not known who is connected. The feeders typically feed residential consumers and 83 of the 100 are purely residential, the others are typically mixes except for one which is known to feed only the landlord lighting of a large office block. The average daily demand across the feeders is approximately 602 kWh and a maximum and minimum daily demand of around 1871 kWh and 107 kWh respectively.

The first decision to be made is how to split the data into testing and training data sets. The data set is reasonably sized, although more data would of course be preferable, especially for residential feeder demand which is expected to have annual seasonality. Ideally, to accurately model annual seasonality several years of data would be available so that the typical year-to-year behaviour could be captured. However, for the purposes of short term forecasts the length of data is sufficient. The final two months were kept over as the out of sample testing set. This consists of 53

days from 1st October 2015 to 22nd November 2015 inclusive. Notice this is just under 9% of the data and is less than the common split of training and testing into a 4:1 ratio (i.e. 20% testing data) as discussed in Sect. 8.1.3. There is a number of reasons for this, firstly this increases the amount of training data whilst retaining a reasonable sized testing set and secondly it ensures that the forecasts are made for some of the colder months of the year, which are typically higher in demand and of particular interest to network operators who are concerned about excessive peak demand.

It should be kept in mind, that since the training is 559 days and only 1.5 years long there may be some limitations in capturing annual seasonalities and therefore the methods here cannot be reliably extended to medium term (one month to a year ahead) or longer term (over 1 year ahead) forecasts.

Hourly temperature forecast data and observed temperature data are also available for the same time period. The forecasts all begin at 7 AM each day and then produce hourly forecasts up to a horizon of 4 days ahead (96 h ahead). This means temperature effects can also be studied but since the forecast origin (where the forecast starts from) is limited to 7 AM each day, they must be treated with caution when using them as inputs to the forecast models. In particular, it would be expected that forecasts become slightly less accurate the further ahead they forecast which means that, e.g. the four hour ahead temperature forecasts (i.e. the ones at 11 AM) will be more accurate than the forecasts five or more hours ahead (i.e. those from noon onwards).

With these datasets several situations can be tested

1. **Case 1:** How does the accuracy of a forecast model change with horizon from 1 h ahead to 96 h (four days) ahead?
2. **Case 2:** Are all residential LV feeders forecast with similar accuracy? If not what are some of the distinguishing factors between them?
3. **Case 3:** What is the effect of including temperature within a forecast model for residential LV network demand?

To allow comparison between models with and without temperature forecast inputs, all forecasts will generate hourly four day ahead forecasts starting at the forecast origin of 7AM of each day of the testing set. This requires aggregating the half-hourly demand time series up to the hourly resolution (see Sect. 6.1.4) to facilitate using temperature data as an input to the models.

To allow comparisons between the different forecasts, some forecast error measures need to be chosen as presented in Chap. 7. To allow comparison between different size feeders, relative measures which don't depend on the size of the feeder (i.e. typically demand magnitude) are required. MAPE is a common relative measure used for demand forecast but since it can be conflated by small values, a modified version of the MAE (see Eq. (7.46) in Chap. 7) is also used which takes the usual MAE but is scaled by the average hourly load of each feeder over the final year of the training data. This will be referred to as the Relative MAE or RMAE. Since the experiment will also include probabilistic forecasts, probabilistic scoring functions will also be required. In this experiment the continuous ranked probability score

(CRPS) is used (see Eq. (7.50) in Chap. 7). The CRPS error for each feeder is also divided by the average hourly demand for that feeder to produce the Relative CRPS or RCRPS.

14.2.2 Data Analysis

As defined here, residential LV feeders are predominantly connected to residential households but may also connect to a small number of shops, offices, churches, schools and other small-to-medium enterprises. For these reasons it would be expected that demand patterns are largely driven by human behaviour and hence contain strong daily, weekly and annual seasonalities. An example of the demand for a few of the feeders is shown in Fig. 14.1 for different numbers of consumers connected (labelled with the variable *NumMpan*s) and different numbers of residential consumers (labelled *NumRes*). The time series plots identify several features. Firstly that there is a wide variety of behaviours, even between the two purely residential feeders (Labelled Feeder 4 and 15 in the plot) with similar numbers of connected consumers (44 and 42). Although they both exhibit annual seasonality with larger demands over the Winter period, there is large periods of low demand during the Christmas and Easter holidays for Feeder 15 but not for Feeder 4. Although this won't be considered in this work, it does suggest holiday periods should be treated as special inputs to the model and this could be an important extension to the more general models presented here (see Sect. 13.6.2). Another important distinction is between the largely residential Feeders (4, 10 and 15) and Feeder 23 which is connected to a single commercial consumer. The commercial consumer doesn't have strong annual seasonalities, and in contrast to the purely residential feeders, has relatively low demand during the Winter period. These time series have identified two important properties of these time series, firstly annual seasonality is an important feature to include in the models, and secondly there is a wide difference between different feeders which suggests there may not be a one-size-fits-all model which will be accurate for all feeders.

The time series plots have identified annual seasonality as an important feature to include in the forecast models. Other inter-annual seasonalities can be identified by considering the autocorrelation function (ACF) plots (see Sect. 6.2.2). In fact, in each of the ACF plots there are relatively large spikes at lags of a day (a lag of 24 h) and a week (lags of 196) and multiples of these. Of these, the weekly periodicities are the strongest autocorrelations as expected. Since there are 100 feeders it is difficult to consider all of their respective autocorrelation plots, instead particularly important lags can be given special focus. The weekly seasonality information is considered in Fig. 14.2, by showing the autocorrelations at lag 196 as a function of the size of each feeder (average daily demand in kilowatthours (kWh)). This shows that the strongest weekly autocorrelations are associated to the largest feeders. One explanation for this is that feeders with larger demands consist of aggregations of larger numbers of residential consumers, increasing the prominent regularities in weekly behaviour.

