TECHNISCHE UNIVERSITÄT WIEN

Diplomarbeit

Error analysis of wind energy prediction models

Ausgeführt am Institut für
Stochastik und Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von
Ao. Univ. Prof. Wolfgang Scherrer

durch
Eni Ballauri
Schönburgstraße 3/8
1040 Wien, Österreich

Wien, Oktober 2016 Eni Ballauri

Acknowledgements

For the development of this thesis I would like to express my deep gratitude to Ao. Univ. Prof. Wolfgang Scherrer, who supervised it and helped me during the process. I would also like to thank Dipl. Ing. Kurt Misak who provided me the opportunity to write this thesis using the data from the Austrian Power Grid as well as Ing. Gernot Wolfram for helping me with questions regarding the prediction models.

I also want to acknowledge all my colleagues, who have accompanied me during the study years and have often helped me with ideas.

Finally I want to thank my family for their unconditional support and Felix for always being there in difficult times.

Abstract

The importance of methods using renewable resources for the energy production has increased drastically in the recent years. In the electricity market the method that has grown most rapidly is the generation of power through wind. Forecasts of this power production are crucial. However, difficulties arise in these forecasts due to the high volatility of the wind and the complexity of the terrain where the wind facilities are installed. This thesis examines five different models for the forecast of wind energy and analyses the behaviour of its errors. The scope is to build a confidence interval for the errors of these forecasts. In this framework two major setups are considered for constructing a confidence interval. The first one is based on applying a GARCH model in order to forecast the conditional variance of the errors. The second uses a regression model to describe the absolute or square errors as a function of explanatory variables like the predicted wind speed and predicted wind power. These models were then used to build the confidence intervals. Both approaches were implemented sucessfully.

Contents

CHAPTER 1

Introduction

During the last years the subject of exploitation and development of renewable energy resources has received much focus due to concerns about increased CO_2 emission, global warming and changing climatic patterns. The price of gas and fossil fuels has become very volatile and insecure lately, which also has economical implications. Another issue concerning oil and gas resources is that they are concentrated in specific regions, so they often have to be imported, whereas energy production through renewable resources is usually based on domestic ground. Additionally to that, non-renewable energy resources are a finite quantity considering the human time-frame, so alternative solutions using renewable resourced need to be adapted in order to satisfy society's increasing energy demand.

In Europe, wind power generation has shown the most rapid growth in the electricity market of renewable sources for the past decade. By the end of 2014 about 10.2% of all European power demand was covered through wind generated power. Austria is also one of the countries that rely on this type of energy. The installed facilities here have the potential to produce 4.5 billion kWh of energy yearly, which could cover the demand of 35% of the Austrian households [15]. The challenge of integrating wind power efficiently in the electric power system though arises in its forecast due to the high variability of wind as well as the complexity of the terrain where the wind facilities are installed. An accurate estimation is very important for grid operators in order to allow an efficient scheduling of power generation according to costumer demand. Both under- and overestimation of generated power may have negative financial consequences for the grid operator:

(a) An overestimation would lead to a lack in actual supply to fulfill the demand and therefore there would be a need to buy the rest of the demanded power capacity.

(b) An underestimation means there is more actual power than predicted, therefore the excess power has to be sold. However, depending on market stocks, it might also occur that the grid operator has to pay to discharge the surplus power, resulting in an economical loss on their behalf.

Generally the prediction frameworks for wind power estimation are categorised into two methods:

(a) *Physical* - this method includes the consideration of physical factors such as topography, terrain complexity, local temperature and pressure to estimate the wind field and then converting the wind speed to power i.e. by using the power curves provided by the turbine's manufacturer

(b) *Statistical* - this implies the use of statistical models to forecast power through using historical data and establishing a relationship between power and other (meteorological) variables.[12]

Often also hybrid methods are used for the forecast, which provide a combination of physical and statistical methods.

In regard to the importance of forecasting accuracy, this thesis will provide an analysis of five predefined wind power forecasting models (ECMWF - two model variations and ALA - three model variations) used from the Austrian Power Grid. The region observed in this framework is that of Burgenland, Austria and it accounts for data from January 1, 2013 up to December 31, 2014. The models that will be analysed in this thesis are used to forecast the power for the upcoming day. The weather prediction data is generated each day at midnight and at 8-9 o'clock the power prediction is made for the day ahead starting from midnight and continuing 24 hours after that in 15 minute-intervals. All the data used in this thesis was provided by the Austrian Power Grid (APG). The predictions of the weather data used for the wind energy forecasts in APG are provided by the central institution of meteorology and geodynamics (Zentralanstalt für Meteorologie und Geodynamic - ZAMG).

The scope of this thesis is to analyse the errors of the five wind energy prediction models used by the Austrian Power Grid and based on the observed patterns to define a confidence interval for these errors. This should help to provide a better view on the interval where the actual power production values are to be found based of the provided estimations. To underline the importance of this subject, figure 1.1 shows the ratio of how much power is misestimated to how much power is produced on a monthly basis. On several months this ratio reaches about 0.5 or above meaning that the amount of power misestimated from the model is about half as much as the overall power production in that specific month.

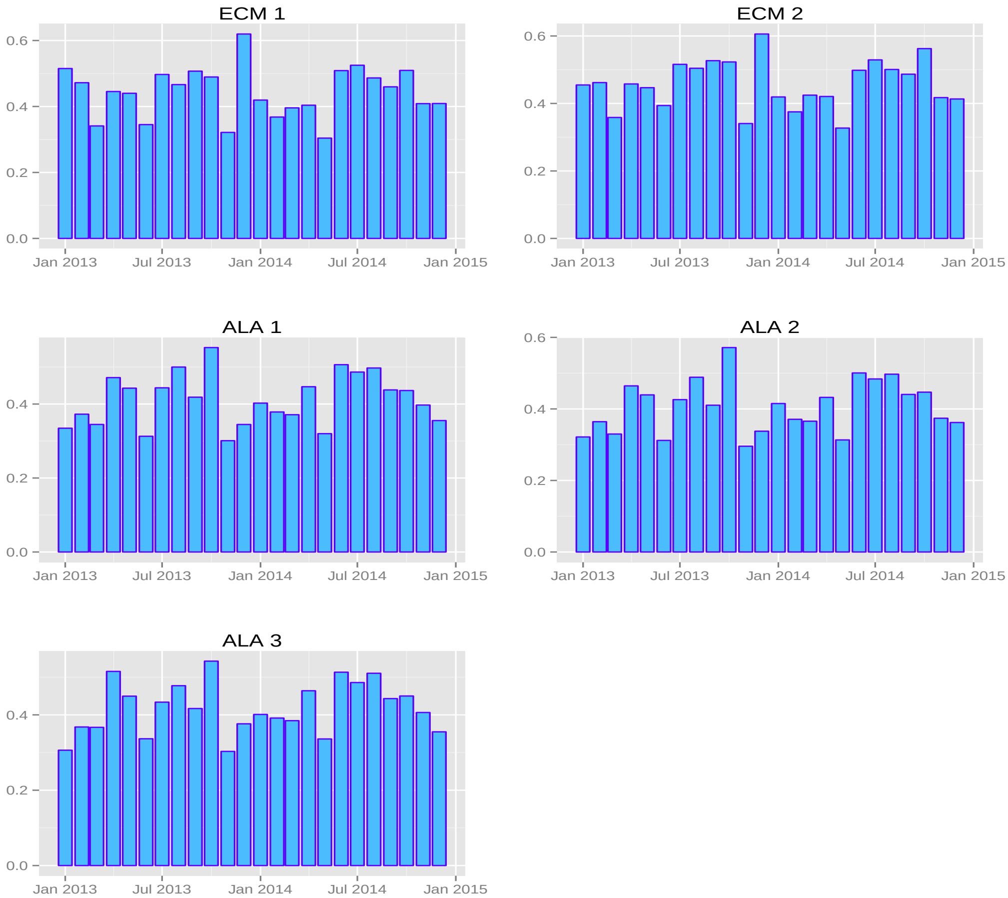

Figure 1.1: Ratio of the absolute misestimated power and actual power production on a monthly basis

This thesis is divided into four main sections:

1. In the first part a basic statistical analysis of the errors of the five models will be made, while trying to find monthly, daily or hourly patterns as well as distribution patterns of the errors.

2. The second part will consider applying a GARCH modeling approach

on the errors in order to predict their volatility.

3. The third part consists in analysing patterns of the daily mean square and absolute errors as well as establishing how they correlate with other variables.

4. The last section involves modeling different confidence intervals for the errors and evaluating their performance.

CHAPTER 2

Analysis of the forecast errors

2.1 Statistical analysis of the day ahead forecasting errors

This section shows a basic analysis of the errors of the five different models for the day ahead power forecast based on the data obtained by the Austrian Power Grid. The actual forecasting models are not provided for this thesis and will not be discussed, however it is provided which variables are used for each of the five models. The scope in this section is to analyse how these errors behave and what correlations exist. The data collected for the purpose of this thesis is given in 15 minute - intervals and the power unit used is megawatt (MW). For a more comparable analysis throughout this thesis the resulting prediction errors are normalised through the installed power (the maximal capacity of the power generator):

$$e_{it} = \frac{\hat{\epsilon}_{it}}{I_t} = \frac{p_t - \hat{p}_{it}}{I_t}$$

where e_{it} represents the normalised errors of model $i \in \{ECM_1, ECM_2, ALA_1, ALA_2, ALA_3\}$ at time t, $\hat{\epsilon}_{it}$ is the error of the forecast model i, I_t the installed power capacity, p_t the measured power and $\hat{p}_{it}$ is the predicted power at time t from model i.

As it can be seen in figure 2.1, the installed capacity has changed drastically from the beginning of 2013 to the end of 2014 and therefore normalising the errors through the installed capacity is important. From now on when talking about errors throughout this thesis, the normalised errors are meant unless stated otherwise.

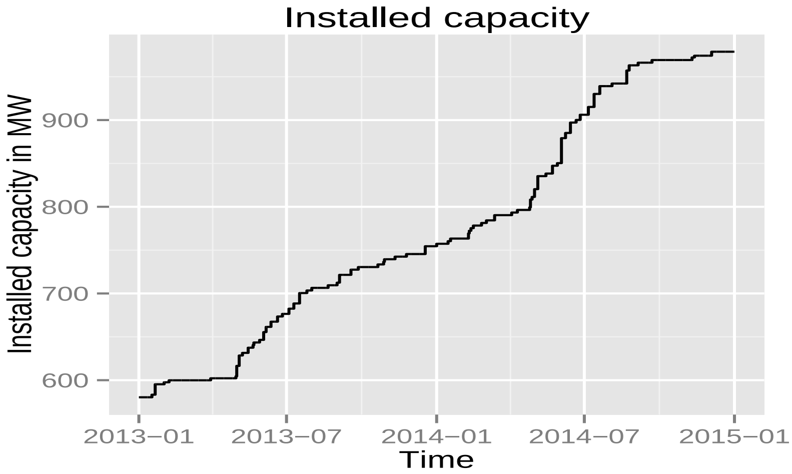

Figure 2.1: Installed power throughout the time

As the actual power forecast models are not provided, a visual representation of the forecasted values is shown in figure 2.2 for a random time interval. The black line indicates the measured power, while the other lines show the forecasted values from the two ECMWF models (labeled as ECM 1 and ECM 2) and the three ALA models (indicated as ALA 1, ALA 2 and ALA 3).

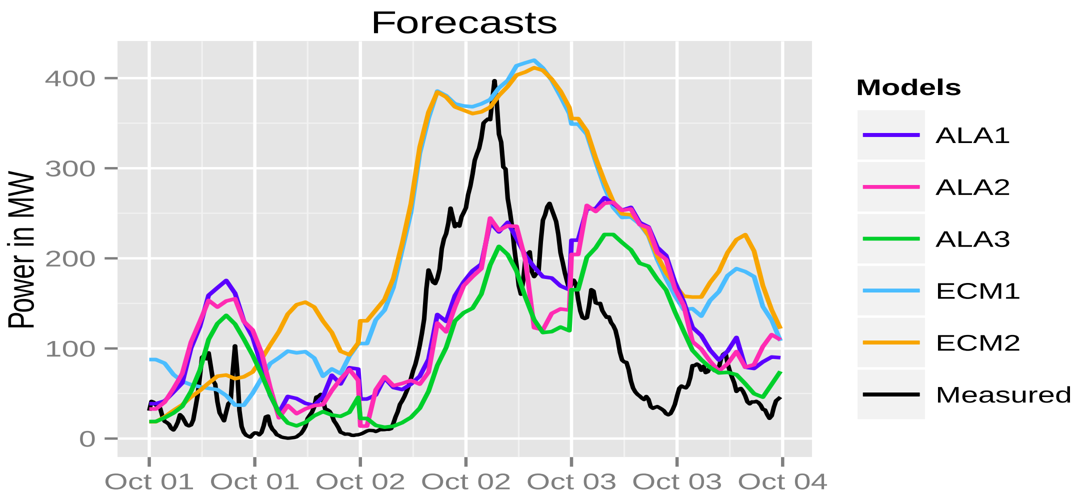

Figure 2.2: The five forecasts for a random time interval of 3 days

At first we look at some basic statistical properties of the calculated normalised errors with sample size T in order to evaluate the five forecasting models:

1. The *mean* value of the errors:

$$\mu = \frac{1}{T}\sum_{t=1}^{T} e_t$$

2. The *variance* and the *mean square error*, which indicate how much the data errors are spread:

$$\sigma^2 = \frac{1}{T-1}\sum_{t=1}^{T}(e_t - \bar{e})^2$$
$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

3. The *mean absolute error*, which gives an indication of how close the predicted values are in regard to the actual measurements:

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |e_t|$$

4. The *coefficient of determination*, which is a measure of the explained variance by the prediction model in relation to the total variation:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(p_t - \hat{p}_t)^2}{\sum_{t=1}^{T}(p_t - \overline{p})^2}$$

where $\overline{p}$ is the average power measured and $\hat{p}_t$ is the estimated power.

5. The *index of agreement*, which evaluates the skill in predicting variations about the observed mean. IOA_1 represents the original definition of the index of agreement as presented by Willmott and Wicks in 1980, while IOA_2 refers to an improvement of the first index as presented in [14]:

$$IOA_1 = 1 - \frac{\sum_{t=1}^{T}(p_t - \hat{p}_t)^2}{\sum_{t=1}^{T}(|p_t - \overline{p}| + |\hat{p}_t - \overline{p}|)^2}$$
$$IOA_2 = 1 - \frac{\sum_{t=1}^{T}|p_t - \hat{p}_t|}{\sum_{t=1}^{T}(|p_t - \overline{p}| + |\hat{p}_t - \overline{p}|)}$$

Both indexes take values in the interval $[0,1]$, where the higher the index, the better the model performance. The advantage of IOA_2 in comparison to IOA_1 is that it approaches the value 1 more slowly as the predicted values $\hat{p}_t$ approach the observed values p_t. This means that when comparing models that perform relatively well, there is a greater separation in index values, so it makes the better performing model more evident. Also IOA_2 is less sensitive to errors that are considered as outliers [14].

6. *Excess kurtosis*, which offers a description over the tail weight of a distribution in comparison to the kurtosis of the normal distribution (=3):

$$kurtosis_{excess} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{e_t - \bar{e}}{\sigma}\right)^4 - 3$$

7. *Skewness*, which measures the asymmetry of the distribution function:

$$skewness = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{e_t - \bar{e}}{\sigma}\right)^3$$

Table 2.1 shows the statistical results mentioned above applied on the normalised errors obtained from the five forecasting models for the time period January 1, 2013 - December 31, 2014.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
μ	0.0007	0.0105	0.0113	0.0093	0.0238
σ^2	0.0209	0.0214	0.0178	0.0173	0.0185
MSE	0.0209	0.0215	0.0179	0.0174	0.0191
MAE	0.1034	0.1063	0.0950	0.0937	0.0968
R^2	0.6853	0.6733	0.7271	0.7332	0.7092
IOA_1	0.9037	0.8855	0.9131	0.9163	0.9071
IOA_2	0.7302	0.7076	0.7439	0.7486	0.7421
$Kurtosis_{excess}$	2.2452	1.4195	1.9933	2.0172	1.9773
$Skewness$	0.3029	0.54595	0.7728	0.7076	0.8085

Table 2.1: Statistical results on the normalised errors of each prediction model

In comparing the mean values, it can be seen that the models ECM 1 and ALA 2 have a rather small mean error. However this could be the case

of more balanced over- and underestimation. The volatilities in between the two ECM models and the three ALA models are respectively quite similar, but it can be noted that the ECM models are more volatile than the ALA models. The coefficients of determination also don't show many discrepancies, varying within a 6% range between each other. The average R^2 for these models is about 70%, which is not very high.
When comparing the two indices of agreement, it can be seen that the IOA_1 values are closer to 1 than the IOA_2 values. Even though the index ranges in between IOA_1 and IOA_2 respectively are not that high, both criteria suggest ALA 2 as the better performing model and ECM 2 as the worst.
The excess kurtosis is positive in all five models, which indicates a leptokurtic distribution function of the errors, meaning that in comparison to the standard normal distribution there is a higher peak and fatter tails, so the distribution is more clustered towards the mean (resulting in a smaller standard deviation). On the other hand, the calculated (positive) skewness in all models points towards a positively skewed error distribution with a longer right tail.

Next we observe the autocorrelation as well as the partial autocorrelation between the normalised errors to see the lagged dependency of the errors. For errors e_t, e_{t-k} with lag k the autocorrelation function is defined as

$$ACF_e(k) = \frac{cov(e_{t-k}, e_t)}{\sigma_e^2} = \frac{\mathbb{E}[(e_{t-k} - \mu_e)(e_t - \mu_e)]}{\sigma_e^2}$$

where μ_e is the mean and σ_e^2 is the variance of the normalised errors e_t. To estimate the sample autocorrelation function, we use the sample mean $\bar{e}$, the sample size T and sample autocorrelation, so that

$$\widehat{ACF}_e(k) = \frac{\frac{1}{T}\sum_{t=k+1}^{T}(e_t - \bar{e})(e_{t-k} - \bar{e})}{\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^2}.$$

The partial autocorrelation function on the other hand measures the linear dependence between e_t and e_{t-k} after removing the effect of the variables in between $(e_{t-k+1} \dots e_{t-1})$.

Figure 2.3 shows the autocorrelation functions of the normalised errors of the five forecast models up to a lag of 100, which refers to data of approximately a day in 15 minute intervals. The ACF (autocorrelation function) for all models indicates a slow decay. This correlation pattern is however to be expected, as the errors are a result of a multi-step forecast. When investigating autocorrelation for further lags, we can see that significant (positive and negative) correlation values reoccur.

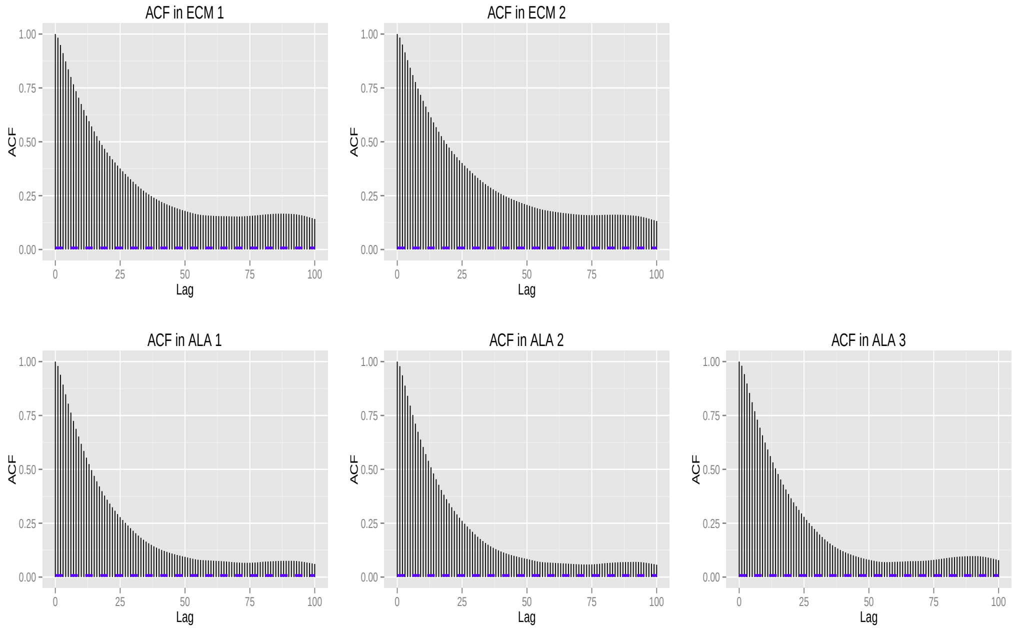

Figure 2.3: Autocorrelation function of the normalised errors for the five models

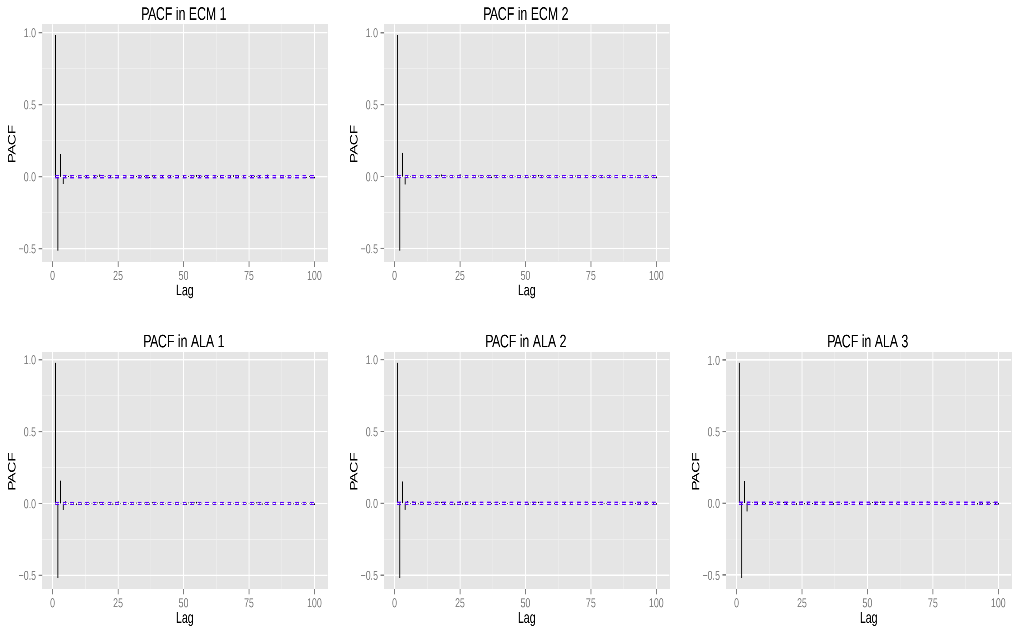

Figure 2.4: Partial autocorrelation function of the normalised errors for the five models

Figure 2.4 on the other hand shows that the partial autocorrelation function decays very quickly. The blue lines in the plots indicate a confidence interval of 95% for the significant levels of the lags, meaning that the (partial) autocorrelation at lag k is significantly different from zero (zero being the case of no autocorrelation) if the value lies outside this confidence interval. The partial autocorrelation function plots for the errors of all five models show a significance up to a lag of 4, which suggests autoregressive behaviour of a fourth order.