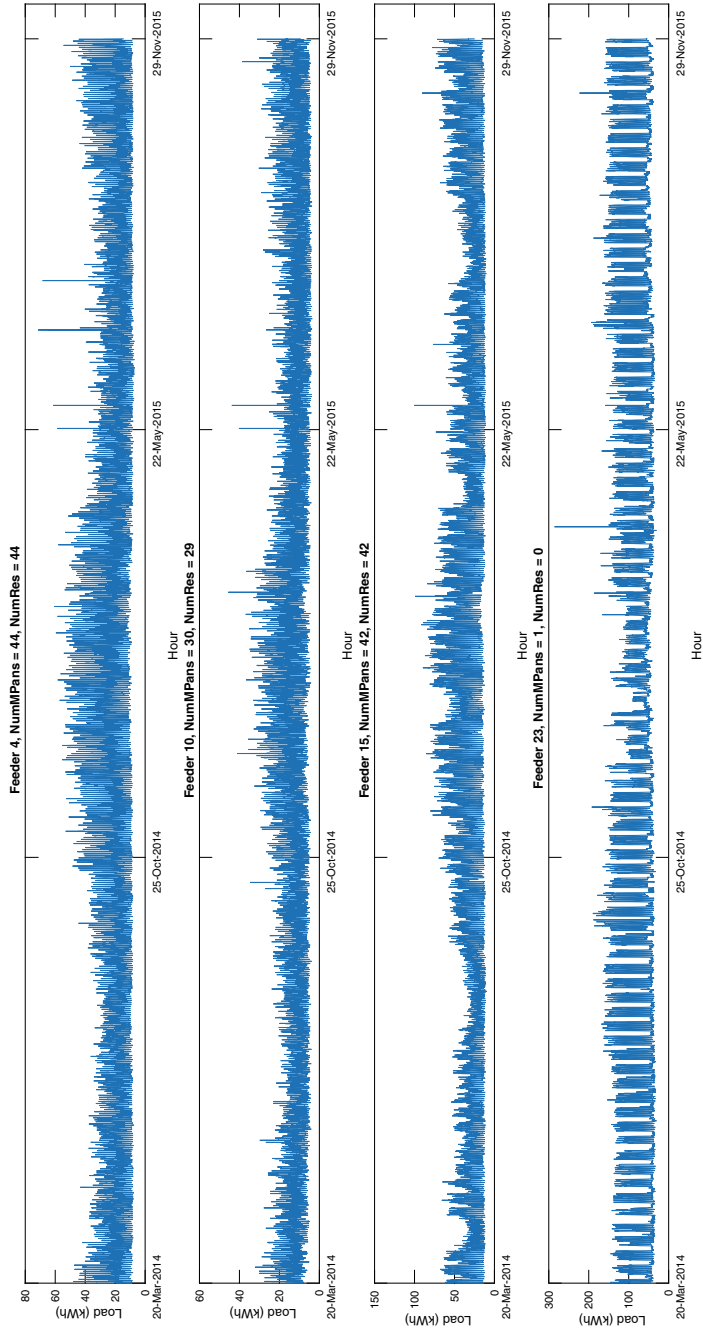


Fig. 14.1 Example of demand time series for four different residential LV feeders. Shown in the title to each is the number of MPANs (consumers), NumMPans, connected to the feeders and also how many of them are residential, NumRes

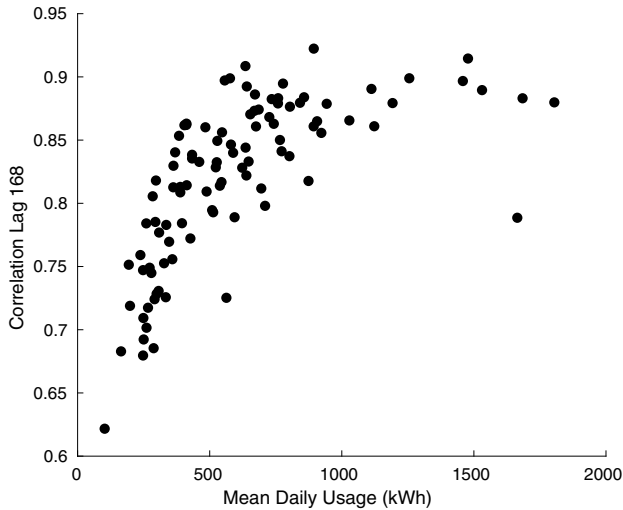


Fig. 14.2 Autocorrelation at lag 168 (weekly seasonal correlation) for all 100 feeders against the mean daily demand. Reprinted from [1] with permission from Elsevier

Given these weekly periodicities what does the average weekly demand look like for an LV feeder? Figure 14.3 shows an example of the normalised average weekly demand for three feeders each with forty consumers connected, two of which are purely residential whilst one consists of a single commercial consumer and 39 residential consumers. The data has been normalised (i.e. divided by the average weekly demand) so that the distribution of demand over the week for different feeders can be compared without being obscured by the magnitude of the demand. The plot indicates some important features:

- Daily and weekly seasonalities are quite prominent.
- Often weekdays (Monday to Friday) are very similar but Saturday and Sunday may be different from weekdays and from each other. This important observations suggests that different days of the week should be treated differently in the models (see later in Sect. 14.2.3).
- The feeder with the single commercial consumer has different patterns and distribution of demand compared to the purely residential feeders. During the week the mixed feeder has its peak demand during the day which indicates that the commercial consumer is likely dominating the demand. In contrast at the weekend, the demand more closely resembles a residential feeder with the peak in the evening which indicates the commercial consumer is no longer dominant and is likely non-operational or has reduced operation during the weekend.

The final observation suggests that there may not be a strong connection between one feeder and another and therefore there is less scope for transferring learning (Sect. 13.4) over low voltage feeders. This will not be tested. Although there is not many anomalous values in the hourly data there are still some missing values due to

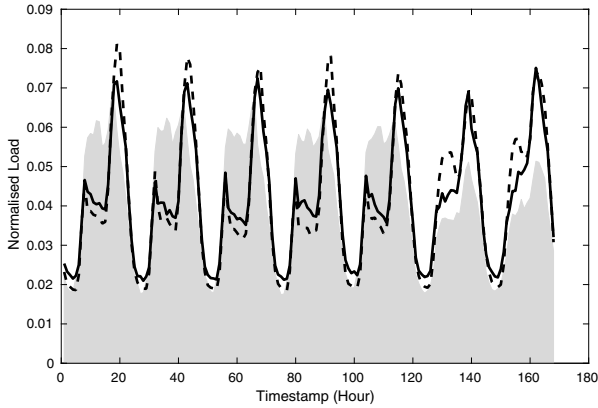


Fig. 14.3 The average normalised weekly demand for three feeders with forty consumers. The shaded profile represents a residential feeder which includes a single commercial consumers, whereas the other profiles (lines) represent feeders with only residential consumers. The profiles start on Monday

communication and sensor faults. Recall from Sect. 6.1 that anomalous/missing values can either be retained and then ignored by the model in the training phase or they can be replaced with informed estimates. The latter simplifies the training process and hence was the chosen option here. The above analysis shows there is strong evidence of weekly periodicities and autocorrelations, and this can be exploited to produce sensible values from which to impute missing data as described in Sect. 6.1.2. Each missing value is replaced with an average of the adjacent hourly demand and the value at the same hour of the previous two weeks. This ensures a final value which is weighted between the magnitude of the weekly seasonality and the locally recent demand.