2.2 Error analysis on monthly basis

In this section the errors are split into months in order to be able to accentuate patterns and the same analysis as in the previous section is conducted.

When looking at the monthly means, it can be noted that for all ECM and ALA models the highest errors occur during the months of March, April and May. What is also noticeable is that the monthly means of the ALA models are mainly positive, which means that the ALA models tend to underestimate. On the other hand, it seems that ECM 1 tends to overestimate during December in comparison to all the other models.

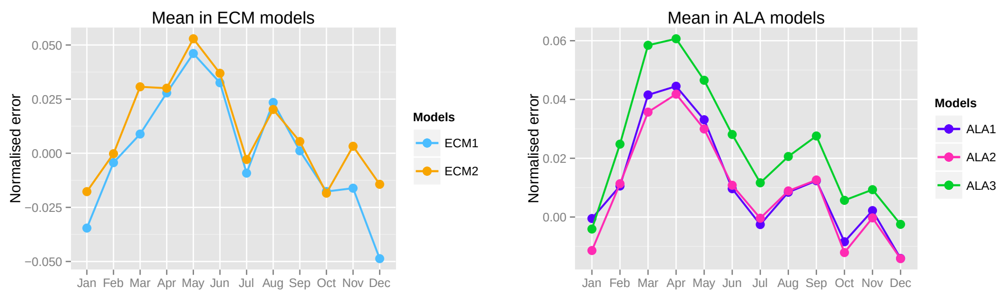

Figure 2.5: Monthly means

The monthly variance of the errors and the mean square errors have a very similar course (though different magnitudes). It can be noted that the ECM models have a relatively high volatility, MSE and MAE (see Figure B.1 and Figure B.2 in appendix) during December in comparison to the other months and also in comparison to the ALA models. On the other hand for the ALA models the volatility and MSE peak is reached during the months of March - May. Generally the ECM models show a slightly higher volatility in comparison to the ALA models (the same applies for the MSE and MAE with exception of April, where the MSE and MAE provided by the ALA models respectively is a bit higher than those provided by the ECM models).

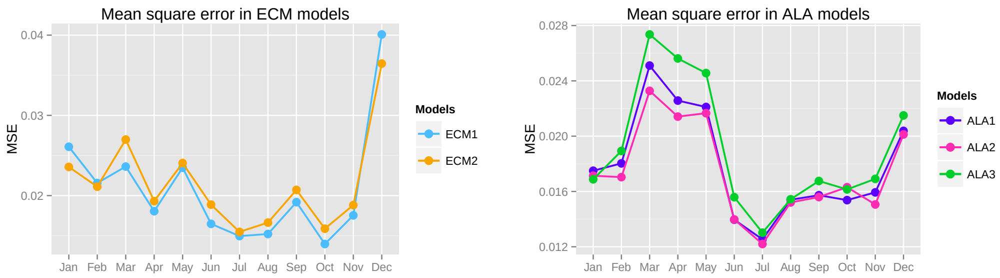

Figure 2.6: Monthly mean square error

Next we look at the coefficients of determination for each month. Generally the monthly coefficients are quite similar amongst the five models, however the ALA models obtain a slightly higher percentage. The months that stand out however are April and December. For April the ECM models offer a noticeably higher coefficient of determination ranging between 70-73%, while the ALA model's coefficients range from 62-67%. During December however, the ECM models indicate an approximately 60% explanation of variance, while the ALA models lie above 75%.

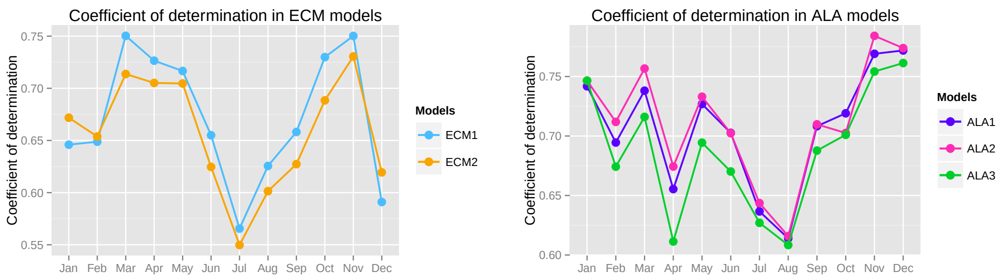

Figure 2.7: Monthly coefficient of determination

The indices of agreement follows the same pattern as the coefficient of determination, where the index values of the ALA models are generally slightly higher than those of the ECM models. The largest differences are again noticed in December, where the ALA models perform significantly better, and in April, where the ECM models show a better performance, especially the ECM 1 model. In August and October ECM 1 also performs slightly better than the ALA models, but not significantly. IOA_2 indicates that ECM 1 also performs better than the rest of the models in May.

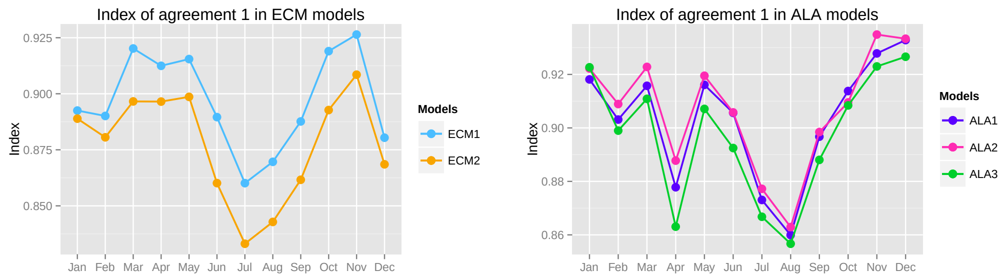

Figure 2.8: Monthly index of agreement 1

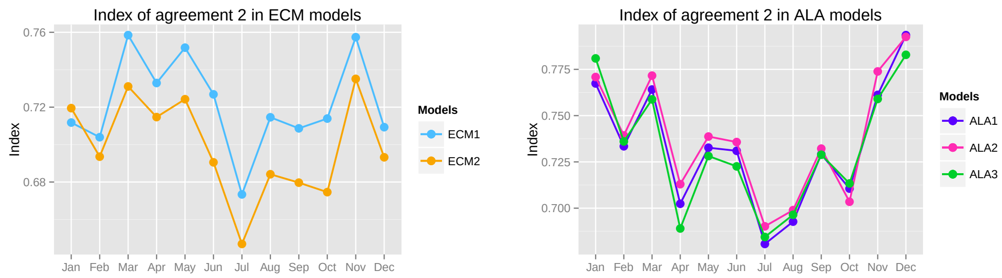

Figure 2.9: Monthly index of agreement 2

General remarks: the ECM 1 model seems to be performing better than the ECM 2 models for most months. An exception show the months of January and December, where the variance, coefficient of determination and the second index of agreement (only in January) are higher for ECM 2. When comparing the ALA models, mainly ALA 2 performs better and ALA 3 worse. For January however, ALA 3 scores higher on both indices of agreement, while the coefficients of determination are almost the same in all three models. For October ALA 1 scores better in regard to variance, R^2 and IOA_1. Another remark is that, with exception of April, the ALA models indicate a better performance than the ECM models.

2.3 Error analysis based on sectors

In this section we will look at the influence the wind front direction has on the five forecasting models. Figure B.3 in the appendix shows the classification in 8 different sections from where the wind is directed (class 0 represents the cases where the classification in any of the 8 sectors was not possible or not made). The scope of this section is to investigate if any of the models shows a different/better performance when wind comes from a specific direction. Similar to the above section we also conduct a statistical analysis with these classes.

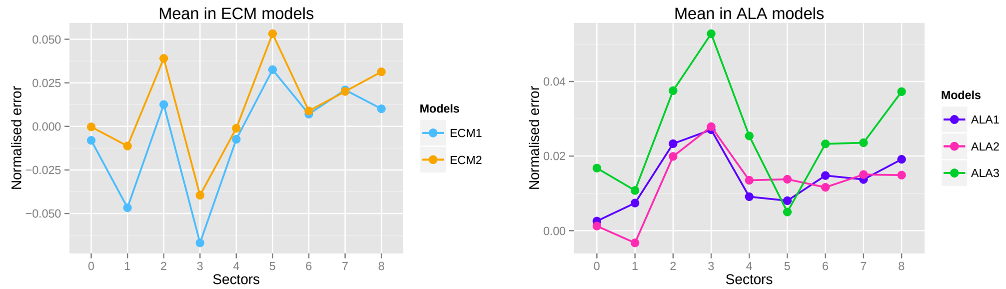

Figure 2.10: Mean error among sectors

The division into sectors again shows that the ALA models tend to underestimate, where ALA 3 shows the highest discrepancies for all sectors with exception of the fifth sector. The ECM 1 obtains the highest errors in sector one and three (overestimation), while the ECM 2 model errors reaches their peak in sectors two, five (underestimation) and thee (overestimation). In comparing the ECM and ALA models between each other, the ECM models generally show higher errors in magnitude.

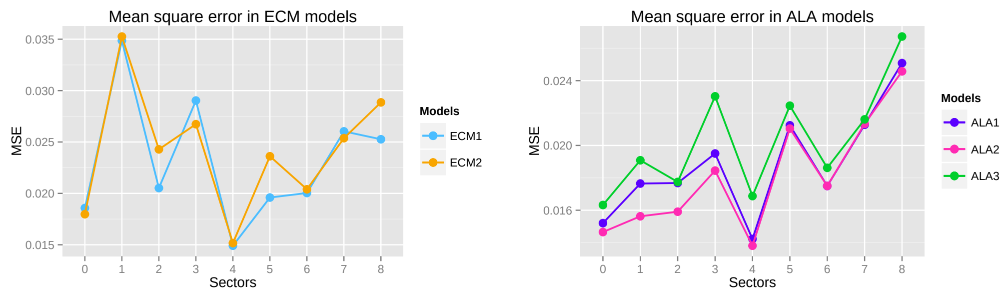

Figure 2.11: Error MSE among sectors

The variance of the errors, MSE and MAE have again a very similiar pattern (Figure B.4 and Figure B.5 showing the variance and MAE can be found in the appendix). The ECM models as well as the ALA models have variances, MSE and MAE values relatively closely to each other within each model group. The ECM 2 model shows a higher variance, MSE and MAE amongst the ECM models in sectors two, five and eight, while the third sector has a higher MSE and MAE value for ECM 1. Amongst the ALA models, ALA 3 ranks higher in values, expecially in sector one and three. ALA 2 on the other hand scores lower than the rest, especially in sectors one and three. In general the ECM models have higher variance, MSE and MAE than the ALA models. An exception in this case is the fifth sector where the ECM 1 model takes the lowest value.

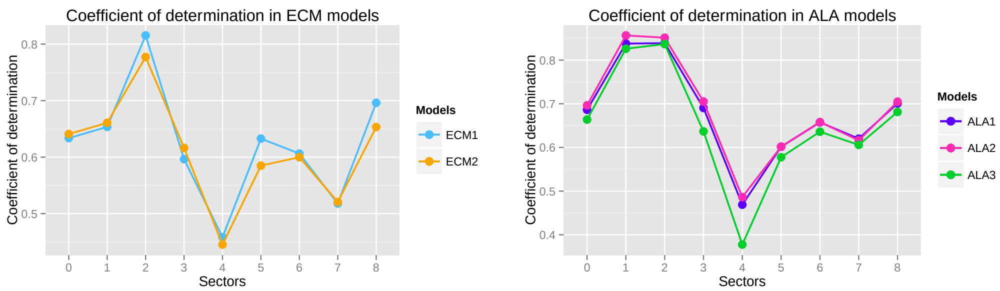

Figure 2.12: Coefficient of determination among sectors

The coefficient of determination amongst the ECM and ALA models respectively is very similiar (ECM 1 has a slightly higher R^2 in sectors two, five and eight than ECM 2). The coefficients of the ALA models are generally higher than those of the ECM model with the exception of sector 5, where ECM 1 obtains the highest value. Sector four shows a very low coefficient of determination, while sector one (only for the ALA models) and two show the highest coefficients (above 80%). The other coefficients lie below 70%.

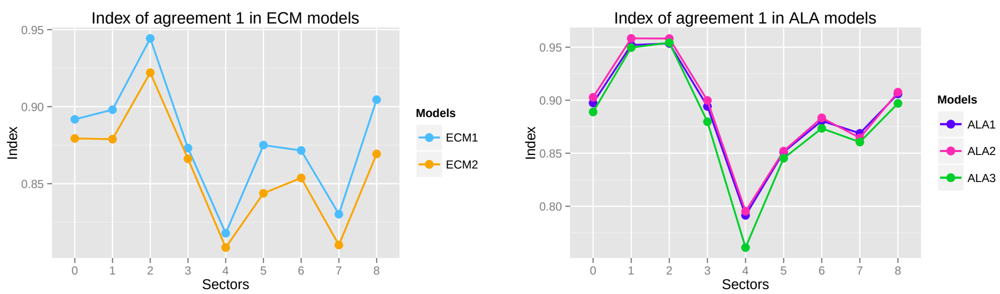

Figure 2.13: Index of agreement 1 among sectors

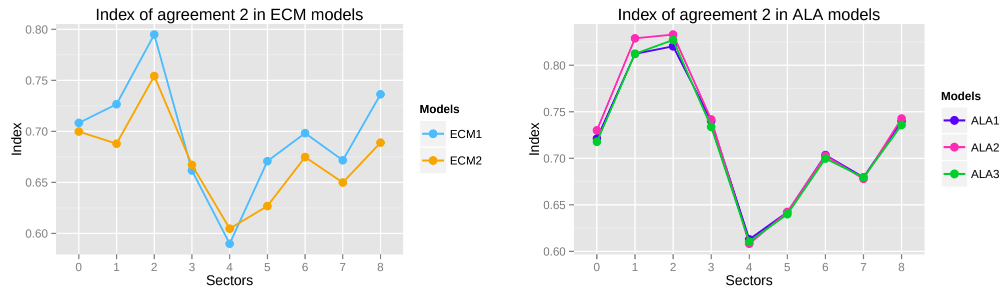

Figure 2.14: Index of agreement 2 among sectors

The indices of agreement (1 and 2) follow a very similiar pattern to that of the coefficient of determination, however the first index of agreement shows higher values on scale. The ALA models (especially ALA 2) have higher indices than the ECM models, with the exception of sectors four (only for IOA_1) and five, where the ECM 1 model shows the leading values. The index of agreement 1 generally takes values above 80%, while the second index of agreement mainly takes values below 75%.

General remarks: also in splitting data according to the wind front directions it seems that the ALA models (especially ALA 2) are generally performing better than the ECM models. However, in sector five the ECM 1 model obtains the highest indices of agreement as well as the highest coefficient of determination. The variance of ECM 1 in the fifth sector is also the lowest of all models. On the other hand, it should be noted that for the considered time frame (01.01.2013 - 31.12.2014) the distribution of data between the sectors is not uniform. The majority of the data is not even specified into a sector (represented by sector 0).

2.4 Error analysis according to day time

Another way we can inspect our data is through investigating if there exist an influence of the time of the day on our forecasting models. Therefore we conduct the same analysis as before on hourly basis.

From what the hourly mean values reveal, the ECM models tend to underestimate power during the hour intervals 0 - 6 and 17 - 23 o'clock and overestimate the rest of the time. The ECM 1 model underestimates less, while ECM 2 overestimates less. The ALA models tend to underestimate for most daytimes, whereas ALA 3 underestimates most and overestimates less.

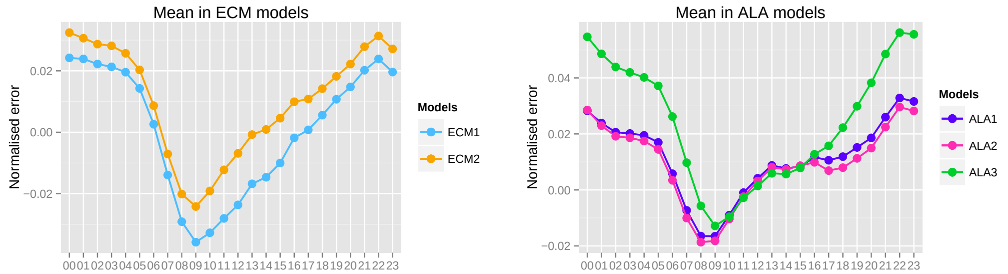

Figure 2.15: Mean of errors according to daytime

As it can be seen from the graphical comparison in figure 2.16, the mean square error of the forecast errors varies according to the daytime. For all time slots the ECM models indicate a higher MSE than the ALA models, where generally ECM 2 obtains the highest MSE and ALA 2 the lowest. Similar patterns and the same results can also be concluded for variance and the mean absolute error (see Figure B.6 and B.7 in the appendix).

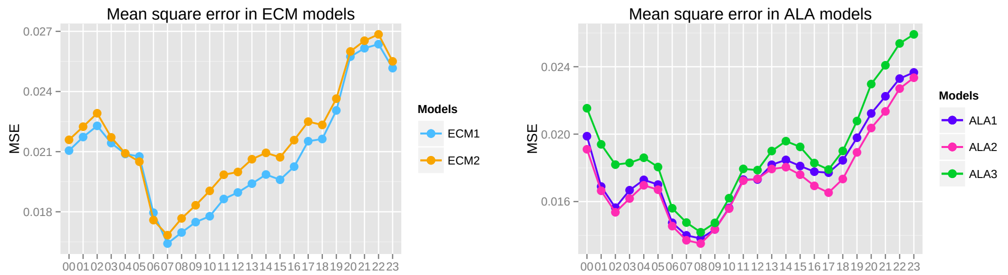

Figure 2.16: MSE according to daytime

Figure 2.17 shows box plots of the errors classified in hours of the day (each hour category includes all four 14-minute intervals contained in that hour, so for example category hour 5 includes all the errors occurring at time 5^{00}, 5^{15}, 5^{30} and 5^{45}). Each category contains many outliers, as it is also observed from the high kurtosis, however in order to have a better view of the box plots many of the outliers are not shown in figure 2.17. The box plots indicate that the normalised errors of all models have a relatively low median compared to the variance. Also it was expected that the estimations are better in the early hours of the day, as those are the nearest estimations as a point in time. However the distribution of the errors doesn't seem to follow this hypothesis, as the errors seem more stable during the morning to midday hours.

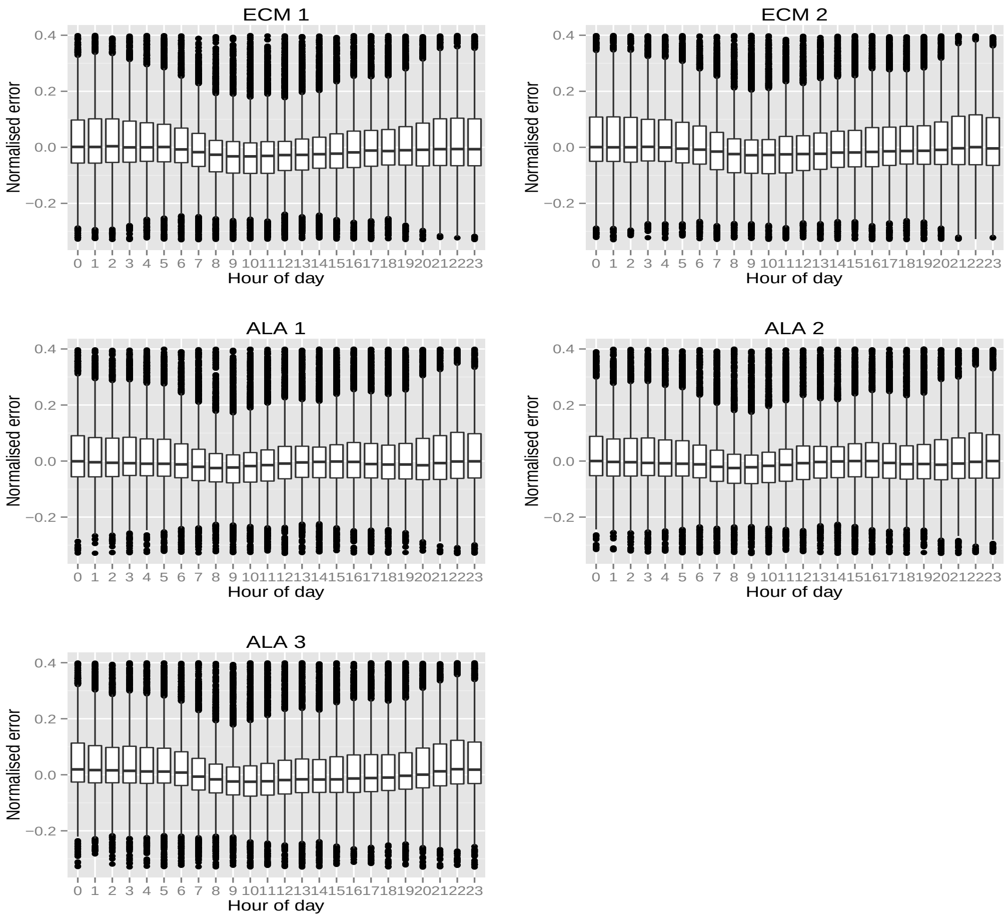

Figure 2.17: Box plots of the normalised errors classified into hour of day

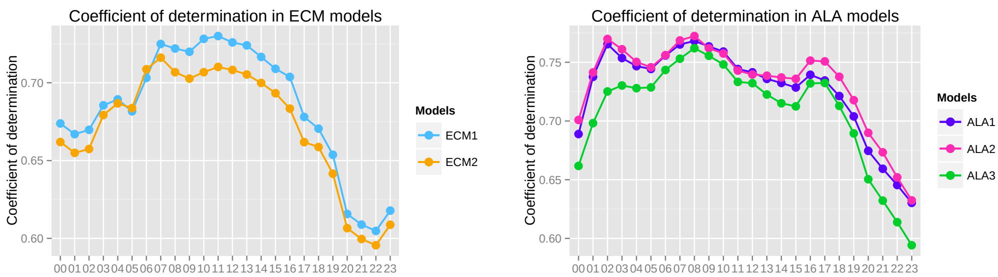

Figure 2.18: Coefficient of determination according to daytime

The coefficient of determination and the first index of agreement show also in this case alike traits. The ECM 1 model take higher values amongst the ECM models for both cases. For the ALA models on the other hand, ALA 2 scores higher. Both the ECM and ALA models have not very high coefficients of determination ranging from approximately between 60-75%. The ALA models though indicate a higher coefficient at each time. The first index of agreement shows relatively high values ranging between 85-95%. Here again the highest indices are obtained from ALA 2, with exception of the 23rd hour, where ECM 1 scores slightly higher.