Having now identified important autoregressive and time period effects in the data it is also worth considering external or exogenous variables. Temperature is often associated with load, for example, in colder temperatures more heating is required and therefore more energy is used [2, 3]. Fortunately for this trial, weather data is readily available from a nearby weather station approximately 16 km from the centre of Bracknell.

The relationship between the demand (in kWh) versus temperature (day ahead) forecasts (in degrees C) is shown in Fig. 14.4, for one of the feeders (which happens to have a particularly strong correlation with temperature) for four different time periods of the day. Also included are the lines of best fit (see Sect. 9.3) and the adjusted coefficient of determination (see Eq. (6.39) in Sect. 8.2.2) which describes how strongly the line explains the relationship. There are a few main observations from this plot

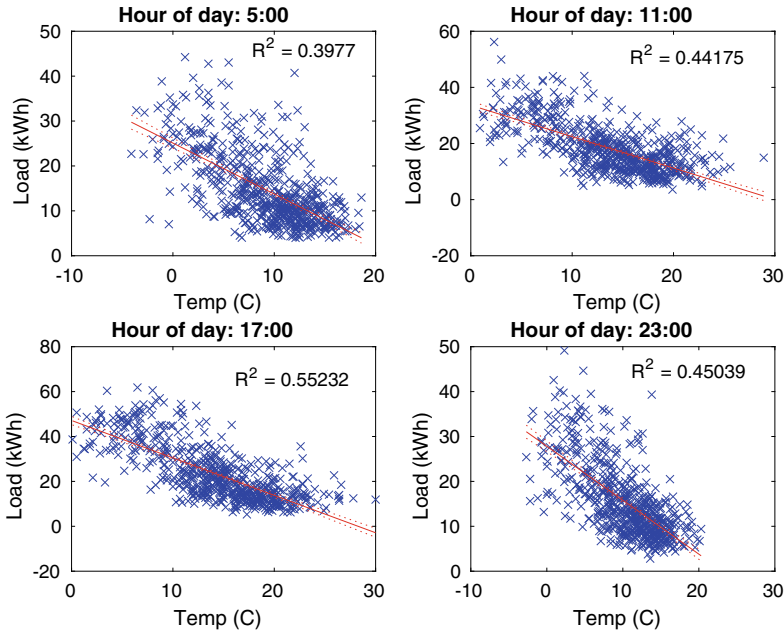


Fig. 14.4 Load versus temperature for a particular feeder for four different hours of the day. Linear fits (red lines) and adjusted R^2 values are also shown

- There is often a negative correlation between demand and temperature. The colder the temperature the more demand required. This is likely due to more electric heating and lighting being required.
- Different hours of the day have different correlations and have different adjusted coefficients of determination. Since heating behaviour is likely driven by whether the house is occupied this explains why some hours are more strongly related to temperature than others.
- Although this feeder has one of the strongest linear correlations with temperature it still isn't particularly strong ($R^2 < 0.56$).

As would be expected, the accuracy of the temperature forecasts reduce with increasing horizon. For the period 31st March 2014 to 28th Nov 2015 the day ahead temperature forecasts have a MAPE of 11.85% this reduces for every subsequent daily horizon up to 23.80% MAPE for four days ahead (i.e. between 73 and 96 h ahead) forecasts. If temperature is one of the most important factors for demand then one would expect that the accuracy of the models would decrease with increasing forecast horizons.

This section has highlighted some features which may be important and will be tested within the forecast models introduced in the next section. Of course, further analysis and techniques could be applied, such as those introduced in Sect. 6.2, to find further features (for example the possible holiday effects as suggested by the

time series plots in Fig. 14.1). However for the purpose of generating forecasts which capture the main features of the demand, the current features should suffice.

14.2.3 Model Selection

The features identified in the data analysis are not only important as inputs to the forecasts models but can be used to inform the choice of models themselves. The forecast models described here are all based on those presented in Chap. 9. Recall the aim is to generate accurate point and probabilistic, four day-ahead forecasts. By comparing the forecast models, insights can be gained on which models are most accurate, but also identify some of the more important features for describing LV level demand. Throughout the section L_1, L_2, \dots , will denote the demand time series with L_t the demand at time step t . For the probabilistic forecasts 99 quantiles will be generated for each time step in the forecast horizon.

To begin, four basic benchmarks will be defined in order to properly assess the inputs and compare to the main forecast models. As described in Sect. 8.1.1 there are several categories of benchmark models. Since there is no state-of-the-art available the focus will be on simple and common benchmarks.

Benchmark 1: Naïve Seasonal Persistence (LW)

As described in Sect. 9.1, for time series with seasonalities, a simple seasonal persistence model can be an effective choice of benchmark. For a series which has a seasonal period of s_1 this is defined as

$$\hat{L}_{t+k} = L_{t+k-s_1}. \quad (14.1)$$

Given that a weekly period is one of strongest auto-correlations in the LV demand time series (Recall Fig. 14.2) $s_1 = 168$ is chosen. This model will be called **LW** to indicate that it is the **Last-Week-as-this-week** persistence forecast.

Benchmark 2: Simple Moving Average (SMA)

The seasonal persistence forecast captures some of the seasonality in the demand time series but as shown in Sect. 9.1 it suffers from the natural variations in the demand from week-to-week. In order to smooth out these deviations the simple moving average was proposed. This simply takes the average over the same period of the week for the previous p weeks. It is defined as

$$\hat{L}_{t+k} = \frac{1}{p} \sum_{i=1}^p L_{t+k-i \times s_1} \quad (14.2)$$

where $s_1 = 168$ is again the weekly period for hourly data. The main parameter p is to be found over the training period and $p = 5$ is found to be the optimal. This

model retains the seasonality of the **LW** method but does not suffer from the random weekly variations. This model is denoted **SMA** or **SMA-pW** to indicate the p weeks of data used in the average.

Benchmark 3: Empirical CDF

A simple probabilistic model can be generated by estimating the distribution of the historical load data. For each period of the week define an empirical distribution function (see Sect. 3.4) using all the load data from the same time period over the final year (using only one year reduces any potential seasonal biases, i.e. selecting one month more than others) of the historical data. In other words to estimate the distribution of points for 2 PM on a Monday, for a particular feeder select all load values from 2 pm on a Monday. From the resultant empirical distribution, quantiles can be selected. The median of this distribution can also be chosen as the corresponding point estimate. For more details on an empirical distribution see Sect. 3.4.

Benchmark 4: Linear Seasonal Trend Model (ST)

The previous benchmarks focus on the weekly seasonal behaviour. A multiple linear model as described in Sect. 9.3 is a relatively simple benchmark which nevertheless can model more sophisticated relationships. Motivated by the analysis in Sect. 14.2.2 a linear model is constructed to produce day ahead forecasts (three other equivalent models are developed to achieve two, three and four day ahead forecasts respectively) which takes into account the annual, weekly and daily effects. One of the easiest ways to include seasonal behaviours is to use sine and cosine functions as basis function (see Sect. 6.2.5), e.g.

$$\sum_{k=1}^H \left(a_k + b_k \eta(t) + \sum_{p=1}^P (c_{k,p}) \sin \left(\frac{2\pi p \eta(t)}{365} \right) + (d_{k,p}) \cos \left(\frac{2\pi p \eta(t)}{365} \right) \right), \quad (14.3)$$

where, H is the number of time steps in a day (24 for the hourly data here), $\eta(t) = \lfloor \frac{t}{H} \rfloor + 1$, is an identifier for the day of the trial (with day 1 the first day of the trial set: 20th March 2014). The function $\lfloor x \rfloor$ here is the *floor function* and rounds down the number to the largest integer less than or equal to x , so for example, $\lfloor 2.1 \rfloor = 2$, $\lfloor -5.4 \rfloor = -6$, and $\lfloor 12 \rfloor = 12$. This simple model is a good start for describing the annual seasonality but it does not take into account the daily seasonality which was observed in the data analysis. The model can therefore be updated using dummy variables (see Sect. 9.3), e.g.

$$\sum_{k=1}^H \mathcal{D}_k(t) \left(a_k + b_k \eta(t) + \sum_{p=1}^P (c_{k,p}) \sin \left(\frac{2\pi p \eta(t)}{365} \right) + (d_{k,p}) \cos \left(\frac{2\pi p \eta(t)}{365} \right) \right), \quad (14.4)$$

where $\mathcal{D}_k(t)$ is the daily effects dummy variable and is defined by

$$\mathcal{D}_j(t) = \begin{cases} 1, & \text{if } j = t + Hk, \text{ for some integer } k \\ 0, & \text{otherwise,} \end{cases}$$

This new model captures annual seasonality by effectively producing 24 models, one model for each hour of the day. The data analysis also showed that there is strong weekly periodicities in the LV demand time series, hence one more adjustment can be applied to give the final model

$$\begin{aligned} \hat{L}_t = & \sum_{k=1}^H \mathcal{D}_k(t) \left(a_k + b_k \eta(t) + \sum_{p=1}^P (c_{k,p}) \sin \left(\frac{2\pi p \eta(t)}{365} \right) + (d_{k,p}) \cos \left(\frac{2\pi p \eta(t)}{365} \right) \right) \\ & + \sum_{l=1}^{7H} f_l \mathcal{W}_l(t), \end{aligned} \quad (14.5)$$

where $\mathcal{W}_l(t)$ a weekly dummy variable defined by

$$\mathcal{W}_j(t) = \begin{cases} 1, & \text{if } j = t + 7Hk, \text{ for some integer } k \\ 0, & \text{otherwise,} \end{cases}$$

The $\sum_{l=1}^{7H} f_l \mathcal{W}_l(t)$ term adjusts each daily hour model depending on the hour of the week and this allows the modelling of different behaviours on the weekends and the weekdays and can capture the features which were observed in Fig. 14.3. One of the hyperparameters to choose is the number of seasonal terms P . For simplicity and to avoid overfitting (see Sect. 8.1.2) this is set to $P = 3$ (Although a validation set as described in Sect. 8.1.3 could be used to properly choose this).

Although the formula looks relatively complicated the model is actually quite simple and is still a multiple linear model. Due to the presence of the dummy variables, there is in fact 168 separate models for each hour of the week for the day ahead forecasts.

The model can be easily extended to include further inputs. In this case the non-linear relationship between temperature, T_t , at time t can be included by adding a simple polynomial of the temperature. In this case, since the relationship between demand and temperature is not too strongly nonlinear a simple cubic is considered: $\alpha_1 T_t + \alpha_2 T_t^2 + \alpha_3 T_t^3$, where $\alpha_1, \alpha_2, \alpha_3$ are the coefficients for the temperature components of the multiple linear model.

Any linear model can be easily extended to a univariate probabilistic model by using the model within a quantile regression for each quantile as described in Sect. 11.4. An example of a quantile regression fit on the 6 PM data in the training set for a specific feeder is shown in Fig. 14.5 for the 10, 50 and 90 percentiles. Notice the main annual seasonality captured by the model and the small increases in demand that occur on weekends. The variation in the demand may not appear to be smooth

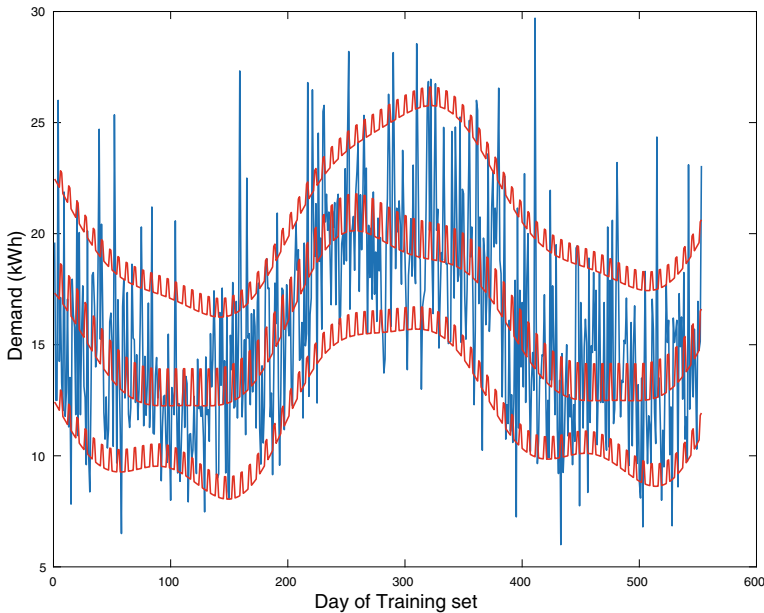


Fig. 14.5 Quantile regression fit of the simple linear model for the demand at 6PM on a specific feeder for the 10, 50 and 90 percentiles

and it may be tempting to try and force the data to have a more simple annual seasonal shape, however it is important to not try and second-guess the patterns in the data as the only true assessment of the model will be on the test set. Besides, since this is a benchmark model it is not necessary to try and make the model perfect. The variation in the annual seasonality could also be due to the small numbers of complete years in the training set. It is likely to be smoother if several years of data were available.

The above benchmarks include a number of important properties that have been discovered by the data analysis, including weather variables, and daily, weekly and annual periodicities. However, they do not include any autoregressive effects. The benchmarks will therefore be compared to a number of slightly more sophisticated models of demand which will include this feature.