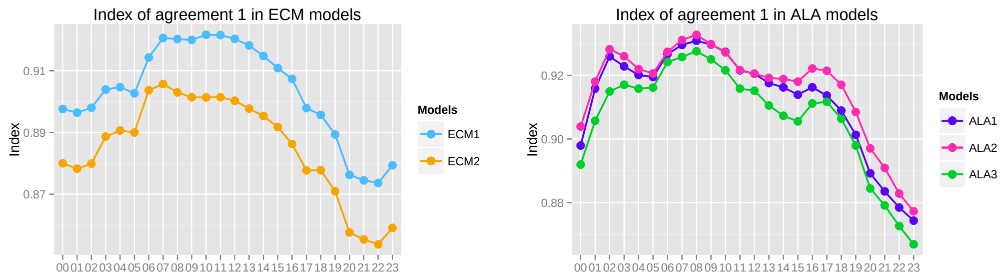

Figure 2.19: Index of agreement 1 according to daytime

The second index of agreement ranges between 66-77%, which is not that high. The indices for the time slots between 2^{00} - 19^{45} are higher for all five models in comparison to the remaining time slots. The ECM models show a similar structure as IOA_1, where ECM 1 scores higher than ECM 2. For the ALA models between 9^{00} - 21^{45} the highest index is that of the ALA 2 model, while for all other times ALA 3 offers the highest index (however with not much discrepancy to the other two indices). Once again the highest indices for all daytimes are scored from the ALA models.

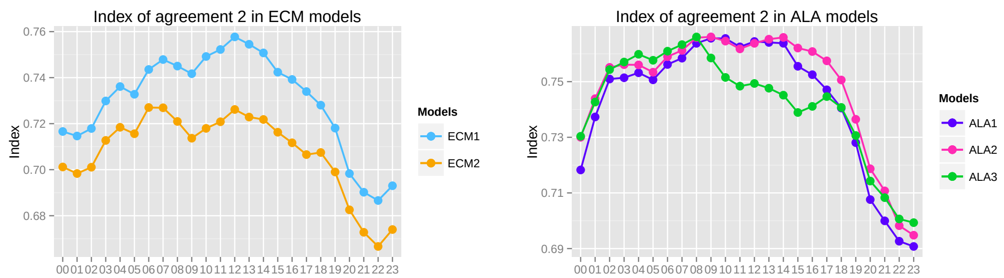

Figure 2.20: Index of agreement 2 according to daytime

General remarks: On an hourly basis we see once more that the ALA models perform better than the ECM models.

2.5 Other statistical tests and error distribution

In this section further tests are applied to review other statistical properties regarding the distribution of the normalised power prediction errors of the five models being investigated. Figure 2.21 shows the structure of the density function for the five model errors. It can be seen that density functions of the two ECM model errors are very similar, whereas the ECM 2 model shows a higher peak than the ECM 1 model. The ALA model errors also have very similar density functions. The ALA 1 and ALA 2 models show an almost identical course, whereas the ALA 3 model shows a higher peak. From the density of the errors of all models it can be seen that they don't follow a normal distribution, as they indicate a skewness and excess kurtosis.

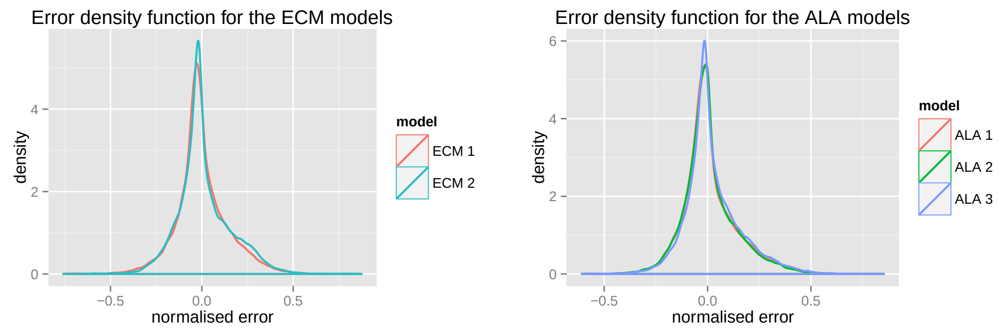

Figure 2.21: Density functions of the normalised errors

However to test the hypothesis whether the errors follow a normal distribution or not, a Kolmogorov Smirnov test was applied to the errors using the function *ks.test()* in the *stats* package in *R*. This test quantifies the distance between the empirical distribution function of the sample (this case the model errors) and the cumulative distribution function of the normal distribution. The test statistic is defined as

$$D_n := \sup_x |F(x) - \hat{F}_n(x)|$$

where n is the number of observations, F is a specified distribution function (the normal distribution function in this case) and $\hat{F}_n$ is the empirical cumulative distribution function of the errors. The null hypothesis

$$H_0: \quad F(x) = \hat{F}_n(x)$$

is then rejected if $\sqrt{n}D_n$ is bigger than a critical value [11]. For the nomalised errors of the five prediction models the null hypothesis of normally distributed errors was rejected in all cases using a significance level of 95% (the p-value was below 5%). A Q-Q (quantile - quantile) plot, which compares the quantiles of the error distributions against those of a normal distribution also indicates that the errors are not normally distributed (refer to Figure 2.22; ideally the points should follow the line if the data were normally distributed).

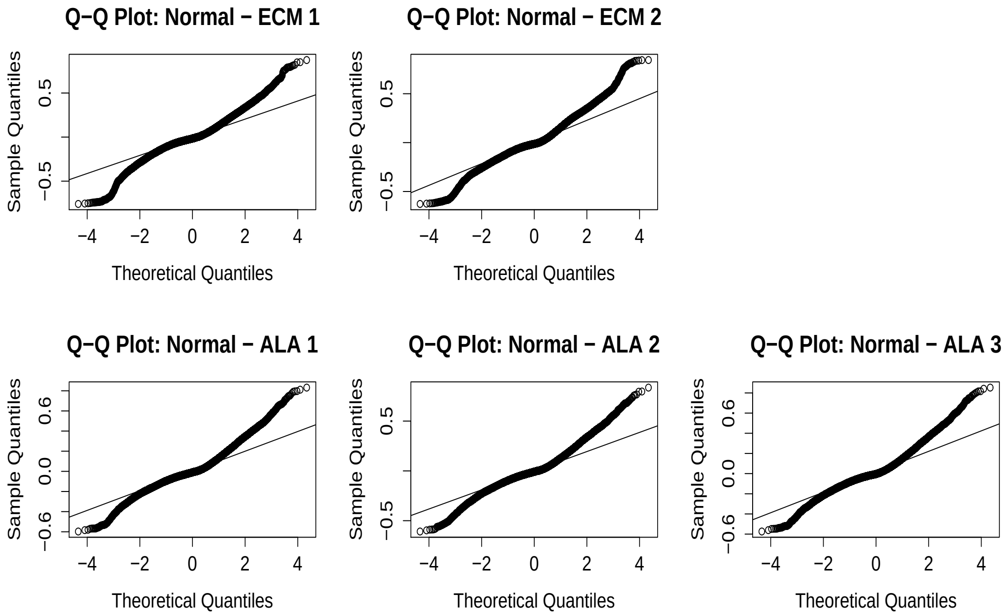

Figure 2.22: Q-Q plots of the normalised errors against the normal distribution

Of interest is also whether the errors of the five models have significantly different means and variances. To test if the means of the errors of the five different models are the same, a t-test from the *stats* package was used, even though the errors are not independent and identically distributed. The p-values were well below 5%, which means that the null hypothesis

$$H_0 : \quad \mu_i = \mu_j, \quad i,j \in \{\text{ECM } 1 \ldots \text{ALA } 3\}, \quad i \neq j$$

was rejected, so the means of the errors in each model are significantly different. One sided t-tests show that

$$\mu_{\text{ALA }3} < \mu_{\text{ALA }1} < \mu_{\text{ALA }2} < \mu_{\text{ECM }2} < \mu_{\text{ECM }1}$$

significantly. In order to see if the variances of the five model errors are the same, a Levene test was performed using the *leveneTest()* function from the *car* package in *R*. The null hypothesis of homoscedasticity between the errors

$$H0: \quad \sigma^2_{\text{ECM 1}} = \cdots = \sigma^2_{\text{ALA 3}}$$

was rejected rejected from the Levene test, meaning that the variance of the errors is significantly different between the models. When comparing the variances of the ECM and ALA models in between them, the results were the same. One sided tests showed that

$$\sigma^2_{\text{ALA 2}} < \sigma^2_{\text{ALA 1}} < \sigma^2_{\text{ALA 3}} < \sigma^2_{\text{ECM 1}} < \sigma^2_{\text{ECM 2}}$$

significantly. The same results as for the variance (in the same order) were also obtained when testing if the mean square error and mean absolute error between the models are the same. Another attempt was made to test for daily homoscedasticity

$$H0: \sigma^2_1 = \cdots = \sigma^2_j = \cdots = \sigma^2_{730}$$

where σ^2_j represents the variance at day $j \in \{1, \ldots, 730\}$ (considering each day of the available data, which covers 2 years), however this null hypothesis was also rejected. A further hypothesis was made to test for daily homoscedasticity within a month, meaning if the daily variance of the errors in a specific month is the same. However also in this case the hypothesis was rejected.

A Levene test was also performed to check if months under the same season have the same variance. It showed that for a confidence level of 5% this hypothesis could not be accepted in neither model. However there were for every model some months with the same variance. The same results can also be said about the daily mean square and mean absolute errors.

CHAPTER 3

(G)ARCH models

In the context of financial markets an important factor is the volatility of the data and how to forecast this volatility. A widely used approach to model and describe periods of changing variance are through (G)ARCH - (General) Autoregressive Conditional Heteroskedasticity - modeling. These are autoregressive models in squared returns, where the conditional variance of the next period is described based on the information given up to this period.

3.1 The ARCH(1) Model

The simplest model to describe conditional heteroscedasticity is the ARCH(1) model. We assume that ε_t is a sequence of independent $N(0,1)$ random variables. In this context an ARCH(1) process a_t is defined as

$$a_t = \sigma_t \varepsilon_t$$

where the conditional variance σ_t^2 is described as a function of past values of the a_t:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2.$$

As a restriction we have $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to ensure positive variance and $\alpha_1 < 1$ for a stationary process. Under the assumptions of an ARCH(1) model we expect high conditional volatility (σ_t) if the residual returns in the previous period (a_{t-1}) is high in magnitude, or low conditional volatility for low returns.

An extension of the ARCH(1) model would be the ARCH(q) model where:

$$a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_q a_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0 \quad \forall i \in \{1, \ldots, q\}$. In this case the required restriction for stationarity is $\sum_{i=1}^{q} \alpha_i < 1$. The variance is given as

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i}$$

(see the following section).

3.2 The GARCH(p,q) model

A further extension (generalisation) of the ARCH model is the GARCH model, which also takes into account the past conditional variances implying the following setting:

$$a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_q a_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2,$$

where $\alpha_0 > 0$, $\alpha_i \geq 0 \ \forall i = \{1, \ldots, q\}$ and $\beta_j \geq 0 \ \forall j = \{1, \ldots, p\}$. It is clear that for $\beta_1 = \cdots = \beta_p = 0$ the GARCH model would be equivalent to an ARCH(q) model. GARCH models have a stationary solution if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. The most important results from this model include the following:

1. Conditional mean and conditional variance of a_t:

$$\begin{aligned}
\mathbb{E}[a_t | a_{t-1}, a_{t-2} \ldots] &= \mathbb{E}[\sigma_t \varepsilon_t | a_{t-1}, \ldots] \\
&= \sigma_t \mathbb{E}[\varepsilon_t | a_{t-1}, \ldots] \\
&= 0
\end{aligned}$$

since $\mathbb{E}[\varepsilon_t | a_{t-1}, \ldots] = 0$, as ε_t is independent of $a_{t-1}, a_{t-2}, \ldots$.

$$\begin{aligned}
\mathbb{V}[a_t | a_{t-1}, a_{t-2} \ldots] &= \mathbb{E}[a_t^2 | a_{t-1}, \ldots] - (\underbrace{\mathbb{E}[a_t | a_{t-1}, \ldots]}_{=0})^2 \\
&= \mathbb{E}[\sigma_t^2 \varepsilon_t^2 | a_{t-1}, \ldots] \\
&= \sigma_t^2 \mathbb{E}[\varepsilon_t^2 | a_{t-1}, \ldots] \\
&= \sigma_t^2
\end{aligned}$$

as $\mathbb{E}[\varepsilon_t^2|a_{t-1},\dots] = \mathbb{V}[\varepsilon_t] = 1$ since $\varepsilon_t \overset{iid}{\sim} N(0,1)$.

2. Unconditional mean and variance of a_t:

$$\begin{aligned}
\mathbb{E}(a_t) &\overset{\text{iterated expectation}}{=} \mathbb{E}[\underbrace{\mathbb{E}[a_t|a_{t-1},a_{t-2},\dots]}_{=0}] = 0 \\
\mathbb{V}[a_t] &= \mathbb{E}[a_t^2] - (\underbrace{\mathbb{E}[a_t]}_{=0})^2 = \mathbb{E}[\sigma_t^2\varepsilon_t^2] \\
&= \mathbb{E}[\sigma_t^2]\underbrace{\mathbb{E}[\varepsilon_t^2]}_{=1} \qquad (\varepsilon_t \text{ is independent of } \sigma_t) \\
&= \alpha_0 + \alpha_1\mathbb{E}[a_{t-1}^2] + \dots + \alpha_q\mathbb{E}[a_{t-q}^2] + \beta_1\mathbb{E}[\sigma_{t-1}^2] + \dots \\
&\quad + \beta_p\mathbb{E}[\sigma_{t-p}^2]
\end{aligned}$$

Under the assumption that a_t is a stationary process we have that

$$\mathbb{E}[a_{t-1}^2] = \dots = \mathbb{E}[a_{t-q}^2] = \mathbb{E}[a_t^2].$$

Furthermore we have

$$\begin{aligned}
\mathbb{E}[\sigma_{t-1}^2] &= \dots = \mathbb{E}[\sigma_{t-p}^2] = \mathbb{E}[\sigma_t^2] = \mathbb{E}[\mathbb{V}[a_t|a_{t-1},\dots]] \\
&= \mathbb{E}[\mathbb{E}[a_t^2|a_{t-1},\dots] - \underbrace{\mathbb{E}[a_t|a_{t-1},\dots]}_{=0}^2] = \mathbb{E}[a_t^2] = \mathbb{V}[a_t].
\end{aligned}$$

Hence the equation above becomes

$$\begin{aligned}
&\mathbb{V}[a_t] = \alpha_0 + (\alpha_1 + \dots + \alpha_q + \beta_1 + \dots + \beta_p)\mathbb{V}[a_t] \\
\Longrightarrow \qquad &\mathbb{V}[a_t] = \frac{\alpha_0}{1 - (\alpha_1 + \dots + \alpha_q + \beta_1 + \dots + \beta_p)}
\end{aligned}$$

where the imposed restriction for stationarity is $\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1$.

3. The covariance function of a_t for a lag $k > 0$ is 0:

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[a_t, a_{t-k}] &= \mathbb{E}[a_t a_{t-k}] \\
&= \mathbb{E}[\mathbb{E}[a_t a_{t-k}|a_{t-1},\dots]] \\
&= \mathbb{E}[a_{t-k}\underbrace{\mathbb{E}[a_t|a_{t-1},\dots]}_{=0}] \\
&= 0 \quad \forall k > 0.
\end{aligned}$$

3.3 ARCH(q) as AR(q)

Under the assumption that the squared residuals a_t^2 are stationary, an ARCH(q) model can also be explained in terms of an autoregressive (AR) model on a_t^2. X_t is generated from an autoregressive model of order q (indicated as AR(q)) when

$$X_t = \omega + \rho_1 X_{t-1} + \rho_2 X_{t-2} + \cdots + \rho_q X_{t-q} + \varepsilon_t$$

where ω is a constant, $\rho_1, \dots \rho_p$ are the parameters of the model and ε_t is a white noise process. Without loss of generality we will show how an ARCH(1) model can be explained as an AR(1) model of the squared residuals (provided they are stationary). By defining the variable v_t so that

$$v_t := a_t^2 - \sigma_t^2,$$

then a_t^2 can be expressed as an autoregressive process of first order (AR(1))

$$a_t^2 = \sigma_t^2 + v_t = \alpha_0 + \alpha_1 a_{t-1}^2 + v_t$$

where v_t fulfills the requirements of a white noise process:

1. $\mathbb{E}[v_t] = 0$:

$$\begin{aligned}\mathbb{E}[v_t] &= \mathbb{E}[\mathbb{E}[v_t|a_{t-1}, \dots]] \\ &= \mathbb{E}[\mathbb{E}[a_t^2 - \sigma_t^2|a_{t-1}, \dots]] \\ &= \mathbb{E}[\underbrace{\mathbb{E}[a_t^2|a_{t-1}, \dots]}_{=\sigma_t^2} - \sigma_t^2] \\ &= 0 \end{aligned} \tag{3.1}$$

2. $\mathbb{V}[v_t] = \sigma_v^2 > 0$:

$$\begin{aligned}\mathbb{V}[v_t] &= \mathbb{E}[v_t^2] - (\underbrace{\mathbb{E}[v_t]}_{=0})^2 \\ &= \mathbb{E}[(\sigma_t^2\varepsilon_t^2 - \sigma_t^2)^2] \\ &= \mathbb{E}[\sigma_t^4(\varepsilon_t^2 - 1)^2] \\ &= \mathbb{E}[\sigma_t^4]\mathbb{E}[(\varepsilon_t^2 - 1)^2] \qquad (\sigma_t \text{ independent of } \varepsilon_t)\end{aligned}$$

As $\varepsilon_t \sim N(0,1)$, we have that ε_t^2 follows a chi squared distribution with one degree of freedom ($\varepsilon_t^2 \sim \chi_1^2$). Therefore $\mathbb{E}[\varepsilon_t^2] = 1$ and $\mathbb{E}[(\varepsilon_t^2 - 1)^2] = \mathbb{V}[\varepsilon_t^2] = 2$.

On the other hand

$$\begin{aligned}
\mathbb{E}[\sigma_t^4] &= \mathbb{E}[(\alpha_0 + \alpha_1 a_{t-1}^2)^2] \\
&= \alpha_0^2 + 2\alpha_0\alpha_1 \mathbb{E}[a_{t-1}^2] + \alpha_1^2 \underbrace{\mathbb{E}[a_{t-1}^4]}_{=\mathbb{E}[a_t^4]} \\
&= \alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0}{1-\alpha_1} + \alpha_1^2 \mathbb{E}[\sigma_t^4]\mathbb{E}[\varepsilon_t^4] \qquad \text{(independence of } \sigma_t \text{ and } \varepsilon_t) \\
&= \alpha_0^2 \frac{1+\alpha_1}{1-\alpha_1} + \alpha_1^2 \mathbb{E}[\sigma_t^4]\,[\underbrace{\mathbb{V}[\varepsilon_t^2]}_{=2} + (\underbrace{\mathbb{E}[\varepsilon_t^2]}_{=1})^2] \\
&= \alpha_0^2 \frac{1+\alpha_1}{1-\alpha_1} + 3\alpha_1^2 \mathbb{E}[\sigma_t^4] \qquad (3.2) \\
\implies \quad \mathbb{E}[\sigma_t^4] &= \alpha_0^2 \frac{1+\alpha_1}{1-\alpha_1} \frac{1}{1-3\alpha_1^2}.
\end{aligned}$$

Therefore:

$$\mathbb{V}[v_t] = 2\alpha_0^2 \frac{1+\alpha_1}{1-\alpha_1} \frac{1}{1-3\alpha_1^2} =: \sigma_v^2 > 0$$

under the sufficient restriction that $\alpha_1^2 < 1/3$.