Main Model 1: Seasonal Exponential Smoothing (HWT)

The double seasonal exponential smoothing method (or Holt-Winter-Taylor (HWT) method after its creators) described in Sect. 9.2 is well suited for LV demand forecasting due to its ability to incorporate two levels of seasonality and incorporate localised autoregressive behaviour. In this case the two periods parameters used are $s_1 = 24$, and $s_2 = 168$, since daily and weekly seasonalities respectively have been shown by the data analysis to be two of the most important components of the demand time series. In the HWT model, recent data contributes more to the final forecast than

older data, this means that it also implicitly models the seasonal component since the overall level of the forecast is based on locally recent information.

Once the parameters are trained for this model, a probabilistic forecast can be generated by bootstrapping the 1-step ahead residuals as described in Sect. 11.6.1. As with the other models, the median is used as the point forecast model.

Main Model 2: Auto-Regressive Models (ARWD, ARWDY)

Another way to incorporate autoregressive information is to generate an AR model on the residuals of a sensible forecast model. This is the same process as described in Sect. 7.5 which describes autoregressive correction for improving forecast models. More generally, any forecast model μ_t which estimates a time series L_t can be improved using this method if there is autocorrelation structure remaining in the residual time series $r_t = L_t - \mu_t$. In this case an autoregressive model is applied to the residual time series

$$r_k = \sum_{k=1}^p \phi_k r_{t-k} + \epsilon_t, \quad (14.6)$$

where ϵ_t is the error, and the most appropriate order p can be found by, e.g. calculating the Akaike Information Criterion (AIC), or other information criterion (see Sect. 8.2.2) for a range of different values $p = 1, \dots, p_{\max}$.

The choice of underlying forecast model μ_t is very general. As the focus is on testing the autoregressive effects, the baseline models will be kept relatively simple.

The analysis showed the importance of weekly seasonality, hence the first choice of forecast model is a simple linear model

$$\mu_t = \sum_{j=1}^{7H} \beta_j \mathcal{W}_j(t). \quad (14.7)$$

where, $H = 24\text{h}$ and $\mathcal{W}_j(t)$ is the period of the week dummy variable

$$\mathcal{W}_j(t) = \begin{cases} 1, & \text{if } j = t + 7Hk, \text{ for some integer } k \\ 0, & \text{otherwise,} \end{cases}$$

as used in the ST benchmark forecast. The parameters to train are the coefficients β_j and these are estimated by simple ordinary least squares (OLS) (see Sect. 8.2) over the initial prior year of historical loads. From this model the residuals are calculated and the estimates from the residual model, \hat{r}_t is added to μ_t to give the final forecast

$$\hat{L}_t = \mu_t + \hat{r}_t. \quad (14.8)$$

This model will be denoted **ARWD** to signify autoregressive model with weekday mean.

A second forecast model is also considered which adds onto the ARWD mean model by including a term for annual seasonality seen in the data analysis, this is given by

$$\mu_t = \sum_{j=1}^{7H} \beta_j \mathcal{W}_j(t) + \sum_{k=1}^K \alpha_{1,k} \sin(2\pi tk/A) + \alpha_{2,k} \cos(2\pi tk/A) \quad (14.9)$$

with parameters β_j and $\alpha_{j,k}$, and $A = 365H$ as annual seasonality. The annual seasonality is modelled by a Fourier approximation of order K which is fixed to $K = 2$ to reduce the complexity. The dummy variable $\mathcal{W}_j(t)$ is as in Eq. (14.7). As with the **ARWD** model the μ_k is estimated by OLS between the model and the training data. This model is used to calculate a new residual time series which is then also trained by OLS and is added to the mean model in Eq. (14.9) to give the final forecast \hat{L}_t given by

$$\hat{L}_t = \mu_t + \hat{r}_t. \quad (14.10)$$

as with the ARWD model. This model is denoted **ARWDY** with the Y signifying the yearly periodicities included through the Fourier terms. Note that separate ARWD/ARWDY models are used depending on whether the forecasts are one, two, three or four days ahead.

Notice the subtle differences between the modelling of the seasonalities in this model versus the ST model. In the ARWDY model the periodicities are not separated for different periods of the day. If there are significant differences in how seasonalities affect different times of day then perhaps the ST will perform slightly better. However, the ARWDY also includes autoregressive effects which now incorporates more interdependencies between hours of the day. As with the ST methods the weather effects are included by adding linear terms to the mean equations.

These regression models will serve as point forecasts. To extend them to probabilistic forecasts the slightly more sophisticated GARCH type model, described in Sect. 11.6.2 will be considered. In this the variance itself will be modelled by considering the final model residuals $\epsilon_t = \hat{L}_t - L_t$ which are assumed to have the form $\epsilon_t = \sigma_t Z_t$ where σ_t is the conditional standard deviation of ϵ_t and $(Z_t)_{t \in \mathbb{Z}}$ is an independent identically distributed random variable with $\mathbb{E}(Z_t) = 0$ with $\mathbb{V}ar(Z_t) = 1$. The method is described in detail in Sect. 11.6.2 and requires a model for the standard deviation. Since the variation in demand is likely to be correlated with the size of the demand (larger demands have more variation) the same mean model for the point forecast will be used for the standard deviation. For example in the case of ARWDY the model will be

$$\sigma_t = \sum_{j=1}^{7H} \tilde{\beta}_j \mathcal{W}_j(t) + \sum_{k=1}^K \tilde{\alpha}_{1,k} \sin(2\pi tk/A) + \tilde{\alpha}_{2,k} \cos(2\pi tk/A). \quad (14.11)$$

where the coefficients $\tilde{\beta}_j$, $\tilde{\alpha}_{1,k}$, $\tilde{\alpha}_{2,k}$ are to be found and the tilde over the parameters is used to distinguish them from the coefficients from the mean model. Once the

standard deviation σ_t and mean \hat{L}_t are found the bootstrap procedure (as introduced in Sect. 11.6.2) can be employed to generate an empirical distribution (see Sect. 3.4) for each time step in the forecast horizon. To do this perform the following operation, for the forecast starting at time step $t = N + 1$

1. Draw a sample \hat{Z}_t from empirical distribution of the random variables Z_t .
2. Scale the variable with the standard deviation to give a residual $\epsilon_{N+1} = \sigma_{N+1} \hat{Z}_t$.
3. Add this to the mean forecast $\tilde{L}_{N+1} = \hat{L}_{N+1} + \epsilon_{N+1}$.
4. Use this current value within the forecast inputs to generate the forecast for the next time step \hat{L}_{N+2} .
5. Repeat this process until forecasts have been generated for all time steps in the forecast horizon.

The bootstraps generate a multivariate probabilistic forecast, but these can be transformed into a univariate probabilistic forecasts for each time step by fitting a distribution or calculating the empirical quantiles at each time step from the generated points (Sect. 3.4).

14.2.4 Testing and Evaluation

A diverse selection of models have been described in Sect. 14.2.3. Since they all have slightly different structures and use different features as inputs they can be used to test a variety of hypothesis and assumptions. As discussed in Sect. 14.2.1 there are the following main questions that can be analysed via the errors on the test set:

1. What is the effect of temperature?
2. How does accuracy change with forecast horizon?
3. How does accuracy change for different feeders?
4. Which features are the most important for an accurate forecast?

The models are all trained on a training set which covers the dates from 20th March 2014 to the 30th September 2015 inclusive and rolling four day-ahead forecasts have been generated for the 53 day testing set starting 1st October 2015. Notice that no validation set has been used in this case (Sect. 8.1.3) this is for two reasons, firstly all the models use a relatively small number of parameters and hence have low chance of being over fitted to the data, hence the model selection step on the validation set has been skipped for this trial. Secondly, although there is around a year and a half of data this is not particularly large for a data set with annual seasonality, hence a validation set would require splitting the data further and would potential reduce the reliability of the results on the test set. Hence the larger training set increases the chance of properly training the model parameters.

To begin consider a comparison of the forecast models generated in Sect. 14.2.3. Table 14.1 shows the average score over all four day-ahead forecasts for the entire 53 day test period for all 100 feeders using the MAPE, RMAE and RCRPS measures (Sect. 14.2).

First consider the point forecasts (MAPE and RMAE scores). The main observations from the results are as follows:

- Of the benchmark models the simple persistence model (LW) is the least accurate whereas the moving average using 5 weeks (SMA-5W) and the simple seasonal regression (ST) are the best performing. This suggests averaging the weekly historical data is useful for producing more accurate models than using a simple last-week-as-this-week value.
- Of the two best benchmarks the ST forecast is slightly more accurate than SMA-5W (only 2% lower MAPE), which suggest including annual seasonality can be beneficial but only slightly.
- The main models (ARWD, ARWDY and HWT) are more accurate than the benchmarks. These models all have autoregressive features and hence suggests there are important temporal interdependencies in the demand time series.
- The ARWD and ARWDY methods are slightly better than the HWT method. One of the main differences between these models is that HWT only explicitly uses the previous lag (although the previous lags are included implicitly as smoothed historical terms) and thus suggests whilst the most recent previous demand is an important indicator of the demand, older lags are also important for determining the current demand.
- There is very little difference between the ARWD and ARWDY forecast models, with the ARWD performing slightly better, on average, across all measures. Thus an explicit seasonality term has limited importance in the forecast accuracy compared to the autoregressive term.

The probabilistic forecasts in fact show the same ranking of the methods as the MAPE and RMAE, with ARWD ranked as the most accurate method, followed by ARWDY, then HWT, then ST and then the Empirical method. This is an encouraging result since it suggests that the accuracy of the point forecasts may be indicative of the accuracy of the probabilistic forecasts. Probabilistic methods are typically more expensive to train and therefore if the point forecasts can be used to rank the probabilistic forecasts this significantly reduces computational cost of identifying and training these methods. However caution must be exercised as this is only an empirical observation and hasn't been established theoretically.

Temperature Effect

Table 14.1 does not consider the influence on temperature on the forecast accuracy of residential LV network demand. As shown in Sect. 14.2.2, there seems to be a relatively low correlation between temperature and the demand despite the fact that the temperature is often seen as strongly connected to electricity demand due to its obvious connection to heating and cooling behaviours. The MAPEs for particular point forecasts that use and don't use temperature forecasts are shown in Table 14.2. The table suggests that the weather is not a strong driver of the demand. The benchmark method, ST, does improve slightly, however the most accurate models, ARWD and ARWDY, both become less accurate when temperature is included. Why would this be the case?

Table 14.1 MAPEs, RMAEs and RCRPSs for all forecast methods over all 4 day-ahead horizons for the entire 53 day test period for all 100 feeders. The lowest errors for each score are highlighted in bold. Reprinted from [1] with permission from Elsevier

Method	Error scores %		
	MAPE	RMAE	RCRPS
LW	18.67	18.93	–
SMA-5W	15.73	16.77	–
Empirical	16.19	16.96	12.62
ST	15.42	15.42	10.97
HWT	14.84	15.01	11.06
ARWD	14.65	14.67	10.32
ARWDY	14.64	14.80	10.44

Table 14.2 MAPEs for the methods showing the effect of including temperature forecast data in a selection of methods. Reprinted from [1] with permission from Elsevier

Method	None	Temperature
ARWD	14.65	16.94
ARWDY	14.6	15.16
ST	15.42	15.16

To further investigate the effect of including temperature as an input, consider what happens when the MAPE scores are split according to forecast horizon (at the daily resolution) as shown in Table 14.3. For comparison, the MAPEs of the temperature forecasts themselves are included. The temperature forecast drops in accuracy by more than 80% from one day-ahead to four days-ahead. Thus if there is a strong dependence on temperature it would be expected that the models trained using the temperature would also drop off in accuracy at a comparable rate. In fact the ST demand forecast accuracy changes very slightly and the ARWDY and ARWD demand forecasts drop in accuracy only by 4.3% and 5.6%. Further experiments can be included, for example including lagged temperature values in the demand forecast models. However, in all cases the results are the same: temperature doesn't appear to have a strong effect on the demand forecast accuracy. If we examine the individual feeders the temperature is only shown to improve the forecasts of 19 out of 100 feeders, and in all cases the MAPEs do not improve by more than 4%.

The inclusion of the temperature also doesn't improve the probabilistic forecasts. In Fig. 14.6 shows the reliability plot (See Chap. 7) for the probabilistic forecasts generated using the ARWDY model using no temperature (solid dots), using actual temperature (unfilled dots) and using the forecast of the temperature (crosses). The diagonal line shows the expected line if the quantiles generated from the model (i.e. the predicted spread of the data) matched the empirical quantiles. It is clear that the model not using any temperature is closest to this line hence showing that including the temperature (whether forecast or actual) does not in fact improve the forecast.