3. $\mathbb{C}\mathrm{ov}(v_t, v_{t+k}) = 0 \quad \forall k > 0$:

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(v_t, v_{t+k}) &= \mathbb{E}[v_t v_{t+k}] \\
&= \mathbb{E}[\mathbb{E}[v_t v_{t+k} | a_{t+k-1}, \dots]] \\
&= \mathbb{E}[v_t \underbrace{\mathbb{E}[v_{t+k} | a_{t+k-1}, \dots]}_{=0}] \\
&= 0 \qquad \forall k > 0. \qquad (3.3)
\end{aligned}$$

This means that all the results of an AR(1) model can be applied to a_t^2. Using the AR(1) correlation function for a lag k we then have

$$\rho_{a_t^2}(k) = \alpha_1^{|k|} \quad \forall k.$$

3.4 GARCH(p,q) as ARMA(p,q)

In a similiar way as in section 3.3, also GARCH(p,q) models can be represented in terms of an ARMA(p,q) model with respect to the squared residuals

a_t^2, given that a_t^2 is stationary and $\mathbb{E}(a_t^4) < \infty$. Assuming w.l.o.g. that $p = q$ and together with

$$v_t := a_t^2 - \sigma_t^2$$

we have:

$$\begin{aligned}
a_t^2 &= \sigma_t^2 + v_t \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_{t-p} a_{t-p}^2 + \beta_1 \underbrace{\sigma_{t-1}^2}_{=a_{t-1}^2 - v_{t-1}} + \dots + \beta_p \underbrace{\sigma_{t-p}^2}_{=a_{t-p}^2 - v_{t-p}} + v_t \\
&= \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + \dots + (\alpha_p + \beta_p) a_{t-p}^2 - \beta_1 v_{t-1} - \dots - \beta_p v_{t-p} + v_t,
\end{aligned}$$

where the term v_t is a white noise given that $\mathbb{E}(a_t^4) < \infty$ [10]. The proof is analogue to equations (3.1) and (3.3).

3.5 Forecasting

Assuming we know the a_t and σ_t^2 values up to time $t-1$ we can calculate the GARCH(p,q) conditional variance at time t using

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_q a_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2.$$

The next period's conditional variance forecast for GARCH(p,q) models (2 step ahead forecast in terms of a_t) can be calculated recursively the following way:

$$\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \mathbb{E}[\sigma_{t+1}^2 | a_{t-1}, \dots] \\
&= \alpha_0 + \alpha_1 \underbrace{\mathbb{E}[a_t^2 | a_{t-1}, \dots]}_{=\sigma_t^2} + \dots + \alpha_q a_{t-q+1}^2 + \beta_1 \sigma_t^2 + \dots + \beta_p \sigma_{t-p+1}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_2 a_{t-1}^2 \cdots + \alpha_q a_{t-q+1}^2 + \beta_2 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p+1}^2
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_{t+2}^2 &= \mathbb{E}[\sigma_{t+2}^2 | a_{t-1}, \dots] \\
&= \alpha_0 + \alpha_1 \underbrace{\mathbb{E}[a_{t+1}^2 | a_{t-1}, \dots]}_{=\hat{\sigma}_{t+1}^2} + \alpha_2 \sigma_t^2 + \dots + \alpha_q a_{t-q+2}^2 + \beta_1 \hat{\sigma}_{t+1}^2 + \dots \\
&\quad + \beta_p \sigma_{t-p+2}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_{t+1}^2 + (\alpha_2 + \beta_2)\sigma_t^2 + \alpha_3 a_{t-1}^2 + \dots + \alpha_q a_{t-q+2}^2 + \\
&\quad \beta_3 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p+2}^2 \\
&\vdots \\
\hat{\sigma}_{t+k}^2 &= \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_{t+k-1}^2 + (\alpha_2 + \beta_2)\hat{\sigma}_{t+k-2}^2 + \dots + \\
&\quad (\alpha_{\max(q,p)} + \beta_{\max(q,p)})\hat{\sigma}_{t+k-\max(q,p)}^2,
\end{aligned}$$

where for $p > q$ we have $\alpha_i = 0$ for $i = q+1, \ldots, p$ and for $q > p$ we have $\beta_i = 0$ for $i = p+1, \ldots, q$.

The conditional variance forecast for the GARCH(1,1) model though can be expressed in a nicer representation. Defining

$$\begin{aligned}
&\sigma^2 := \mathbb{E}[\sigma_t^2] = \mathbb{E}[\mathbb{E}[a_t^2|a_{t-1},\ldots]] = \mathbb{E}[a_t^2] = \mathbb{V}[a_t] = \frac{\alpha_0}{1-(\alpha_1+\beta_1)} \\
\Longrightarrow \qquad &\alpha_0 \overset{(*)}{=} \sigma^2(1-(\alpha_1+\beta_1))
\end{aligned}$$

we have:

$$\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \alpha_0 + \alpha_1 \underbrace{\mathbb{E}[a_t^2|a_{t-1},\ldots]}_{=\sigma_t^2} + \beta_1\sigma_t^2 = \alpha_0 + (\alpha_1+\beta_1)\sigma_t^2 \\
&\overset{(*)}{=} \sigma^2(1-(\alpha_1+\beta_1)) + (\alpha_1+\beta_1)\sigma_t^2 \\
&= \sigma^2 + (\alpha_1+\beta_1)(\sigma_t^2 - \sigma^2) \\
\\
\hat{\sigma}_{t+2}^2 &= \alpha_0 + \alpha_1 \underbrace{\mathbb{E}[a_{t+1}^2|a_{t-1},\ldots]}_{=\hat{\sigma}_{t+1}^2} + \beta_1\hat{\sigma}_{t+1}^2 = \alpha_0 + (\alpha_1+\beta_1)\hat{\sigma}_{t+1}^2 \\
&= \sigma^2(1-(\alpha_1+\beta_1)) + (\alpha_1+\beta_1)\hat{\sigma}_{t+1}^2 = \sigma^2 + (\alpha_1+\beta_1)(\hat{\sigma}_{t+1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha_1+\beta_1)[\sigma^2 + (\alpha_1+\beta_1)(\sigma_t^2-\sigma^2) - \sigma^2] \\
&= \sigma^2 + (\alpha_1+\beta_1)^2(\sigma_t^2-\sigma^2) \\
&\vdots \\
\hat{\sigma}_{t+k}^2 &= \sigma^2 + (\alpha_1+\beta_1)^k(\sigma_t^2-\sigma^2).
\end{aligned}$$

3.6 Parameter estimation

The estimation of the parameters of the GARCH model is obtained through the maximization of the likelihood function with respect to the parameters. The likelihood function $L(\cdot)$ is a joint probability density function of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$[1] given the data $A_1, \ldots, A_n$.

For an iid sample $A_1, \ldots, A_n$ let f represent their joint density function with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$. Given the initial data of a GARCH(p,q) model ($\boldsymbol{a_0}$,

[1] $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors containing the parameters $(\alpha_1, \ldots, \alpha_q)$ and $(\beta_1, \ldots, \beta_p)$ respectively

which includes the initial q values of a_t and $\boldsymbol{\sigma_0}$, which includes the initial p values of the conditional variance) and with

$$\varepsilon_t = \frac{a_t}{\sigma_t},$$

the likelihood function can be written as

$$\begin{aligned}
L(\boldsymbol{\alpha}, \boldsymbol{\beta}|A_n, \dots, A_1) &= f_{A_1,\dots,A_n|\boldsymbol{a_0},\boldsymbol{\sigma_0}}(a_1, \dots, a_n) \\
&= f_{A_2,\dots,A_n|A_1=a_1,\boldsymbol{a_0},\boldsymbol{\sigma_0}}(a_2, \dots, a_n) f_{A_1|\boldsymbol{a_0},\boldsymbol{\sigma_0}}(a_1) \\
&= \dots = \prod_{t=1}^{n} f_{A_t|A_{t-1}=a_{t-1},\dots,A_1=a_1,\boldsymbol{a_0},\boldsymbol{\sigma_0}}(a_t) \\
&= \prod_{t=1}^{n} \frac{1}{\sigma_t} f_{E_t}(\varepsilon_t) \qquad (\varepsilon_t \text{ are independent of each other})
\end{aligned}$$

where f_{E_t} indicates the density function of ε_t. The log-likelihood function (logathithmic value of the likelihood function) then becomes

$$\ln(L(\boldsymbol{\alpha}, \boldsymbol{\beta}|A_n, \dots, A_1)) = \sum_{t=1}^{n} \ln(f_{E_t}(\varepsilon_t)) - \ln(\sigma_t).$$

For $\varepsilon_t \sim N(0,1)$ we have $f_{E_t}(\varepsilon_t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\varepsilon_t^2}{2})$, therefore the log likelihood function can be expressed as

$$\begin{aligned}
\ln(L(\boldsymbol{\alpha}, \boldsymbol{\beta}|A_n, \dots, A_1)) &= \sum_{t=1}^{n} \left[-\frac{\varepsilon_t^2}{2} - \ln(\sqrt{2\pi}) - \ln(\sigma_t) \right] \\
&= \sum_{t=1}^{n} \left[-\frac{1}{2}(\varepsilon_t^2 + \ln(\sigma_t^2)) - \ln(\sqrt{2\pi}) \right].
\end{aligned}$$

A problem is that the initial conditional variances are unknown, so initial values are assigned. However, in the long run, if the time series is long enough, the estimation of the initial conditional variance will not have a major impact [8]. The rest of the σ_t are calculated as specified from the order of the GARCH model.

3.7 Application of GARCH models on data

3.7.1 Data settings

This section includes the validation of different (G)ARCH models applied to the normalised errors of wind energy prediction for the five prediction models used by Austrian Power Grid. It has to be noted that these errors are correlated with each other, although the GARCH model assumes no correlation. However this issue will be neglected to avoid further complexity. The normalised errors of every model are tested for stationarity. A Kwiatkowski-Phillips-Schmidt-Shin test was done on the errors using the *kpss.test()* function in the *tseries* package in *R*. This tests for the null hypothesis that the series is stationary against the alternative hypothesis of a unit root. The tests rejected the null hypothesis of stationarity for the models with the exception of ECM 2.

The evaluation of the GARCH models throughout this thesis is done using the "Rugarch" package in R. The function *ugarchgspec*() is used to specify the type of GARCH modeling. Eg. for a GARCH(1,1) model with normaly distributed errors we have

> ugarchspec(variance.model = list(garchOrder = c(1,1)), distribution= "norm", mean.model = list(armaOrder=c(0,0), include.mean=F))

The fitting of model is then done through the function $ugarchfit()$ using the specification model and the data that is to be fitted (in this case the normalised error increments a_t):

> ugarchfit(ugarchspec(), data)

For the estimation the condition of stationarity is imposed. The function $ugarchfit()$ does the estimation of the coefficients through maximizing the likelihood function using the augmented Lagrange solver 'solnp'. As an initial variance this algorithm calculates a feasible set of starting points and then estimates the model parameters [3].

3.7.2 Model selection

In this section the normalised errors are evaluated through ARCH and GARCH models of different orders. Specifically the models ARCH(1), ARCH(2), ARCH(3), GARCH(1,1), GARCH(1,2), GARCH(2,1) and GARCH(2,2) will be evaluated for each model and their Akaike information criteria (AIC) will be compared in order to specify the order of the (G)ARCH model. The AIC is defined as

$$AIC = \frac{2k - 2\ln(L)}{T},$$

where k is the number of parameters to be estimated in the model, L represents the value of the likelihood function and T is the length of the time series. This criteria makes a trade off between the goodnes of fit of the model against its complexity (the number of parameters used) in order to enable a model selection. Table 3.1 shows the AIC values for the above mentioned (G)ARCH models for all five prediction models. The model with the lowest AIC is chosen.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
ARCH(1)	-2.35236	-2.31497	-2.51221	-2.53149	-2.52092
ARCH(2)	-2.36693	-2.33497	-2.53492	-2.56018	-2.56125
ARCH(3)	-2.37639	-2.34296	-2.55105	-2.57447	-2.57318
GARCH(1,1)	-2.37182	-2.34247	-2.54398	-2.56834	-2.57001
GARCH(1,2)	-2.37179	-2.34243	-2.54395	-2.56831	-2.56999
GARCH(2,1)	**-2.38073**	**-2.35041**	**-2.56206**	**-2.58433**	**-2.58184**
GARCH(2,2)	-2.38070	-2.35038	-2.56204	-2.58430	-2.58181

Table 3.1: Akaike information criterion on the different GARCH models

The numbers in bold in table 3.1 show which (G)ARCH model has the least AIC value for each of the models, meaning which model should be selected. In this case the GARCH(2,1) model was selected for the errors of each model. Therefore the normalised errors e_t of each of the prediction models will be described as

$$e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2.$$

Table 3.2 shows the estimated coefficients of the GARCH(2,1) model for each of the normalised prediction errors. As it can be seen the sum of the coefficients in every model is close to 1, which is the limit for stationarity.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
α_0	0.00016	0.00012	0.00008	0.00009	0.00009
α_1	0.88652	0.82395	0.83816	0.84173	0.84130
β_1	0.01654	0.06275	0.01442	0.01840	0.04157
β_2	0.09594	0.11230	0.14641	0.13887	0.11613

Table 3.2: GARCH(2,1) model coefficient estimations for the normalised errors of each prediction model

To visualise the obtained results, we will build a confidence interval for the power prediction using the estimated variance from the GARCH(2,1) models. Assuming the mean value of the normalised errors is 0 and with their estimated conditional standard deviation $\hat{\sigma}_t$ we build the confidence intervals

$$[\,\hat{p}_t - I_t \cdot \hat{\sigma}_t,\ \hat{p}_t + I_t \cdot \hat{\sigma}_t\,] \tag{3.4}$$

where $\hat{p}_t$ is the predicted power at time t for each of the models and I_t is the installed power at time t [2]. As a final confidence interval we will take the union of the intervals (3.4) for all the models:

$$[\,\min_i\{\hat{p}_{it} - I_t \cdot \hat{\sigma}_{it},\ \max_i\{\hat{p}_{it} + I_t \cdot \hat{\sigma}_{it}\}\,] \tag{3.5}$$

where $\hat{p}_{it}$ is the power predicted from model $i \in \{$ EMC 1, EMC 2, ALA 1, ALA 2, ALA 3 $\}$ at time t and $\hat{\sigma}_{it}$ is the conditional standard deviation of the normalised errors of model i at time t ($\hat{\sigma}_{it}$ is obtained from a 1-step calculation of the GARCH(2,1) models given the information up to period $t-1$).

Figure 3.1 shows the measured power for a random time interval marked with a black line and the confidence interval (3.4) provided by the ALA 2 model marked with a light blue shade.

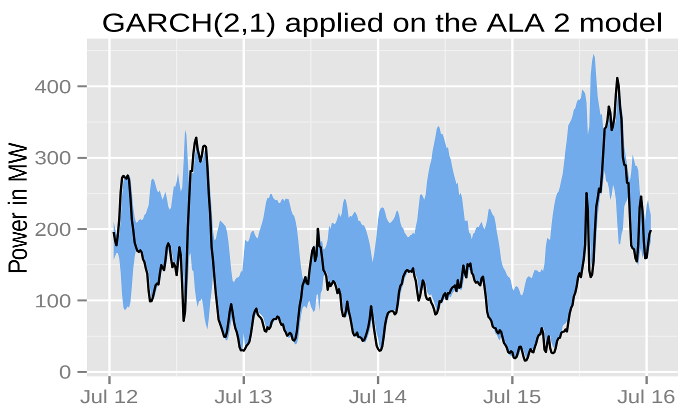

Figure 3.1: Confidence interval $[\,\hat{p}_{\text{ALA 2},t} - I_t \cdot \hat{\sigma}_{\text{ALA 2},t},\ \hat{p}_{\text{ALA 2},t} + I_t \cdot \hat{\sigma}_{\text{ALA 2},t}\,]$ provided from the ALA 2 model marked in a blue shade and the measured power produced marked with a black line.

[2]For the GARCH modeling the normalised errors have been used (normalised through the installed power - refer to chapter 2), therefore to build the confidence interval the errors need to be scaled back.

On the other hand, figure 3.2 shows the measured power for a random time interval and the confidence interval (3.5) around it, as provided by the union of the confidence intervals of all five models. As it can already be seen from this graphic, most of the produced power lies within the introduced confidence interval, in fact it covers 82.8% of the power measured.

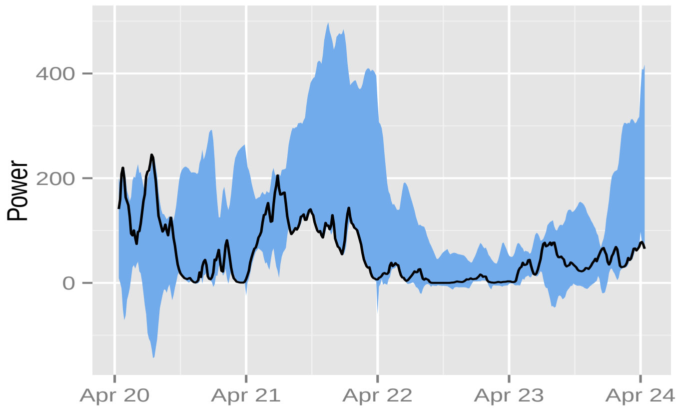

Figure 3.2: Confidence interval $[\,\min_i\{\hat{p}_{it} - I_t \cdot \hat{\sigma}_{it},\ \max_i\{\hat{p}_{it} + I_t \cdot \hat{\sigma}_{it}\}\,]$ marked in a blue shade and the measured power produced marked with a black line.

However, by adding a factor 1.17 to the confidence interval so that

$$[\,\min\{\hat{p}_{it} - 1.17 \cdot I_t \cdot \hat{\sigma}_{it},\ \max\{\hat{p}_{it} + 1.17 \cdot I_t \cdot \hat{\sigma}_{it}\}\,] \tag{3.6}$$

the interval (3.6) covers 95% of the measured power data.

3.7.3 Testing the fit

The values fitted from the GARCH(2,1) modeling show quite good results as a confidence interval. However to see how this model really performs a cross validation will be done:

1. the data containing the normalised errors e_{it}, $t = 1, \dots, T$ is split into a training data set with size T_1 (e_{it}, $t = 1, \dots, T_1$) and a test data set with size $T - T_1$ (e_{it}, $t = T_1 + 1, \dots, T$)

2. using only the data from the training set, a GARCH(2,1) model is estimated and the parameters of the model specified

3. using these estimated parameters a forecast of the conditional variance is done on the test data set

4. an evaluation of the performance of the GARCH model is done by seeing what percentage of the power measured lies within the interval

$$[\min_i\{\hat{p}_{it} - 1.17 \cdot I_t \cdot \hat{\sigma}_{it},\ \max_i\{\hat{p}_{it} + 1.17 \cdot I_t \cdot \hat{\sigma}_{it}\}\,] \tag{3.7}$$

where $\hat{\sigma}_{it}$ is the forecast of the test data set.

Assuming that the forecast is done each day at 8 o'clock for the day ahead, we need a forecasting horizon of 16 - 40 hours ahead. However as the data is measured in 15 minute intervals, the needed forecast horizon is 64-159 steps ahead.

The results of this forecasts are shown in figure 3.3. The x-axis shows the number of days used for the training set, while the y-axis shows the percentage of the measured power data in the test set that lies in the confidence interval (3.7). The percentages range between 93%-95.5% for each test set, which is quite a good result.

Figure 3.4 on the other hand shows the average range of the confidence interval constructed with the forecasted values of each test test. As it can be seen the high percentage coverage of the power data is due to the fact that the intervals are quite wide.

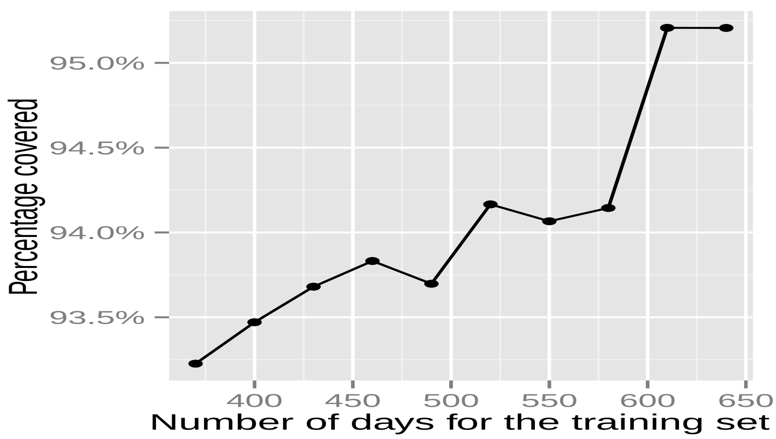

Figure 3.3: Cross validation with different training sets on the confidence interval
$[\min_i\{\hat{p}_{it}-1.17\cdot I_t\cdot\hat{\sigma}_{it},\ \max_i\{\hat{p}_{it}+1.17\cdot I_t\cdot\hat{\sigma}_{it}\}\,]$ which shows what percentage of the data is covered from the test set.

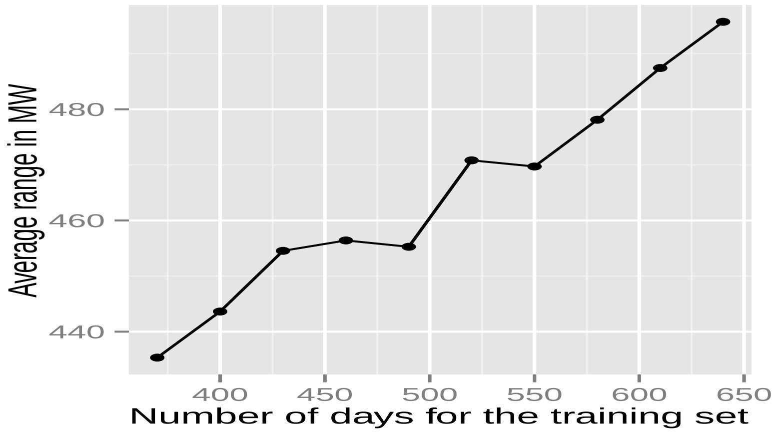

Figure 3.4: The average range of the confidence interval $[\,\min_i\{\hat{p}_{it} - I_t \cdot \hat{\sigma}_{it},\ \max_i\{\hat{p}_{it} + I_t \cdot \hat{\sigma}_{it}\}\,]$ for the different test sets.