Table 14.3 MAPE Scores for different day ahead horizons for a selection of methods which use the forecast temperature values as inputs. Also for comparison is the average MAPE for the temperature forecast themselves. Reprinted from [1] with permission from Elsevier

Method	MAPE			
	Day 1	Day 2	Day 3	Day 4
ARWD	16.51	17.26	17.12	16.89
ARWDY	14.75	15.11	15.31	15.46
ST	15.12	15.21	15.16	15.16
Temperature	8.98	10.57	13.46	16.47

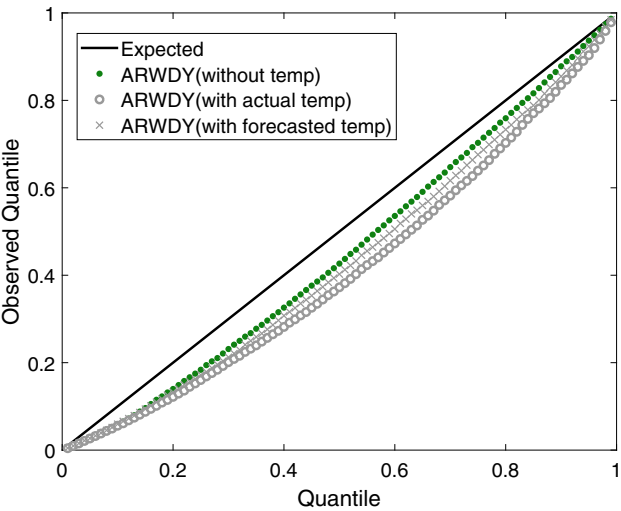


Fig. 14.6 Reliability plot for the ARWDY forecast for different temperature inputs. Reprinted from [1] with permission from Elsevier

There could be many explanations for the lack of influence of the temperature data, for example, the seasonality could be the main driver of demand and the perceived correlation between demand and temperature is actual only due to the collinearity between seasonality and temperature. In fact seasonality may be a confounding variable in this situation (Sect. 13.6.1). In addition, it could be that much of the heating for the consumers on these feeders use gas instead of electrical boilers and hence temperature will only have a minimal effect. However, regardless of the reason, for a forecaster the results for this particular data and test set indicate that temperature is not a particularly important input for these forecast models. This result also highlights an important lesson: even if there is strong intuitive reasons for an explanatory feature to be important, it does not necessarily translate to importance for the forecast model. Going forward with the analysis the temperature is not going to be considered any further.

Forecast Accuracy and Horizon

Most forecasts become less accurate the further ahead they predict (Sect. 7.3). One reason for this is that often recent values can be strong indicators of the near future demand. Further into the future, the most recent observations are much older and hence not as useful at making accurate predictions. If a forecast can retain its accuracy for longer time horizons into the future then they can be useful for more longer term planning. For storage applications this means longer term plans can be made for when to charge and discharge the device.

Table 14.4 shows the forecast accuracy (MAPE) for selected methods as a function of days ahead. In other words the ‘Day 1’ column means the average error from forecasting between 1 and 24h ahead, the ‘Day 2’ column indicates the average error from forecasting between 25 and 48h ahead etc. For ARWDY the MAPE errors only increase from 14.34% to 14.87%, i.e. a 3.7% increase. In fact there is only a small drop in the accuracy for any of these models and this indicates that the models can offer similar accuracy for estimating tomorrows demand as they do for four days time. Computationally speaking, this can be quite advantageous as it can reduce the cost of model retraining with minimal impact on the forecast accuracy.

How does the accuracy change at the hourly level? This time consider the probabilistic forecasts (recall the results are qualitatively similar whether the point or probabilistic forecasts are considered). Figure 14.7 shows the relative CRPS errors as a function of hourly horizon for selected methods. Recall, the forecasts all begin at 7 AM of each day and hence the first horizon point corresponds to the period 8–9 AM. The following observations can be made

- The intraday shape of the errors are similar. Hence the period of the day is a major indicator of the accuracy of the forecasts. Notice that the areas of highest errors (largest CRPS) correspond to periods typically associated to high demand (and hence high volatility), e.g. around the evening period. The overnight periods (around 10PM until 5AM) have the lowest errors. These are typically periods of low activity. This supports using a GARCH type model where the volatility (standard deviation) is correlated with the average demand.
- Although the shapes are similar, there is a small trend of increasing error from one day to the next.
- Different forecasts are more accurate for different periods of the day. For example, although ARWD is generally the most accurate model, for some evening periods

Table 14.4 MAPE Scores for each method over each day ahead horizon. Reprinted from [1] with permission from Elsevier

Method	MAPE			
	Day 1	Day 2	Day 3	Day 4
HWT	14.56	14.83	14.95	15.04
ARWDY	14.34	14.59	14.75	14.87
ST	15.36	15.41	15.44	15.49

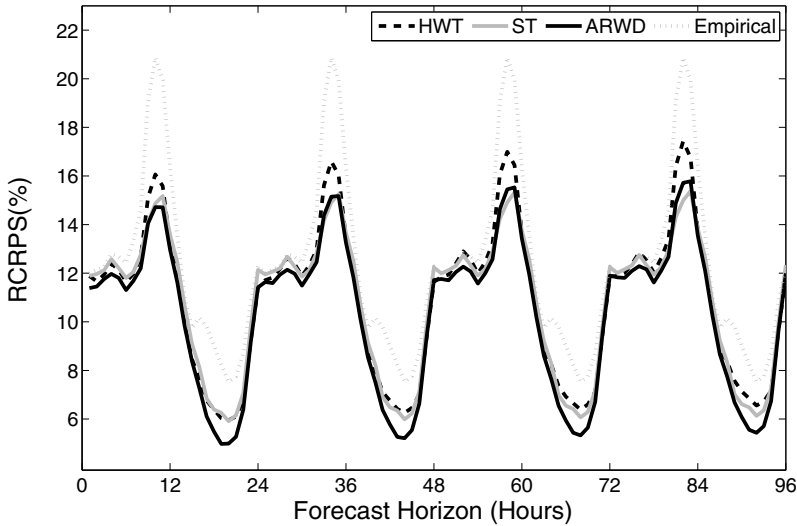


Fig. 14.7 Plot of average Normalised CRPS for selected methods for horizons from 1 h to 96. Reprinted from [1] with permission from Elsevier

the ST benchmark is actually more accurate. Hence one potential way of generating a more accurate forecast could be to take a combination of several models (see Sect. 13.1).

Forecast Accuracy and Feeder Size

The final piece of analysis concerns comparing the accuracy across different feeders. As previously mentioned the feeders come in all shapes and sizes. Some have up to 109 customers connected whereas some have as few as one. Further there is a diversity of the types of customers. Some are commercial, but most are domestic and even amongst the domestic customers there is a wide diversity in their behaviours. Are you like your neighbour? In addition, there are other loads that are not monitored: street lights, elevators, cameras, landlord lighting etc. which all contribute to the load shape and diversity.

Figure 14.8 shows the MAPEs for the ARWDY model for each individual feeder as a function of the average daily demand. Each point represents a different feeder and some of them have been given different icons to signify different categories of feeders. An instant observation is that 88 of the feeders closely fit a power law curve (the bold curve in the plot). The smaller feeders tend to be more volatile and therefore have larger relative error values. In contrast the larger feeders have smaller errors. This can be explained largely by the law of large numbers. The increased size of the feeder corresponds to larger numbers of customers connected to the feeders, which result in smoother and more regular demands that are easier to forecast.