Figure 3.5 shows a comparison between the sample variance of the ALA 2 model errors (black line) and the forecasted conditional variances for that model (blue line) as calculated from the 64-159 step-ahead forecast using the GARCH(2,1) model[3]. It can be seen that each of the day-ahead forecasts tends to converge towards the sample variance as the forecasting horizon increases (meaning with each additional step-ahead).

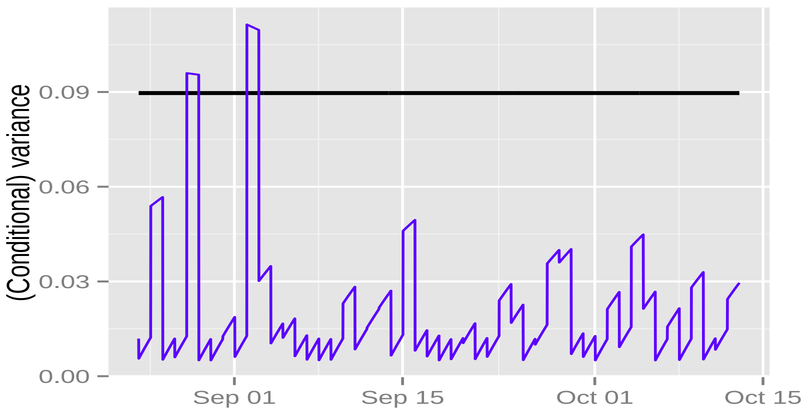

Figure 3.5: Comparison between the conditional and unconditional variance (marked in blue and black respectively) of the errors of the ALA 2 model as calculated by the GARCH(2,1) model using a training set of 490 days.

[3]The training set used for this figure consisted of 490 days.

3.7.4 GARCH model on hourly basis

Another attempt was made to forecast the conditional variance on an hourly basis, so that the forecasts made are only 16-40 steps ahead. For this purpose a GARCH model was applied to a subset of the data, which included only the errors of every hour at the hour (so only the first of the four 15-minute intervals of each hour). In comparing the AIC values of the different (G)ARCH models (ARCH(1), ARCH(2), ARCH(3), GARCH(1,1), GARCH(1,2), GARCH(2,1) and GARCH(2,2)), also in this case the lowest AIC value was provided by the GARCH(2,1) model. The estimated coefficients of the GARCH(2,1) modeling for each of the five errors are shown in table 3.3.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
α_0	0.00146	0.00130	0.00103	0.00099	0.00104
α_1	0.79357	0.77151	0.74817	0.73345	0.75879
β_1	0.12077	0.12244	0.13018	0.12056	0.10870
β_2	0.08466	0.10505	0.12065	0.14499	0.13151

Table 3.3: GARCH(2,1) model coefficient estimations for the normalised errors of each prediction model (applied on an hour basis)

Also in this case we will provide a confidence interval similar to (3.4). In this case the conditional variance of each hour provided from the GARCH model will be used for the three consecutive 15-minute intervals included in that hour (e.g. the estimated $\hat{\sigma}_t$ for 10^{00} is also used for 10^{15}, 10^{30} and 10^{45}). The confidence interval is then

$$[\,\min_i\{\hat{p}_{ith_q} - I_t \cdot \hat{\sigma}_{ith},\ \max_i\{\hat{p}_{ith_q} + I_t \cdot \hat{\sigma}_{ith}\}\,] \tag{3.8}$$

where $\hat{p}_{ith_q}$ is the predicted power of model i at day t at hour h and quarter q, I_t the installed power and $\hat{\sigma}^2_{ith}$ the conditional variance of model i at day t at hour h, as provided by the GARCH model. The interval (3.8) is wide enough to cover 87% of the measured power data. However by adding a factor $c = 1.38$, the confidence interval

$$[\,\min_i\{\hat{p}_{ith_q} - 1.38 \cdot I_t \cdot \hat{\sigma}_{ith},\ \max_i\{\hat{p}_{ith_q} + 1.38 \cdot I_t \cdot \hat{\sigma}_{ith}\}\,] \tag{3.9}$$

covers 95% of the measured power (what we opt for). Also for this model a cross validation will be done, similarly to the one mentioned in the previous section. Figure 3.6 and figure 3.7 show the results of this cross validation. As it can be seen the percentage of the power data covered from every test set is well above 95%, however the average range of these confidence intervals is mainly above 600 MW, which is a lot more than the average range of the confidence interval (3.7). A possible reason for these intervals being a lot wider compared to the confidence intervals validated in section 3.7.3 can be due to the estimated GARCH model parameters being higher (compare table 3.3 and table 3.2, especially the parameters α_0 and β_1).

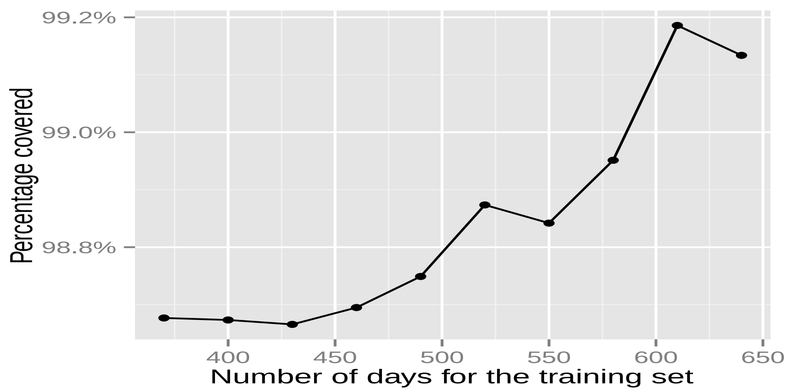

Figure 3.6: Cross validation on the confidence interval
$[\,\min_i\{\hat{p}_{ith_q} - 1.38 \cdot I_t \cdot \hat{\sigma}_{ith},\ \max_i\{\hat{p}_{ith_q} + 1.38 \cdot I_t \cdot \hat{\sigma}_{ith}\}\,]$ which shows what percentage of the data is covered from the test set.

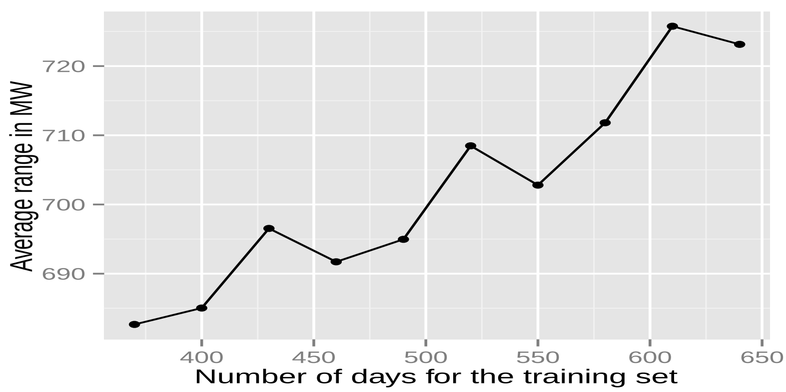

Figure 3.7: The average range of the confidence interval
$[\,\min_i\{\hat{p}_{ith_q} - 1.38 \cdot I_t \cdot \hat{\sigma}_{ith},\ \max_i\{\hat{p}_{ith_q} + 1.38 \cdot I_t \cdot \hat{\sigma}_{ith}\}\,]$ for the different test sets.

CHAPTER 4

Analysis of the daily mean square and mean absolute errors

In this chapter we will analyse the daily mean square errors (MSE) and daily mean absolute errors (MAE) of the five power prediction models. The scope is to see how different factors, such as the meteorological data, affect the daily MSE and MAE as well as to identify any occurring patterns.

4.1 Variables effecting the daily mean square and absolute errors

For the scope of this thesis the five prediction models are not identified, however it is known which meteorological predictions are used as variables in the different power prediction models. The three meteorological variables that are used in all power prediction models are the temperature in Andau region as well as the wind speed and direction in Neusiedler See. The ALA 2 model also uses the wind speed and wind direction in Andau, while the ECM 1 model additionally uses the wind speed and direction in Zwerndorf and Unterlaa. The power prediction models use predicted weather data, whereas the ECM and ALA models obtain different predictions for the wind speed and direction. However the temperature predictions are identical. As the power prediction models are fed with predicted weather data, it is to be expected that the power prediction errors are bigger when the weather prediction errors are bigger. To evaluate this statement we will look at the relationship between the daily mean square/absolute error of the power estimation against the daily mean square/absolute error of the weather estimations. The measured weather data is given in 10-minute intervals, while the weather prediction data is given hourly. In this case to estimate the errors, both data was interpolated quarter hourly in order to match with the intervals of the power measurement and estimation and the daily MSE/MAE was calculated from this data[1]. Figures 4.1 and 4.2 show the relationship between the daily MSE and MAE of the power prediction from the ALA 2 model compared to the daily MSE/MAE of the wind speed in Neusiedler See and the temperature in Andau (the results for the other models are similar). While the power prediction errors seems to increase with higher the wind speed prediction errors, while the relationship between the temperature and power errors seems to be more random.

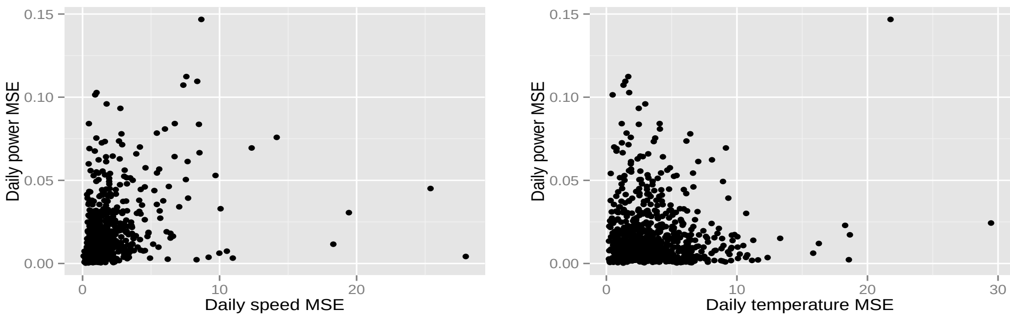

Figure 4.1: Relationship between daily mean square errors of power the power prediction from the ALA 2 model and wind speed at Neusidler See prediction and temperature at Andau prediction respectively.

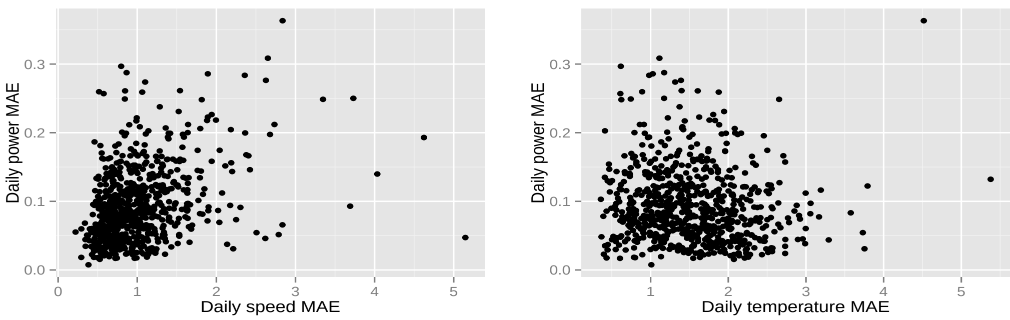

Figure 4.2: Relationship between daily mean absolute errors of power the power prediction from the ALA 2 model and wind speed at Neusidler See prediction and temperature at Andau prediction respectively.

[1]The predicted wind speed and wind directions are given for two different altitudes (at 10m as well as at 100m). For the following results the estimations at 10m height were used

As next it will be investigated whether the daily MSE and MAE of the power prediction models depends on the actual daily energy production. It is expected that when the actual energy production is higher, there is more room for errors and hence the daily MSE and MAE should be higher. Figure 4.3 and figure 4.4 shows this relationship for the daily MSE and MAE respectivly, whereas the daily power production is categorised in 10 GW intervals. The figures generally indicate a positive relationship between these two variables, meaning that on the days where more power is produced, the estimation errors are also higher. An exception to this trend makes the category of 50-60 GW, where the daily MSE and MAE show lower values, which could mean that the forecast is more stable for these levels of energy produced.

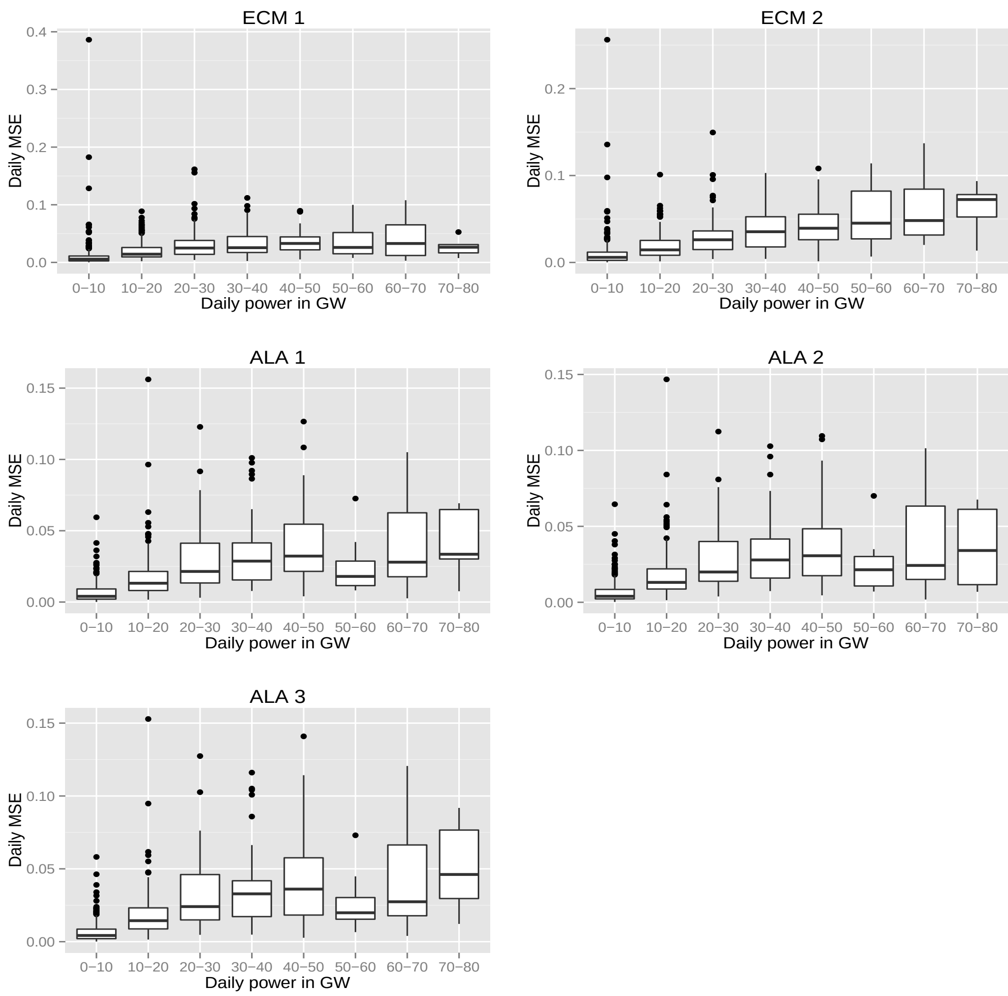

Figure 4.3: Relationship between daily MSE and daily power production

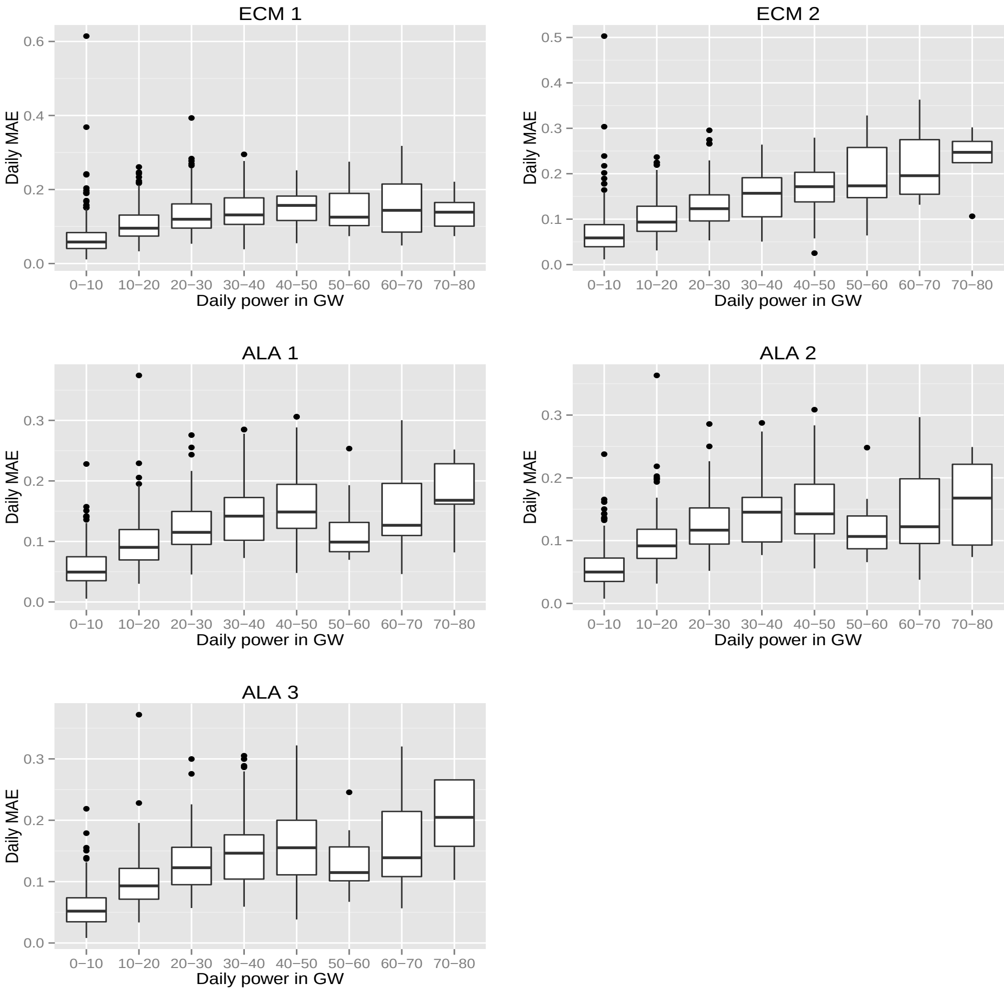

Figure 4.4: Relationship between daily MAE and daily power production

However, it is to be taken into account that the higher the power production, the less observations fall under the category - on most days the power produced is below 20 GW. Also it can be observed that in general the higher the power production, the more disperse the daily MSE/MAE is under this category. The positive linear relationship seen on the above plots is also backed up through statistical testing of linear modeling, which indicates a positive and significant slope for all models. For the model

$$d_t = \alpha_0 + \alpha_1 p_t + u_t$$

where d_t indicated the daily MSE (or MAE), p_t the daily power production at day t and u_t an error term, the parameter α_1 is positive and significantly different from zero.

Similar results are obtained when observing the relationship between the daily MSE/MAE of the power prediction and the daily estimation of power. This means that for days where more power production is estimated, it can be expected that the MSE and MAE will also be higher. Here we also mention that under the categories where a higher power production is estimated there are less observations and they are more disperse. A linear modeling for these two cases also shows a positive and significant relationship between these variables for all prediction models.

4.2 Patterns of the daily MSE and MAE of the power prediction models

In this section we want to see if the daily mean square/absolute error of the power estimated thought the ECM and ALA models follows any seasonal patterns. As only the data of two years is provided, applying a seasonal filter to the data is not possible, hence we will look for a more basic seasonal estimation. At first we look how the daily MSE and MAE is distributed throughout the seasons[2].

Note: one outlier has been ommited from the daily MSE ECM plots in order to have a better view of the data distribution.

As it can be seen from figures 4.5 and 4.6 generally the higher daily MSE/MAE values occur during the spring and winter season. For these two seasons the values also seem to be more disperse and not very concentrated around the median. The ALA models show that the daily MSE/MAE takes for both years the highest values in spring, while summer seems to be a more stable season. The ECM models show similar results, with the exception that winter 2013 shows the highest daily MSE/MAE throughout both years. As the available data to analyse includes only 2 years, we will try to fit a simple seasonal pattern considering the seasonal peaks. To evaluate this seasonal component, a composition of sine and cosine oscillation functions will be used.

[2]The season categorisation in this thesis follows the meteorological definition of seasons: Spring includes the months March - May; Summer includes June-August; Autumn includes September - November; Winter includes December - February

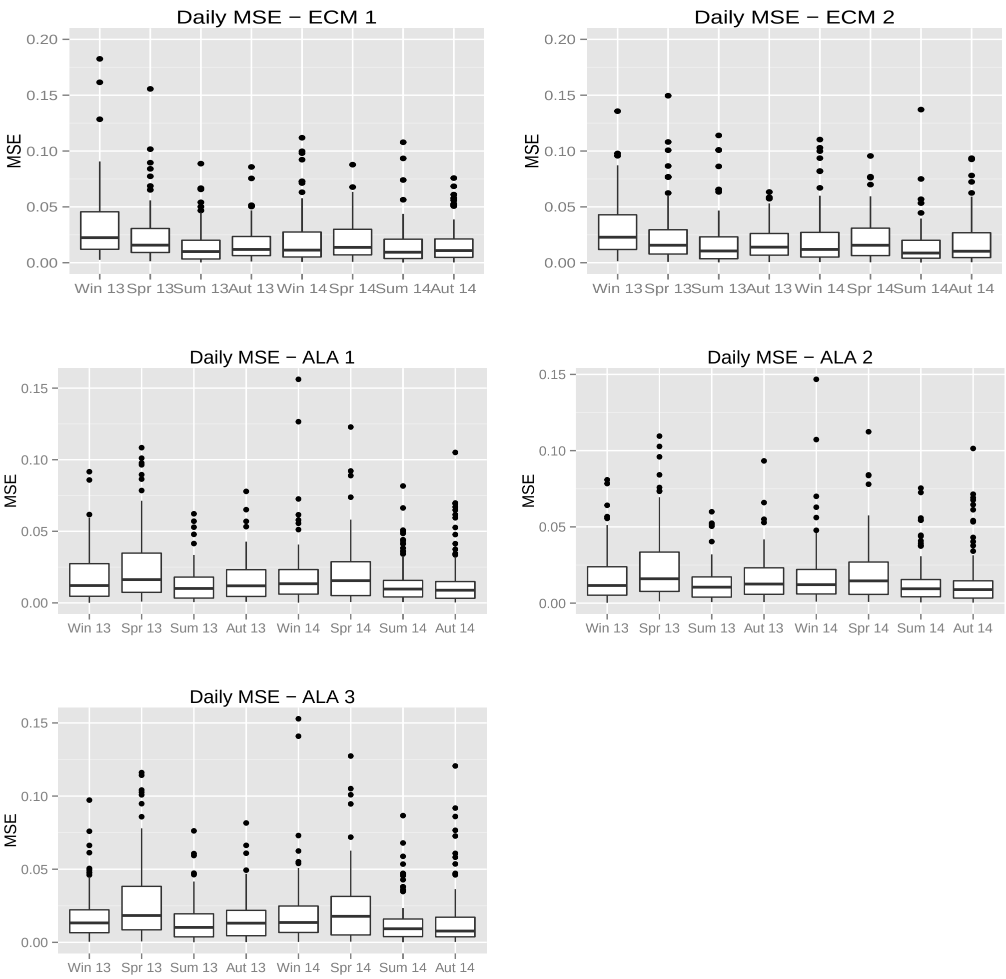

Figure 4.5: Daily MSE in different seasons

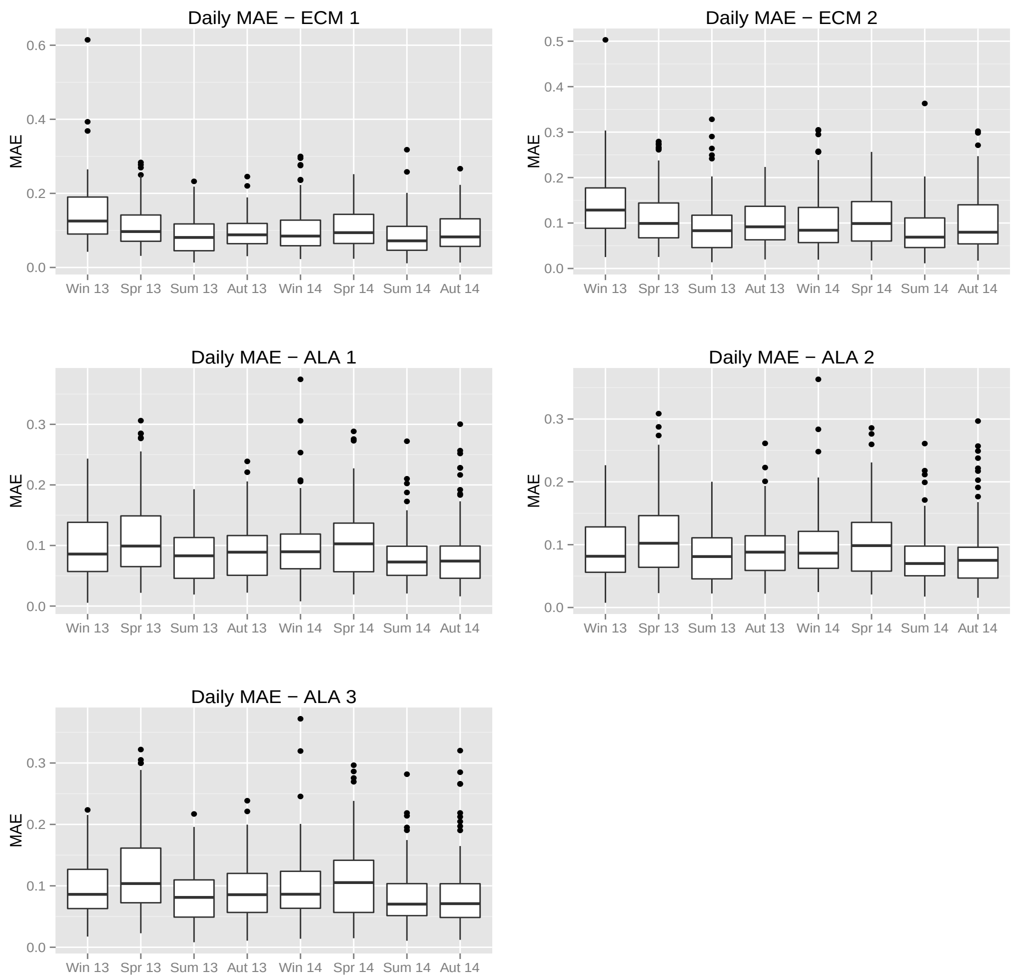

Figure 4.6: Daily MAE in different seasons

The first modeling to be tried is

$$\begin{aligned} M_t = &\alpha + \beta_1 \cos(t \cdot 360/365) + \beta_2 \sin(t \cdot 360/365) + \\ &\gamma_1 \cos(2t \cdot 360/365) + \gamma_2 \sin(2t \cdot 360/365) + u_t \end{aligned}$$

where M_t is the daily mean absolute or square error of the power estimation at day t whereas $t \in \{1, \dots, 365\}$ represents each day of the year and u_t is an error term. The arguments of the sine and cosine functions are given in degrees and the factor 360/365 is used as a scaling factor to fit the 365 days of the year in a full period of 360 degrees. The model fit was done using the $lm()$ function from the *stats* package in R, which estimates the coefficients using the ordinary least square method. However, when testing whether the coefficients from the linear model are significantly different from zero, this hypothesis is rejected for γ_1 and γ_2. Therefore the reduced model

$$M_t = \alpha + \beta \cos(t \cdot 360/365) + \gamma \sin(t \cdot 360/365) + u_t \tag{4.1}$$

will be used to fit a seasonal trend. The results of this fit are shown in figures 4.7 and 4.8: the dots represent the daily MSE/MAE of the power estimation and the red line represents the fitted values from the model described in (4.1). It should be noted though, that the coefficient of determination (R^2) in each model lies around 5%.

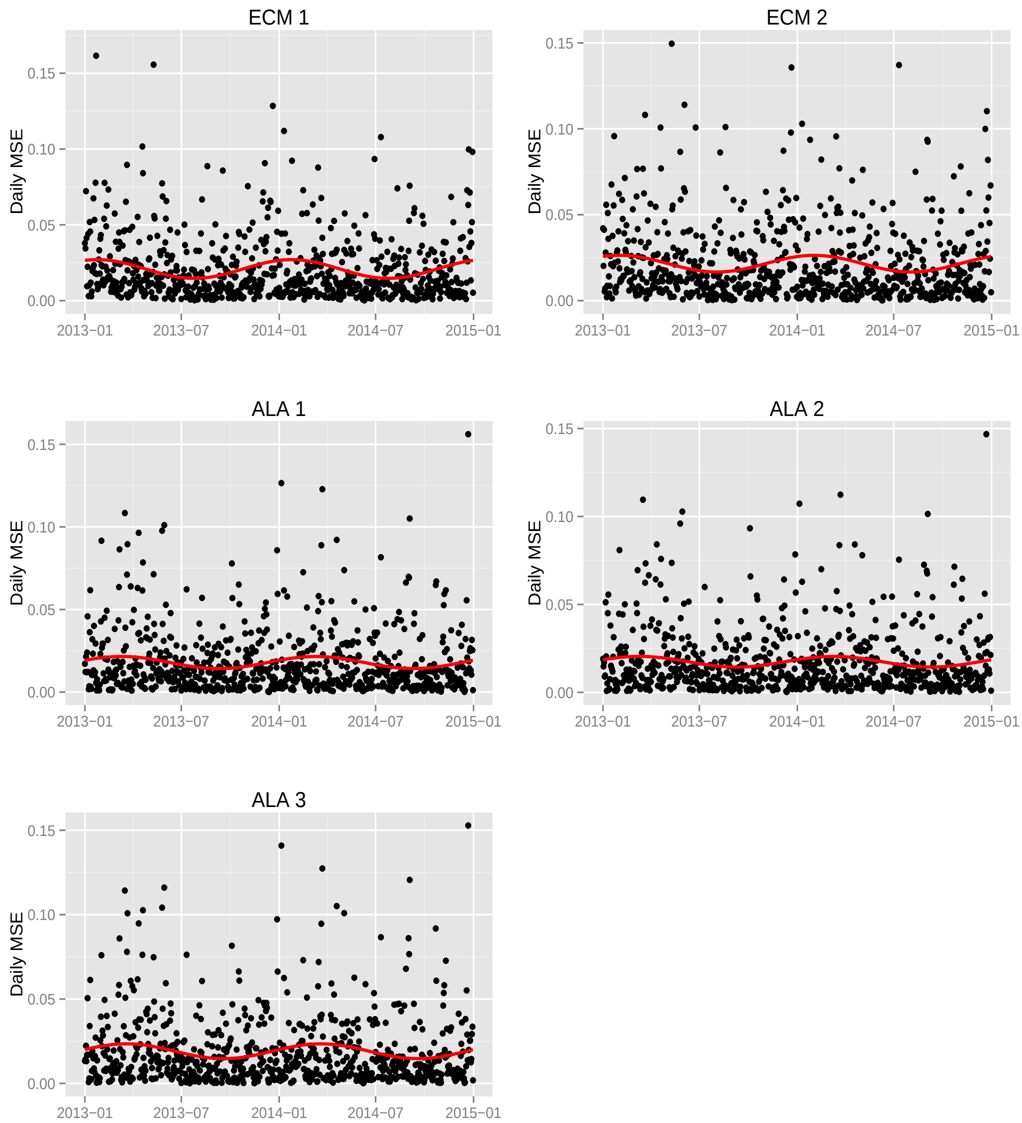

Figure 4.7: Fitting seasonal trend for daily MSE

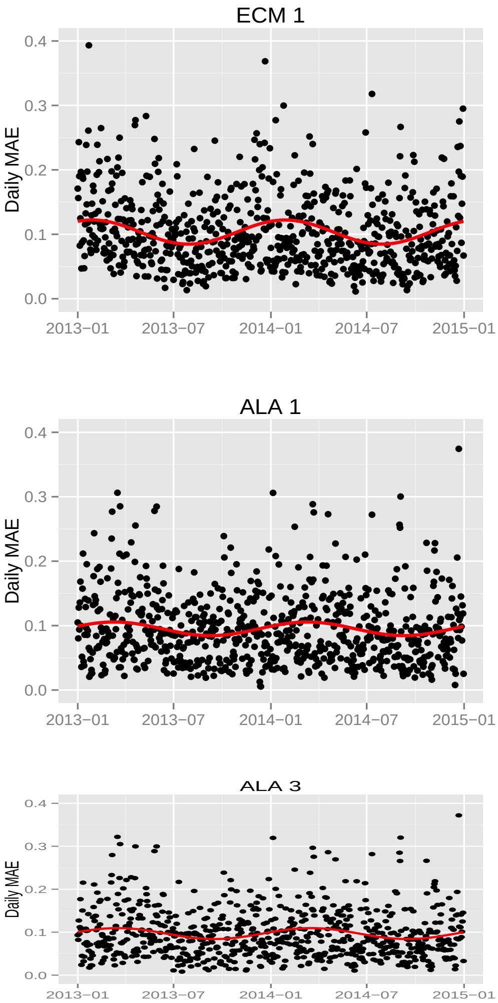

Figure 4.8: Fitting seasonal trend for daily MAE

CHAPTER 5

Confidence intervals

5.1 Current confidence interval

The scope of this thesis is to deliver a confidence interval for the errors of the power prediction models. Currently the confidence interval used for the wind power estimation is the range delivered from the predictions of the five power prediction models

$$[\,\min_i\{\hat{p}_{it}\}, \quad \max_i\{\hat{p}_{it}\}\,] \tag{5.1}$$

where $\hat{p}_{it}$ is the estimated power from model $i \in \{\text{ECM } 1 \ldots \text{ALA } 3\}$ at the time t.

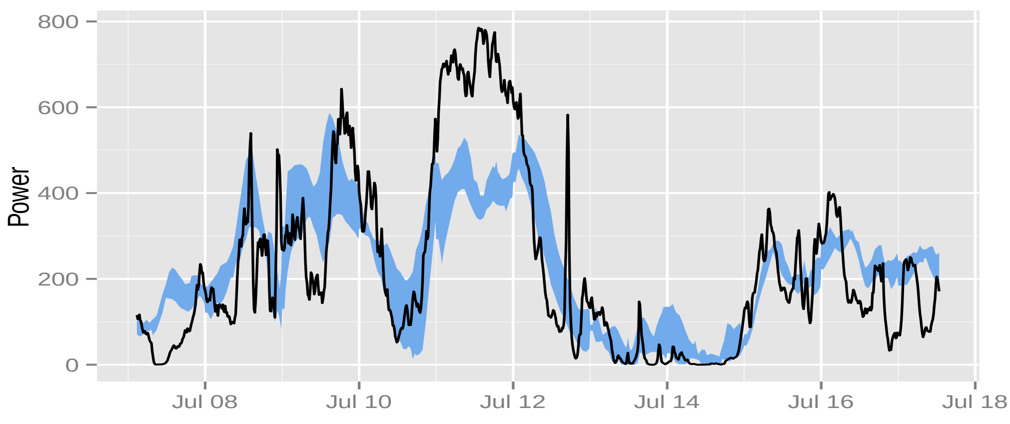

Figure 5.1: Measured power and confidence interval in current use for a random time interval

Figure 5.1 shows the measured power production from the wind energy with a black line and the currently used confidence interval with a shaded blue area for a random time interval. Already in this figure it can be seen that on many occasions the range provided by the prediction models is too narrow and/or far from the actual power produced. The data measured in the time intervals between the years 2013 to 2014 shows that only about 37% of the actual measured power values lie within this range.

The following sections will show alternative confidence intervals to use instead based on the behaviour of the error terms. The goal is to build a confidence interval based on the five existing power prediction models, so that 95% of the measured power values lie within its range.

5.2 Confidence interval based on the daily mean absolute error

The first idea to be tested is using the daily mean absolute error as a confidence interval for the errors. In this case it would be opted that the produced power p_{tq} at day t and time q lies within the interval

$$[\, \min_i \{\hat{p}_{itq} - mae_{it}\}, \quad \max_i \{\hat{p}_{itq} + mae_{it}\} \,]$$

where $\hat{p}_{itq}$ is the estimated power from model $i \in \{\text{ECM } 1 \ldots \text{ALA } 3\}$ at time q in day t and mae_{it} is the mean absolute error of model i at day t. This means that for every quarter-hourly interval in a day each model i respectively delivers the same amount of distance around the predicted power value $\hat{p}_{itq}$. As a final confidence interval for this modeling, the union of the five intervals provided by each model will be used taking the minimum of the boundaries between the models $i \in \{\text{ECM } 1 \ldots \text{ALA } 3\}$ as the lowest boundary of the confidence interval and the maximum of the five boundaries as the upper boundary. Depending on the magnitude of the errors it may occur that the daily mean absolute error is bigger than the predicted error, in which case the lower boundary of the confidence interval would result in a negative value. This would however not make sense as there can not be a negative power, therefore the confidence interval is

$$[\, \max\{0,\ \min_i \{\hat{p}_{itq} - mae_{it}\}\}, \quad \max_i \{\hat{p}_{itq} + mae_{it}\} \,]. \tag{5.2}$$

By evaluating this confidence interval with the calculated daily mean absolute error for the time period January 2013 - December 2014, it can be determined that 81.7% of the actual wind power measurements lies within this interval. This is quite a good result in comparison to the original confidence interval used, which only covers 37% of the data. Figure 5.2 shows the comparison between the confidence interval used currently

$$[\min_i\{\hat{p}_{it}\}, \quad \max_i\{\hat{p}_{it}\}]$$

which is marked with a dark blue colour and the confidence interval as described in (5.2), which is marked with a light blue colour for a random time interval. The black line on the other hand shows the actual measured power for that time interval.

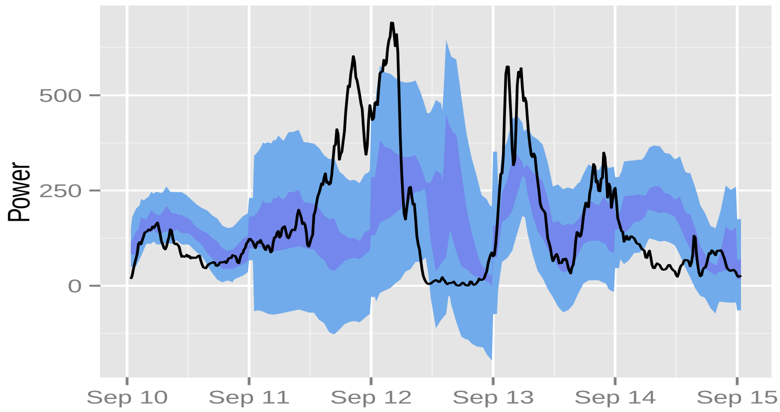

Figure 5.2: Comparison of the confidence intervals $[\min_i\{\hat{p}_{it}\}, \max_i\{\hat{p}_{it}\}]$ marked in dark blue and $[\min_i\{\hat{p}_{itq} - mae_{it}\}, \max_i\{\hat{p}_{itq} + mae_{it}\}]$ marked in light blue for a random time interval.

As the daily mean absolute error is an unknown quantity, the two following sections will show two different models to describe its behaviour. Using the predicted values $\widehat{mae}_{it}$ from these models, we will then build a confidence interval

$$[\min_i\{\hat{p}_{itq} - c \cdot \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + c \cdot \widehat{mae}_{it}\}] \tag{5.3}$$

where the parameter c will be determined as such that the interval (5.3) covers 95% of the power data measured.

5.2.1 Daily mean absolute errors - Model I

As next we will try to model the daily mean absolute error using a seasonal trend throughout th year. To be noted is that when modeling the daily mean absolute errors, the normalised errors (through the installed capacity) are considered, therefore when building the confidence intervals, the errors need to be scaled back by multiplying with the installed capacity at the point in time, which means

$$\widehat{mae}_{it} = I_t \cdot \widehat{MAE}_{it}$$

where I_t is the installed capacity at day t and MAE_{it} the daily mean of the normalised absolute errors of model i at day t.

To ensure that the modeled values for the daily mean absolute error are positive, the model estimation will be done on the logarithmic values of the daily normalised mean absolute errors for each model i. The model to be considered is the linear model

$$\ln(MAE_{it}) = \alpha_i + \beta_i \cos(t \cdot 360/365) + \gamma_i \sin(t \cdot 360/365) + u_{it} \tag{5.4}$$

with correlated errors following an AR(1) process

$$u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it} \tag{5.5}$$

where ε_{it} is a white noise. The estimation of the parameters α_i, β_i, γ_i and ρ_i is done through the *gls()* function in the *nlme* package, which makes a general least square (GLS) estimation of the model. The estimated parameters for all these models are presented in table 5.1.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\alpha}$	-2.4346	-2.4241	-2.5280	-2.5345	-2.5207
$\hat{\beta}$	0.1651	0.1359	0.0178	0.0332	0.0317
$\hat{\gamma}$	0.1037	0.0928	0.1074	0.1008	0.1379
$\hat{\rho}$	0.2726	0.2780	0.3179	0.2850	0.3190

Table 5.1: Estimated coefficients of (5.4), (5.5) for each of the five models

Given the data information up to time t, the one-step forecast for the model (5.4) is given as follows:

$$\widehat{\ln(MAE_{i,t+1})} = \hat{\alpha}_i + \hat{\beta}_i \cos\left(\frac{(t+1)\cdot 360}{365}\right) + \hat{\gamma}_i \sin\left(\frac{(t+1)\cdot 360}{365}\right) + \hat{\rho}_i \hat{u}_{it} \tag{5.6}$$

However, the residuals $\hat{u}_{it}$ are not known at time t, but they can be estimated using their AR(1) structure as $\hat{\rho}_i \hat{u}_{i,t-1}$. Using the information up to time $t-1$, we then have

$$\begin{aligned}\widehat{\ln(MAE_{i,t+1})} = &\hat{\alpha}_i + \hat{\beta}_i \cos\left(\frac{(t+1)\cdot 360}{365}\right) + \hat{\gamma}_i \sin\left(\frac{(t+1)\cdot 360}{365}\right) + \\ &\hat{\rho}_i^2 \hat{u}_{i,t-1} \end{aligned} \tag{5.7}$$

The forecasted values in (5.7) are the logarithmic values of the daily mean absolute (normalised) errors, so for the confidence interval they should be transformed back. To obtain an unbiased estimation of the daily mean absolute value at time $t+1$ given the information up to time $t-1$ we use[1]

$$\widehat{mae}_{i,t+1} = I_{t+1} \cdot \exp(\widehat{\ln(MAE_{i,t+1})}) \cdot \exp(\hat{\sigma}_i^2/2), \tag{5.8}$$

where $\hat{\sigma}_i^2$ is the estimated variance of $\rho\varepsilon_{it} + \varepsilon_{i,t+1}$.