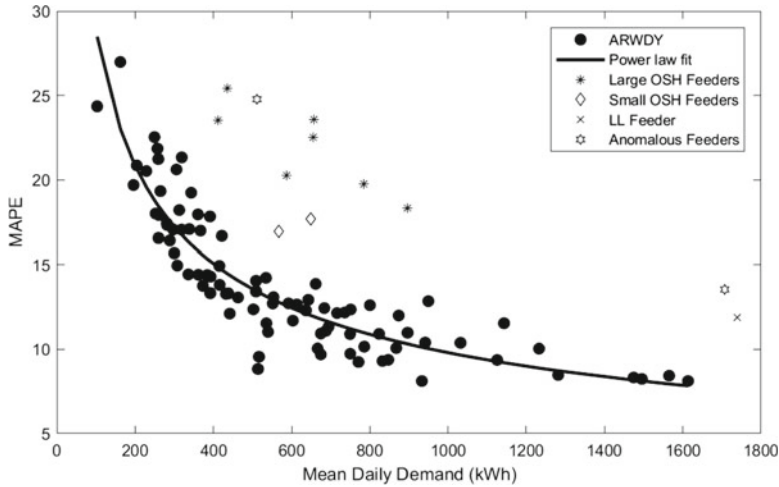


Fig. 14.8 Scatter plot of the relationship between MAPE and mean daily demand for two different forecasting methods. Feeders which apparently have overnight storage heaters or have usually larger errors have been labelled separately as OSH and anomalous respectively. Also shown is a power law fit to the non-OSH/anomalous feeders. Reprinted from [1] with permission from Elsevier

There are twelve feeders which don't fit the power law relationship. This unusual behaviour should prompt the investigator to try and better understand why these feeders don't fit the general trend. By looking at the average profile of seven feeders an immediate observation is that the feeders have unusually large overnight demand. This prompted a consideration of what type of consumers were on these feeders. In fact it was found that 75–85% of customers on each of these seven feeders had large numbers of overnight storage heaters (OSH). Overnight storage heaters are heaters which use energy during the night to store up heat and then release this energy during the day. These are labelled “Large OSH Feeders” in the plot. Further to this, two other feeders had smaller overnight demands, their feeders had 62% and 75% of their customers with OSHs, these are labelled “Small OSH Feeders”. This in itself isn't enough to explain these feeders not obeying the power law. It also must be confirmed that none of the other 88 feeders also have large proportions (greater than 60%) of OSH or profiles which have large overnight demands. In fact this was found to be the case with these feeders and hence strongly suggests the presence of a high proportion of OSH effects the accuracy of the chosen forecast models in unexpected ways. What about the other three anomalous feeders? At least one of them was further found to be unique. In fact the largest feeder was found to be a landlord lighting connection for a large office block. The final two are inconclusive as the connectivity information is incomplete.

These results are important on a number of levels:

1. The power law relationship suggests that forecasts are more accurate for larger feeders than smaller feeders. For storage applications this can help make decisions on where to use a storage device. For example, on larger feeders an accurate forecast can be generated which means a storage device is more likely to be optimally controlled. In contrast a storage device may be unsuitable for a smaller feeder since the demand is too volatile.
2. It suggests more bespoke methods are required for those feeders which have unusually large errors (those with large deviations from the power law relationship). Identifying these feeders creates opportunities for developing improved forecasts which will increase the opportunities for applications such as storage control to a wider class of substations and cases.

14.3 Example Code

To demonstrate some of the methods and techniques described in this book, and to show how they are implemented in real code, a python notebook has been shared to show some of the steps in analysing data and developing a model. The code can be found at the following repository: <https://github.com/low-voltage-loadforecasting/book-case-study>.

The notebook will briefly demonstrate topics including:

1. **Exploratory data analysis** using common Python plotting libraries matplotlib¹ and Seaborn,²
2. **Feature modelling** using common Python data library Pandas,³
3. **Cross-validation** using machine learning library Scikit-learn,⁴
4. **Model fitting and selection** (including simple benchmarks) in Python. In contrast to the previous section the code will focus more on machine learning models (Chap. 10). In particular, common machine learning packages/libraries such as Scikit-learn,⁵ and TensorFlow⁶ will be presented.
5. **Model evaluation and diagnosis**.

¹ See <https://matplotlib.org/stable/index.html>.

² See <https://seaborn.pydata.org/index.html>.

³ See <https://pandas.pydata.org/>.

⁴ Found at <https://scikit-learn.org/stable/>.

⁵ Found at <https://scikit-learn.org/stable/>.

⁶ See <https://www.tensorflow.org/>.

14.4 Summary

This chapter has highlighted the major components to creating and analysing a successful demand forecast. The chapter has shown, through a residential LV network application, how to apply the techniques and methods described in the previous chapters in order to properly design a forecast trial given the available data. Basic plots such as ACFs and scatter plots have been used to identify key relationships, and these have been supported through various statistical summaries. Point and probabilistic forecasts have been considered and compared using a range of error measures. The comparison of these models was used to identify some of the important features and key relationships.

The chapter has also highlighted the importance of benchmarks for better understanding the forecasts, the role data analysis plays in creating the models but most importantly the chapter has shown the importance of questioning basic assumptions about data and explanatory variables. For example, for load forecasting of residential demand, temperature is often included in all models as it is assumed to be a driving factor for the demand. However, at least in this specific example, including temperature could reduce the forecast accuracy.

A code has been shared with this book and described in Sect. 14.3. This helps to demonstrate how to implement some of the methods and techniques in practice. The reader is encouraged to experiment with generating their own forecasts. A guided walk-through is given in Appendix C which can be used to go through the main steps, from data cleaning to testing.

14.5 Questions

For this section, the ask is to run your own forecast trial. You can follow the same procedure as in Sect. 14.2, or follow the more extensive steps given in Chap. 12. You can also follow the step-by-step walk-through in Appendix C. To perform the experiment select a demand time series from one of those shared in Appendix D.4.

It is also recommend running the code linked to in Sect. 14.3 to get some ideas for practical analysis and implementation.

References

1. S. Haben, G. Giasemidis, F. Ziel, S. Arora, Short term load forecasting and the effect of temperature at the low voltage level. *Int. J. Forecast.* **35**, 1469–1484 (2019)
2. S. Rahman, Formulation and analysis of a rule-based short-term load forecasting algorithm. *Proc. IEEE* **78**, 805–816 (1990)
3. M. Bessec, J. Fouquau, The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. *Energy Econ.* **30**(5), 2705–2721 (2008)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