To evaluate the goodness of the estimation presented in (5.7) and (5.8), it will be calculated how much of the power measured is covered from the interval

$$[\ \min_i\{\hat{p}_{itq} - \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + \widehat{mae}_{it}\}\ ].$$

Calculations with this confidence interval show that it covers 82.8% of the measured data. This value doesn't account for the opted 95%, however adding the factor $c = 2.05$ the new confidence interval

$$[\ \min_i\{\hat{p}_{itq} - 2.05\cdot\widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 2.05\cdot\widehat{mae}_{it}\}\ ] \tag{5.9}$$

[1] Given the AR(1) structure of the errors, $\ln(MAE_t)$ can be written as

$$\ln(MAE_{t+1}) = \underbrace{\alpha + \beta\cos\left(\frac{(t+1)\cdot 360}{365}\right) + \gamma\sin\left(\frac{(t+1)\cdot 360}{365}\right)}_{=:x_{t+1}\xi} + \underbrace{u_{t+1}}_{=\rho^2 u_{t-1}+\rho\varepsilon_t+\varepsilon_{t+1}}.$$

Hence the expected value of MAE_{t+1} given the information up to point $t-1$ (indicated by $\mathcal{F}_{t-1}$) can be written as

$$\begin{aligned}\mathbb{E}[MAE_{t+1}|\mathcal{F}_{t-1}] &= \mathbb{E}[e^{\ln(MAE)_{t+1}}|\mathcal{F}_{t-1}] \\ &= \mathbb{E}[e^{x_{t+1}\xi+\rho^2 u_{t-1}+\rho\varepsilon_t+\varepsilon_{t+1}}|\mathcal{F}_{t-1}] \\ &= e^{x_{t+1}\xi+\rho^2 u_{t-1}} + \mathbb{E}[e^{\rho\varepsilon_t+\varepsilon_{t+1}}|\mathcal{F}_{t-1}].\end{aligned}$$

Given that $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, then $e^{\rho\varepsilon_t+\varepsilon_{t+1}} \sim LN(0,\sigma^2)$, where $\sigma^2 = (1+\rho^2)\sigma_\varepsilon^2$. So then $\mathbb{E}[e^{\rho\varepsilon_t+\varepsilon_{t+1}}|\mathcal{F}_{t-1}] = e^{\sigma^2/2}$.

covers 95% of the measured power data. Figure 5.3 shows a comparison of the confidence interval with the fitted daily mean absolute values as in (5.9) marked with a light blue shaded area against the currently used confidence interval (dark blue area). The black line represents the measured power.

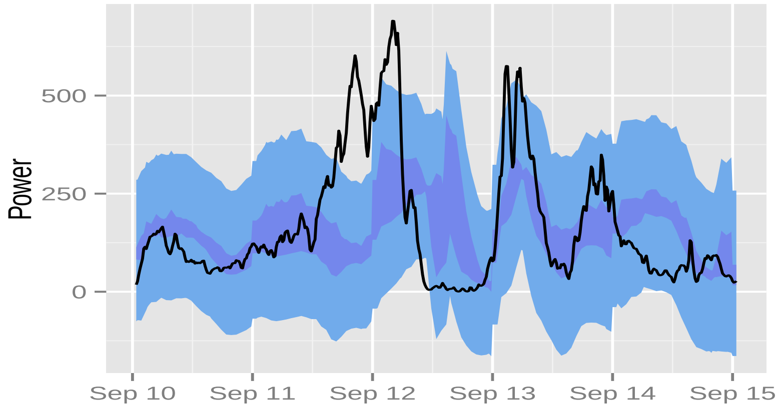

Figure 5.3: Confidence interval
$[\ \min_i\{\hat{p}_{itq} - 2.05 \cdot \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 2.05 \cdot \widehat{mae}_{it}\}\ ]$ on a random time interval marked in light blue against the currently used confidence interval $[\min_i\{\hat{p}_{it}\}, \quad \max_i\{\hat{p}_{it}\}]$ marked in dark blue

To see if this model really works well, a cross validation will be done. The data will be split into a training set (data for the days $1, \ldots, T_1$) and a test data set (days $T_1, \ldots, T$). The parameters of the model are estimated using the training set. Day ahead forecasts (for day $t+1$) are then made for the test data set using information available up to day $t-1$ as mentioned above. The evaluation of the performance of the model is done by seeing what percentage of the power measured lies within the forecasted interval for the days $T_1, \ldots, T$.

Figure 5.4 shows the results of the cross validation. The data available for this thesis is two years, so the daily mean absolute error analysis was done for 730 days and the results were applied on 15-minute intervals each day. The x-axis shows how many days were used as a training set (T_1), while the y-axis shows the percentage of the measured power data covered from the confidence interval predicted for the test set. The percentages in the cross validation range from around 95% to 98%, which is quite a good result.

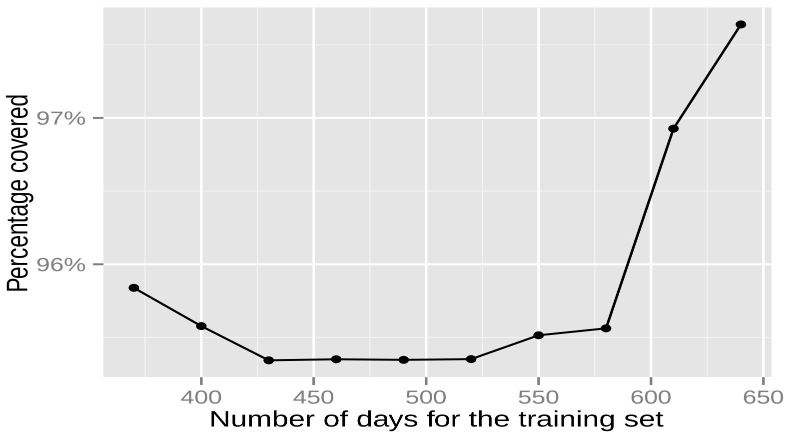

Figure 5.4: Cross validation results on the confidence interval $[\,\min_i\{\hat{p}_{itq} - 2.05 \cdot \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 2.05 \cdot \widehat{mae}_{it}\}\,]$ where the daily mean absolute errors are modeled according to (5.4) with AR(1) residuals.

5.2.2 Daily mean absolute error - Model II

In this section it will be attempted to model the (normalised) daily mean absolute error by also taking into account the predicted value of the wind speed for Neusiedler See as well as the mean value of the power predicted:

$$\begin{aligned}\ln(MAE)_{it} =& \alpha_i + \beta_i \cos(t \cdot 360/365) + \gamma_i \sin(t \cdot 360/365) + \\ & \eta_i \ln(\widehat{mw}_{it}) + \zeta_i \ln(\widehat{mp}_{it}) + u_{it}\end{aligned} \tag{5.10}$$

where $\widehat{mw}_{it}$ is the mean of the wind speed (in Neusiedler See) predicted by model i for day t and $\widehat{mp}_{it}$ is the mean power predicted by model i for day t. Also in this case $u_{it} \sim AR(1)$, so

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}. \tag{5.11}$$

The parameter estimation is again done through a GLS estimation. The forecasted values $\widehat{mae}_{i,t+1}$ are done following the same analogy as in the previous section, so based on information up to time $t-1$. The confidence interval

$$[\,\min_i\{\hat{p}_{itq} - \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + \widehat{mae}_{it}\}\,],$$

where the daily mean absolute error is modeled according to (5.10) and (5.11) covers 84.2% of the measured power. However, using a factor $c = 1.82$, so that

$$[\,\min_i\{\hat{p}_{itq} - 1.82 \cdot \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 1.82 \cdot \widehat{mae}_{it}\}\,] \tag{5.12}$$

offers the 95% coverage of the power data. The estimated parameters for the model (5.10) and (5.11) are presented in table 5.2.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\alpha}$	-4.3946	-4.6143	-4.8527	-4.9195	-3.7667
$\hat{\beta}$	0.0350	0.0033	-0.0462	-0.0451	-0.0326
$\hat{\gamma}$	0.0773	0.0664	0.0922	0.0876	0.1000
$\hat{\eta}$	0.2153	0.3821	0.1308	0.0544	0.6026
$\hat{\zeta}$	0.3507	0.3572	0.4602	0.4846	0.1682
$\hat{\rho}$	0.2108	0.2371	0.2341	0.2467	0.2331

Table 5.2: Estimated coefficients of (5.10),(5.11) for each of the five models

The cross validation results on this model are shown in figure 5.5. The results are quite good, as the percentage of data covered is above 95% for all the test data sets.

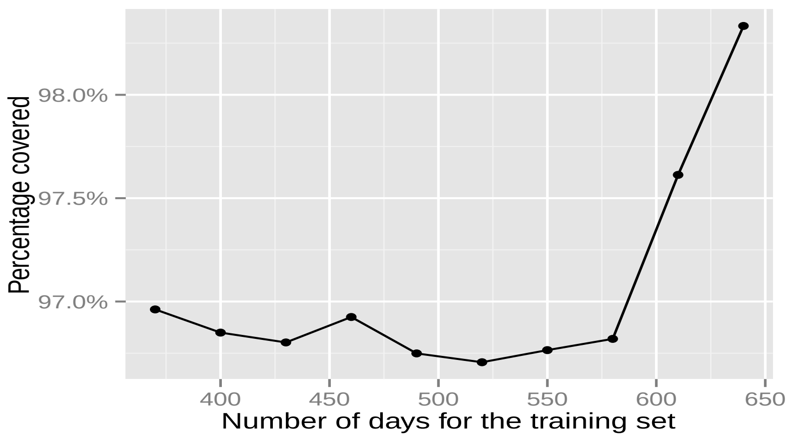

Figure 5.5: Cross validation results on the confidence interval $[\,\min_i\{\hat{p}_{itq} - 1.82 \cdot \widehat{mae}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 1.82 \cdot \widehat{mae}_{it}\}\,]$ where the daily mean absolute errors are modeled according to (5.10) with AR(1) residuals.

5.3 Confidence interval based on the daily mean square error

Similarly to the structure of section 5.2, this section will evaluate a confidence interval based however on the daily mean square errors. This confidence interval will involve building an interval using the standard deviation of the errors around the predicted value of power:

$$[\,\min_i\{\hat{p}_{itq} - \sqrt{mse_{it}}\},\quad \max_i\{\hat{p}_{itq} + \sqrt{mse_{it}}\}\,] \tag{5.13}$$

where $\hat{p}_{itq}$ is the estimated power from model $i \in \{\text{ECM } 1 \ldots \text{ALA } 3\}$ at time q in day t and mse_{it} is the mean square error of model i at day t. Here again every model i uses the same deviance (mse_{it}) for all quarter-hour predictions within a day. Using the calculated daily mean square errors for the time interval between the years 2013 to 2014 in (5.13) it can be measured that 86.2% of the power measured lies within this interval. Figure 5.6 shows a comparison of the confidence interval (5.13) with the calculated daily mean square errors against the currently used confidence interval for a random time interval.

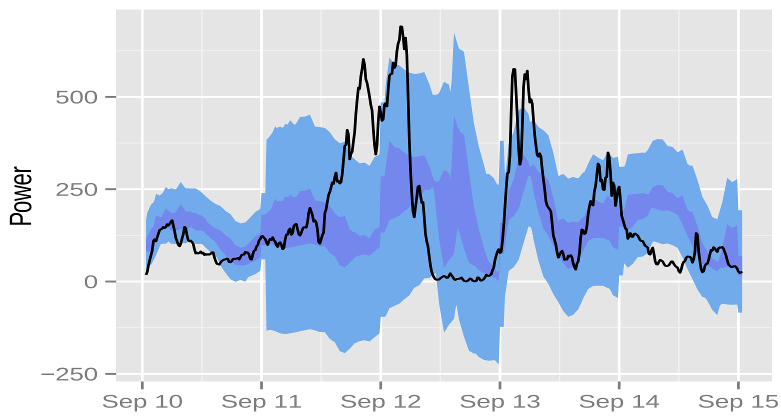

Figure 5.6: Comparison of the confidence intervals $[\,\min_i\{\hat{p}_{itq} - \sqrt{mse_{it}}\},\quad \max_i\{\hat{p}_{itq} + \sqrt{mse_{it}}\}\,]$ marked in light blue and $[\,\min_i\{\hat{p}_{itq}\},\quad \max_i\{\hat{p}_{itq}\}\,]$ marked in dark blue. The black line represents the power measured.

The two following sections offer two models for the estimation of the daily mean absolute error similar to the ones presented in section 5.2.

5.3.1 Modeling the daily mean square error - Model I

At first we will try to model the daily mean square errors using only the seasonal trend similar to section 5.2.1. Also in this case the daily mean square normalised errors are modeled, so

$$\widehat{mse}_{it} = I_t^2 \cdot \widehat{MSE}_{it}$$

where I_t is the installed capacity and MAE_{it} the mean of the normalised square errors for model i at day t. To assure positive prediction values for the daily mean square error, the model is again build using the logarithmic values

$$\ln(MSE_{it}) = \alpha_i + \beta_i \cos(t \cdot 360/365) + \gamma_i \sin(t \cdot 360/365) + u_{it} \tag{5.14}$$

where $u_{it} \sim AR(1)$. To obtain an unbiased estimation from the transformed values of the daily mean square errors at time $t+1$ given the information up to time $t-1$ we use

$$\widehat{mse}_{i,t+1} = I_{t+1} \cdot \exp(\widehat{\ln(MSE_{i,t+1})}) \cdot \exp(\hat{\sigma}_i^2/2), \tag{5.15}$$

where $\hat{\sigma}_i^2$ is the estimated variance of $\rho\varepsilon_{it} + \varepsilon_{i,t+1}$ (similar to section 5.2.1). Using (5.14) and (5.15) to build the confidence interval

$$[\min_i\{\hat{p}_{itq} - \sqrt{\widehat{mse}_{it}}\}, \quad \max_i\{\hat{p}_{itq} + \sqrt{\widehat{mse}_{it}}\}] \tag{5.16}$$

shows that 89.7% of the power measured is covered from this confidence interval. In this case a 95% coverage of the measured data is given through the interval

$$[\min_i\{\hat{p}_{itq} - 1.41 \cdot \sqrt{\widehat{mse}_{it}}\}, \quad \max_i\{\hat{p}_{itq} + 1.41 \cdot \sqrt{\widehat{mse}_{it}}\}]. \tag{5.17}$$

Table 5.3 shows the estimated parameters of the model (5.14), while figure 5.7 shows the cross validation results of using the confidence interval (5.17) and modeling the daily mean square error according to (5.14). Also in this case the results of the cross validation are quite good covering above 95% of the measured power data for each test data set.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\alpha}$	-4.4331	-4.4208	-4.5936	-4.6003	-4.5745
$\hat{\beta}$	0.2686	0.2384	0.0211	0.0424	0.0375
$\hat{\gamma}$	0.1965	0.1855	0.2094	0.1924	0.2701
$\hat{\rho}$	0.2883	0.3019	0.3212	0.3019	0.2955

Table 5.3: Estimated coefficients of (5.14) for each of the five models

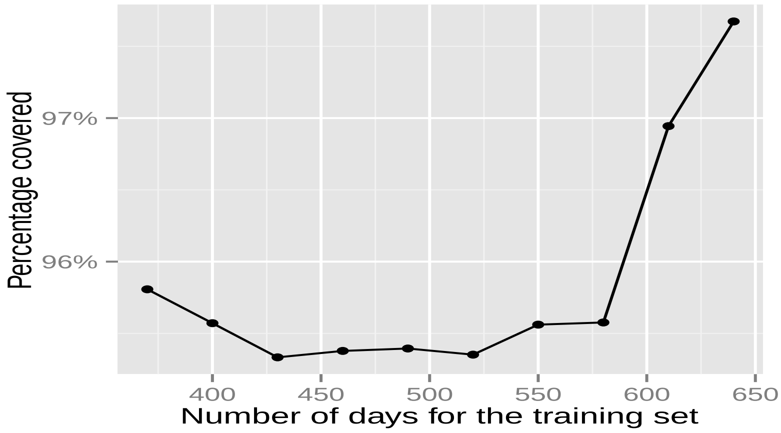

Figure 5.7: Cross validation results on the confidence interval $[\min_i\{\hat{p}_{itq} - 1.41 \cdot \widehat{mse}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 1.41 \cdot \widehat{mse}_{it}\}]$ where the daily mean square errors are modeled according to (5.14).

5.3.2 Modeling the daily mean square error - Model II

Similarly to section 5.2.2, it will also be attempted to model the (normalised) daily mean square errors by also taking into account the predicted value of the wind speed for Neusiedler See as well as the mean value of the power predicted by each of the models:

$$\begin{aligned}\ln(MSE_{it}) =& \alpha_i + \beta_i \cos(t \cdot 360/365) + \gamma_i \sin(t \cdot 360/365) + \\ & \eta_i \ln(\widehat{mw}_{it}) + \zeta_i \ln(\widehat{mp}_{it}) + u_{it}\end{aligned} \tag{5.18}$$

with $u_{it} \sim AR(1) \quad \forall i$, where $\widehat{mw}_{it}$ is the mean of the wind speed (in Neusiedler See) predicted by model i for day t and $\widehat{mp}_{it}$ is the mean power predicted by model i for day t. This modeling of the daily mean square errors shows slightly better results compared to the modeling based only on the seasonal pattern, as the confidence interval

$$[\min_i\{\hat{p}_{itq} - \sqrt{\widehat{mse}_{it}}\}, \quad \max_i\{\hat{p}_{itq} + \sqrt{\widehat{mse}_{it}}\}] \tag{5.19}$$

where $\widehat{mse}_{it}$ is delivered through (5.18) covers 90.5% of the observed power data. The 95% coverage for this model is obtained through the confidence interval

$$[\min_i\{\hat{p}_{itq} - 1.35 \cdot \sqrt{\widehat{mse}_{it}}\}, \quad \max_i\{\hat{p}_{itq} + 1.35 \cdot \sqrt{\widehat{mse}_{it}}\}]. \tag{5.20}$$

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\alpha}$	-7.7324	-8.5125	-8.7562	-8.9224	-7.3571
$\hat{\beta}$	0.0414	-0.0005	-0.0932	-0.0984	-0.0911
$\hat{\gamma}$	0.1424	0.1450	0.1826	0.1705	0.2137
$\hat{\eta}$	0.5650	0.5530	0.2276	0.0635	0.8528
$\hat{\zeta}$	0.5417	0.7059	0.8251	0.8835	0.4527
$\hat{\rho}$	0.2217	0.2583	0.2445	0.2709	0.2414

Table 5.4: Estimated coefficients of (5.18) for each of the five models

Table 5.4 shows the estimated parameters of (5.18), while figure 5.8 shows the cross validation results, which are over 96% for each of the test sets.

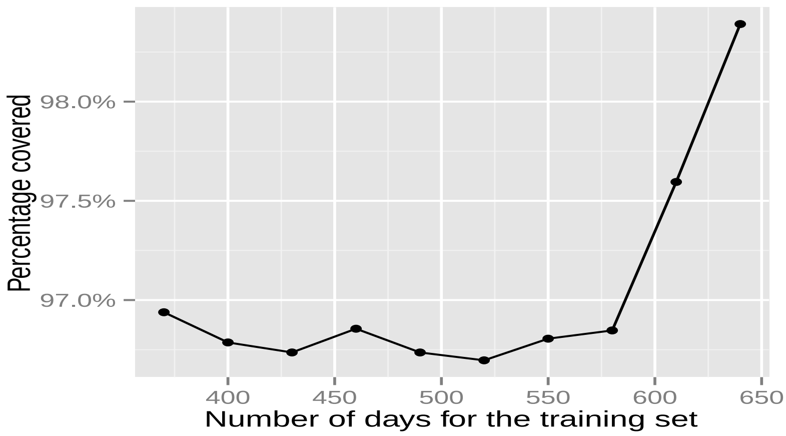

Figure 5.8: Cross validation results on the confidence interval $[\,\min_i\{\hat{p}_{itq} - 1.35 \cdot \widehat{mse}_{it}\}, \quad \max_i\{\hat{p}_{itq} + 1.35 \cdot \widehat{mse}_{it}\}\,]$ where the daily mean square errors are modeled according to (5.19).

5.4 Confidence intervals on 15-minute basis

The confidence intervals shown in the sections above use the same measure of uncertainty (daily mean square error or daily mean absolute error) for the whole day. In this section we will attempt to build a confidence interval of the errors on quarter-hourly basis using the absolute and squared errors.

5.4.1 Confidence interval using the absolute error

The first confidence interval to be investigated is an interval build on the absolute error around the predicted power:

$$[\ \min_i\{\hat{p}_{itq} - a_{itq}\}, \quad \max_i\{\hat{p}_{itq} + a_{itq}\}\ ] \tag{5.21}$$

where a_{itq} is the absolute error of model i at day t at time q. Obviously the measured power p_{tq} at any point in time is included in this interval. However we need an estimate on how the absolute error behaves. In this case we will try to generate the (normalised) absolute error as a product of the daily mean absolute error (mae_t) and a factor r_{tq}, which depends on the estimated power, the estimated wind speed in Neusiedler See, the time of the day and the range given by the currently used confidence interval. The variables involved in r_{tq} were chosen based on their importance and how they seem to affect the error terms. The modeling will be done on the normalised absolute errors A_{itq}, which in the end will be scaled back by the installed power I_t so that

$$a_{itq} = I_t \cdot A_{itq}.$$

The model that will need to be estimated is

$$A_{itq} = mae_{it} \cdot r_{itq} \tag{5.22}$$

whereas the daily mean absolute error mae_{it} is modeled as in (5.4) with the autoregressive residuals of first order. To assure a positive value for the term r_{itq}, its square root value will be modeled[2]

$$\sqrt{r_{itq}} = \delta_q \omega_{iq} + \hat{w}_{itq}\theta_i + \hat{p}_{itq}\xi_i + K_{tq}\tau_i + u_{itq} \tag{5.23}$$

where δ_q represents a factor (dummy variable) indicating each quarter hour in a day {00:00, 00:15, ... , 23:45}, $\hat{w}_{itq}$ represent the estimated wind speed in Neusiedler See by model i at day t time q, $\hat{p}_{itq}$ is the estimated power and K_{tq} is the range of the currently used confidence interval (see (5.1)) at the point in time. The errors u_{itq} in (5.23) are also correlated. In this case the estimation of the parameters in (5.23) was done using the Cochrane-Orcutt procedure: for a linear model

$$y_t = x_t\beta + u_t \tag{5.24}$$

with autoregressive errors of first order

$$u_t = \rho u_{t-1} + \varepsilon_t$$

[2]A transformation with the logarithmic function did not provide good results in this case.

the transformed model using the difference

$$\tilde{y}_t := y_t - \rho y_{t-1} = \underbrace{(x_t - x_{t-1})}_{=:\tilde{x}_t}\beta + \underbrace{u_t - \rho u_{t-1}}_{=\varepsilon_t} = \tilde{x}_t\beta + \varepsilon_t \tag{5.25}$$

can be used to estimate β with OLS (ordinary least squares), as the error term ε_t in (5.25) is white noise. This estimated $\hat{\beta}$ can then be used to estimate (5.24). The AR coefficient ρ used for the Cochrane-Orcutt procedure on (5.23) was estimated from the errors of an OLS regression on (5.23). The forecast for $\sqrt{r_{itq}}$ at any time tq is given as

$$\widehat{\sqrt{r_{itq}}} = \delta_q\hat{\omega}_{iq} + \hat{w}_{itq}\hat{\theta}_i + \hat{p}_{itq}\hat{\xi}_i + K_{tq}\hat{\tau}_i. \tag{5.26}$$

To obtain an unbiased estimation for r_{itq} we use

$$\widehat{r_{itq}} = \widehat{\sqrt{r_{itq}}}^2 + \hat{\sigma}_2^2$$

where $\hat{\sigma}_2^2$ is the estimated variance of the model residuals $\hat{u}_{itq}$.[3] So in the end the predicted values for $\hat{a}_{i,t+1,q}$, $q = 00:00, \ldots, 23:45$ given the information up to day $t-1$ will be given as

$$\hat{a}_{i,t+1,q} = I_{t+1} \cdot \exp(\widehat{\ln(MAE_{i,t+1})}) \cdot \exp(\hat{\sigma}_i^2/2) \cdot (\widehat{\sqrt{r_{i,t+1,q}}}^2 + \hat{\sigma}_{i2}^2). \tag{5.27}$$

Using the day ahead estimates given the information up to the day before as provided by (5.22) to estimate a confidence interval according to (5.21), we look for a constant c so that 95% of the measured power data lies in the interval

$$[\ \min_i\{\hat{p}_{itq} - c\cdot\hat{a}_{itq}\}, \quad \max_i\{\hat{p}_{itq} + c\cdot\hat{a}_{itq}\}\ ].$$

$c = 1.57$ offers such a confidence interval. A cross validation similarly to the one explained in the previous sections is then done on the confidence interval

$$[\ \min_i\{\hat{p}_{itq} - 1.57\cdot\hat{a}_{itq}\}, \quad \max_i\{\hat{p}_{itq} + 1.57\cdot\hat{a}_{itq}\}\ ] \tag{5.28}$$

[3] For this type of model we have $\sqrt{r_{t+1}} = \widehat{\sqrt{r_{t+1}}} + \hat{u}_{t+1}$. The expected value of r_{t+1} given the information up to time t is

$$\begin{aligned}\mathbb{E}[r_{t+1}|\mathcal{F}_t] &= \mathbb{E}[(\widehat{\sqrt{r_{t+1}}} + \hat{u}_{t+1})^2|\mathcal{F}_t] \\ &= \widehat{\sqrt{r_{t+1}}}^2 + 2\widehat{\sqrt{r_{t+1}}}\underbrace{\mathbb{E}[\hat{u}_{t+1}|\mathcal{F}_t]}_{=0} + \mathbb{E}[\hat{u}_{t+1}^2|\mathcal{F}_t].\end{aligned}$$

$\mathbb{E}[\hat{u}_{t+1}^2|\mathcal{F}_t]$ can be estimated as the variance of the residuals $\hat{u}_t$.

using the above mentioned modeling of the absolute value and the results are shown in figure 5.9. The most test data sets covered less than 95% of the measurements, however above 94%, which is a very good result.

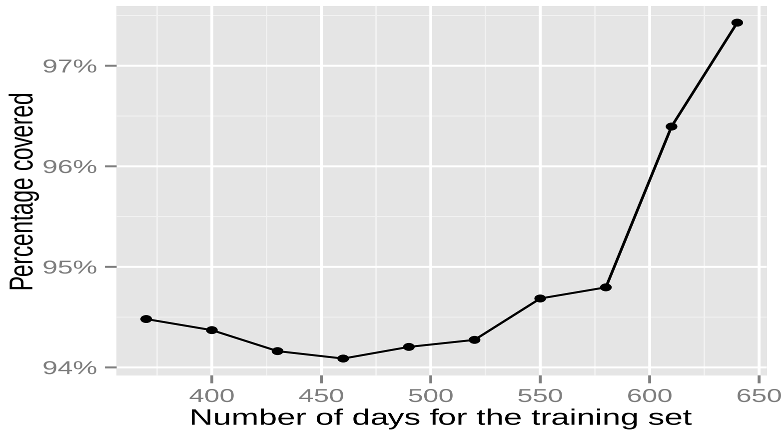

Figure 5.9: Cross validation results on the confidence interval $[\,\min_i\{\hat{p}_{itq} - 1.57 \cdot \hat{a}_{itq}\}, \quad \max_i\{\hat{p}_{itq} + 1.57 \cdot \hat{a}_{itq}\}\,]$ where the absolute errors are modeled according to (5.27).

The estimated parameters of the mae_{it} part of this model are the same as in table 5.1, while the estimated parameters of (5.23) are presented in table 5.5 ($\hat{\theta}_i, \hat{\xi}_i$ and $\hat{\tau}_i$) and figure 5.10 ($\hat{\omega}_{iq}$ for all quarter hours q).

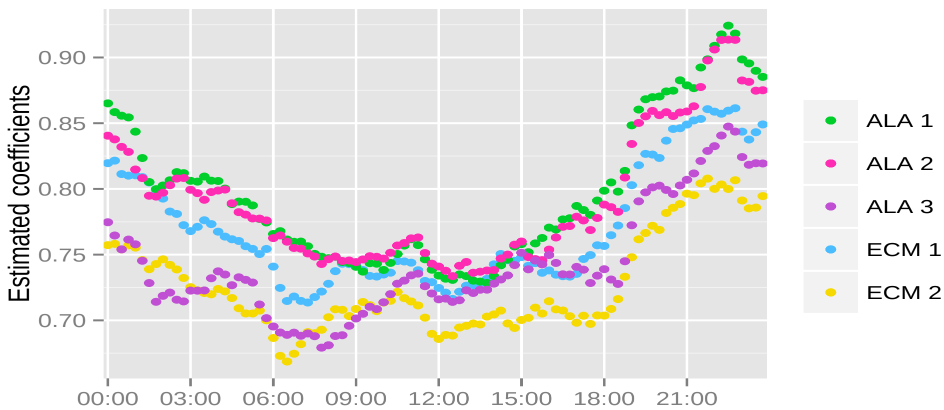

Figure 5.10: Coefficients $\hat{\omega}_{iq}$ of each quarter hour (q) for all models $i \in \{$ ECM 1, ... , ALA 3$\}$.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\theta}_i$	0.027	0.051	0.008	-0.003	0.087
$\hat{\xi}_i$	0.00004	-0.0001	0.0003	0.0004	-0.0004
$\hat{\tau}_i$	0.001	0.0005	0.0005	0.001	0.0005

Table 5.5: Estimated coefficients $\hat{\theta}_i, \hat{\xi}_i$ and $\hat{\tau}_i$ of the model $\sqrt{r_{itq}} = \delta_q \omega_{iq} + \hat{w}_{itq}\theta_i + \hat{p}_{itq}\xi_i + K_{tq}\tau_i + u_{itq}$

5.4.2 Confidence interval using the square errors

On the other hand we will provide a similar confidence interval using the squared errors of each 15-minute time interval. In this case the used confidence interval which covers 95% of the wind power measurements is

$$[\ \min_i\{\hat{p}_{itq} - 1.16 \cdot \sqrt{\hat{s}_{itq}}\}, \quad \max_i\{\hat{p}_{itq} + 1.16 \cdot \sqrt{\hat{s}_{itq}}\}\ ] \tag{5.29}$$

where s_{itq} is the squared error of model i at day t and time q. The estimation of the squared errors s_{itq} comes from the estimation of the normalised square errors S_{itq}:

$$\hat{s}_{itq} = I_t^2 \cdot \hat{S}_{itq},$$

where similarly as in section 5.4.1

$$\hat{S}_{itq} = \widehat{mse}_{it} \cdot \hat{r}_{itq}. \tag{5.30}$$

The daily mean square error term mse_{it} is modeled as in (5.14) with autoregressive errors of first order and the term r_{itq} is modeled as

$$\sqrt{r_{itq}} = \delta_q \omega_{iq} + \hat{w}_{itq}\theta_i + \hat{p}_{itq}\xi_i + K_{tq}\tau_i + u_{itq} \tag{5.31}$$

where δ_q is again a factor (dummy variable) indicating each quarter hour in day ($q \in \{00{:}00,\ 00{:}15,\ \ldots\ ,\ 23{:}45\}$), $\hat{w}_{itq}$ represents the estimated wind speed in Neusiedler See by model i at day t time q, $\hat{p}_{itq}$ is the estimated power and K_{tq} is the range of the currently used confidence interval (see (5.1)) at the point in time. The estimation of the parameters in (5.31) was also done through the Cochrane-Orcutt procedure (as in section 5.4.1) and the estimation of $\hat{r}_{itq}$ from $\widehat{\sqrt{r_{itq}}}$ was also done in an analogue way as described in section 5.4.1.

Figure 5.11 shows the cross validation on this model. The 95% coverage through these modeled intervals is only obtained with two of the test sets, however the others also show good results, covering over 94% of the measured power data.

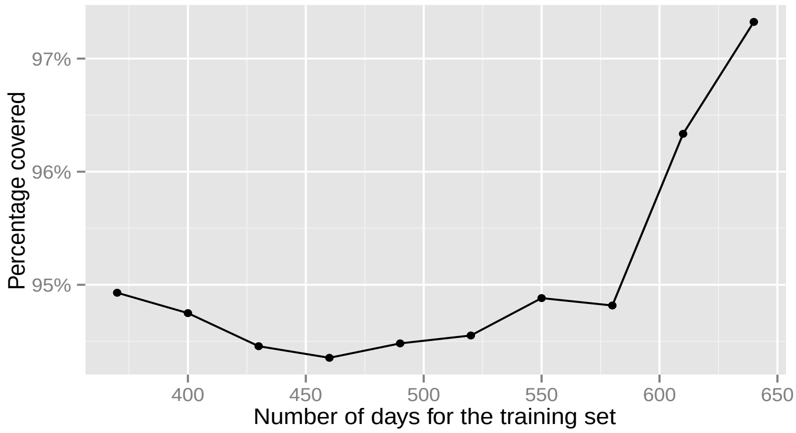

Figure 5.11: Cross validation results on the confidence interval $[\,\min_i\{\hat{p}_{itq} - 1.16 \cdot \sqrt{\hat{s}_{itq}}\}, \quad \max_i\{\hat{p}_{itq} + 1.16 \cdot \sqrt{\hat{s}_{itq}}\}\,]$ where the squares errors are modeled according to (5.30).

The estimated parameters of the mse_{it} part of this model are the same as in table 5.3, the estimated parameters $\hat{\theta}_i, \hat{\xi}_i$ and $\hat{\tau}_i$ of (5.31) are presented in table 5.5 and the $\hat{\omega}_{iq}$ estimated parameters are shown in figure 5.12.

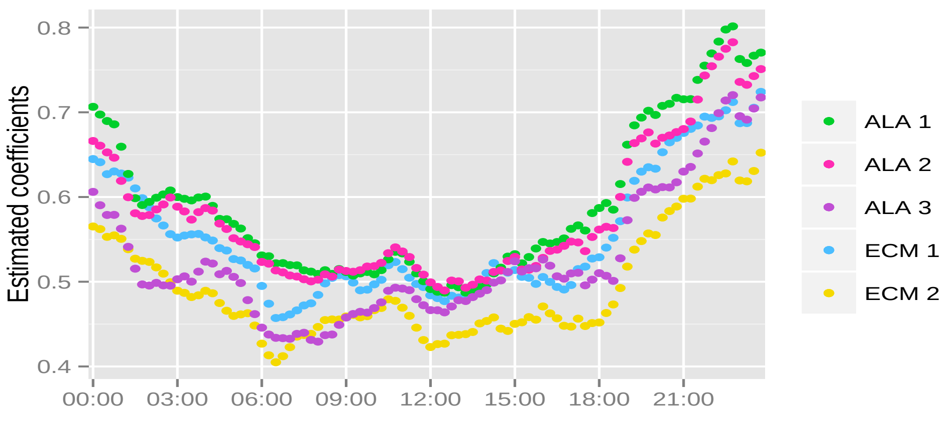

Figure 5.12: Coefficients $\hat{\omega}_{iq}$ of each quarter hour (q) for all models $i \in \{$ ECM 1, ... , ALA 3$\}$.

	ECM 1	ECM 2	ALA 1	ALA 2	ALA 3
$\hat{\theta}_i$	0.049	0.088	0.035	0.013	0.133
$\hat{\xi}_i$	-0.00003	-0.0003	0.0003	0.0005	-0.001
$\hat{\tau}_i$	0.001	0.001	0.001	0.001	0.001

Table 5.6: Estimated coefficients $\hat{\theta}_i, \hat{\xi}_i$ and $\hat{\tau}_i$ of the model $\sqrt{r_{itq}} = \delta_q \omega_{iq} + \hat{w}_{itq}\theta_i + \hat{p}_{itq}\xi_i + K_{tq}\tau_i + u_{itq}$

5.5 Comparison of the confidence intervals

Sections 5.2, 5.3 and 5.4 show different confidence intervals for the errors of each prediction model. They all seem to perform quite well according to the cross validations done, as the majority of the power produced is included in these modeled confidence intervals. Furthermore, all the newly defined intervals show a better performance in comparison to the confidence interval currently used, which only covers 37% of the power measured.

On the other hand, confidence intervals covering more of the measured data are also wider (the range is bigger), so a trade off between the data covered and interval width needs to be made. In this aspect we will do a comparison of the range of each of the confidence intervals presented in the sections above.

Table 5.7 shows the average range (width) in MW of each of the constructed confidence intervals. The estimated variables in these intervals are the predicted values of the variables for the day ahead ($t + 1$) given the information up to day $t - 1$, as explained in the sections above.

As it can be seen from the results in table 5.7, the confidence intervals with the lowest average ranges are

$$[\min_i\{\hat{p}_{itq} - 1.35 \cdot \sqrt{\widehat{mse}_{it}}\}, \max_i\{\hat{p}_{itq} + 1.35 \cdot \sqrt{\widehat{mse}_{it}}\}] \tag{5.32}$$

and

$$[\ \min_i\{\hat{p}_{itq} - 1.82 \cdot \widehat{mae}_{it}\},\ \max_i\{\hat{p}_{itq} + 1.82 \cdot \widehat{mae}_{it}\}\ ] \tag{5.33}$$

where the daily mean square error $\widehat{mse}_{it}$ and the daily mean absolute error $\widehat{mae}_{it}$ are modeled according to the seasonal trend, the daily average power

Confidence interval	**% covered**	**Mean range in MW**
$[\min_i\{\hat{p}_{itq}\}, \max_i\{\hat{p}_{itq}\}]$ the currently used confidence interval	37%	87.6
$[\min_i\{\hat{p}_{itq} - 2.05\widehat{mae}_{it}\}, \max_i\{\hat{p}_{itq} + 2.05\widehat{mae}_{it}\}]$, where $\widehat{mae}_{it}$ is modeled as in section 5.2.1	95%	412.4
$[\min_i\{\hat{p}_{itq} - 1.82\widehat{mae}_{it}\}, \max_i\{\hat{p}_{itq} + 1.82\widehat{mae}_{it}\}]$, where $\widehat{mae}_{it}$ is modeled as in section 5.2.2	95%	378.4
$[\min_i\{\hat{p}_{itq} - 1.41\sqrt{\widehat{mse}_{it}}\}, \max_i\{\hat{p}_{itq} + 1.41\sqrt{\widehat{mse}_{it}}\}]$, where $\widehat{mse}_{it}$ is modeled as in section 5.3.1	95%	413.6
$[\min_i\{\hat{p}_{itq} - 1.35\sqrt{\widehat{mse}_{it}}\}, \max_i\{\hat{p}_{itq} + 1.35\sqrt{\widehat{mse}_{it}}\}]$, where $\widehat{mse}_{it}$ is modeled as in section 5.3.2	95%	378.1
$[\min_i\{\hat{p}_{itq} - 1.57\hat{a}_{itq}\}, \max_i\{\hat{p}_{itq} + 1.57\hat{a}_{itq}\}]$, where $\hat{a}_{itq}$ is modeled as in section 5.4.1	95%	390.5
$[\min_i\{\hat{p}_{itq} - 1.16\sqrt{\hat{s}_{itq}}\}, \max_i\{\hat{p}_{itq} + 1.16\sqrt{\hat{s}_{itq}}\}]$, where $\hat{s}_{itq}$ is modeled as in section 5.4.2	95%	396.4

Table 5.7: Comparison of the different confidence intervals estimated for the whole data set

estimation and the daily average wind speed estimation. The confidence intervals

$$[\,\min_i\{\hat{p}_{itq} - 1.16 \cdot \sqrt{\hat{s}_{itq}}\},\ \max_i\{\hat{p}_{itq} + 1.16 \cdot \sqrt{\hat{s}_{itq}}\}\,] \tag{5.34}$$

and

$$[\,\min_i\{\hat{p}_{itq} - 1.57 \cdot \hat{a}_{itq}\},\ \max_i\{\hat{p}_{itq} + 1.57 \cdot \hat{a}_{itq}\}\,] \tag{5.35}$$

where the square error $\hat{s}_{itq}$ and the absolute error $\hat{a}_{itq}$ are modeled on a quarter hourly basis also show very similar results with one another, whereas their average interval range is about 12-18 MW wider than that of (5.32) and (5.33). The widest average interval range is obtained from the intervals

$$[\,\min_i\{\hat{p}_{itq} - 1.41 \cdot \sqrt{\widehat{mse}_{it}}\},\ \max_i\{\hat{p}_{itq} + 1.41 \cdot \sqrt{\widehat{mse}_{it}}\}\,] \tag{5.36}$$

and

$$[\,\min_i\{\hat{p}_{itq} - 2.05 \cdot \widehat{mae}_{it}\},\ \max_i\{\hat{p}_{itq} + 2.05 \cdot \widehat{mae}_{it}\}\,], \tag{5.37}$$

where the daily mean square error $\widehat{mse}_{it}$ and the daily mean absolute error $\widehat{mae}_{it}$ are modeled only according to the seasonal trend with autoregressive residuals.

As next this we compare the results of the cross validations of each of the confidence intervals. Figure 5.13 shows the comparison of the percentage of power data covered by each test set in the cross validation. The confidence interval (5.37) is represented by the line *KI 1* in the plot, the interval (5.33) by *KI 2*, (5.36) by *KI 3*, (5.32) by *KI 4*, (5.35) by *KI 5* and (5.34) by *KI 6*. Figure 5.14 on the other hand shows the average range of each of the intervals resulting from the cross validation of each of the test sets[4].

The results of the confidence intervals (5.33) and (5.32) show that they perform quite similarly and have the highest percentage of power data covered, which is even above 96%. However they are among the widest intervals. The intervals (5.37) and (5.36) cover above 95% of the power data, however the average range of these intervals is quite close to those of (5.33) and (5.32). The confidence intervals (5.34) and (5.35) are the ones that cover less percentage of the power data in comparison to the others, even though each test set covers more than 94% of the data. On the other hand, these two intervals have the lowest average range.

[4] The test and data sets used for this comparison are the same as mentioned in the sections above

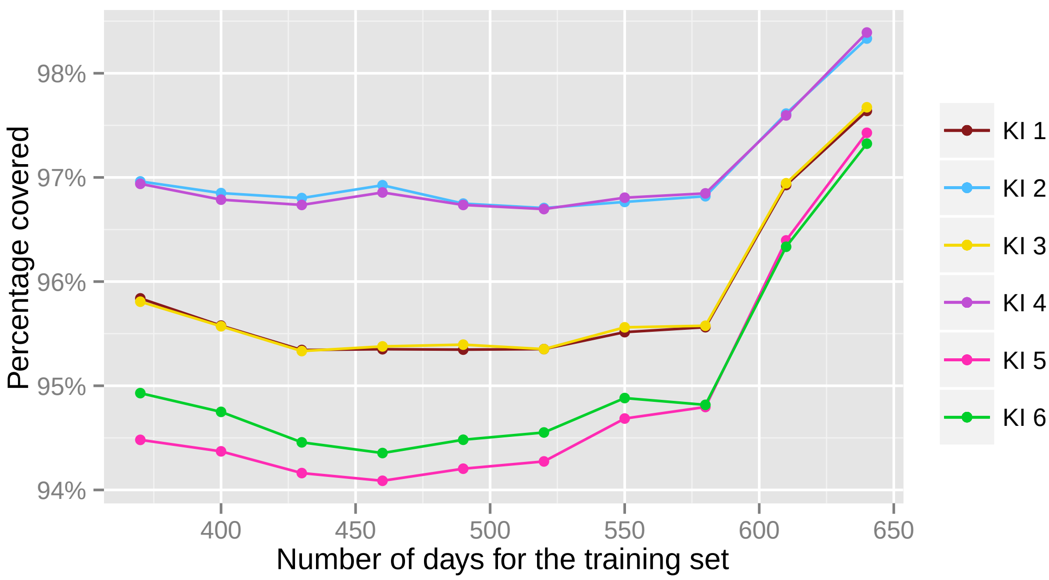

Figure 5.13: Cross validation results showing what percentage of the measured power is covered by each confidence interval.

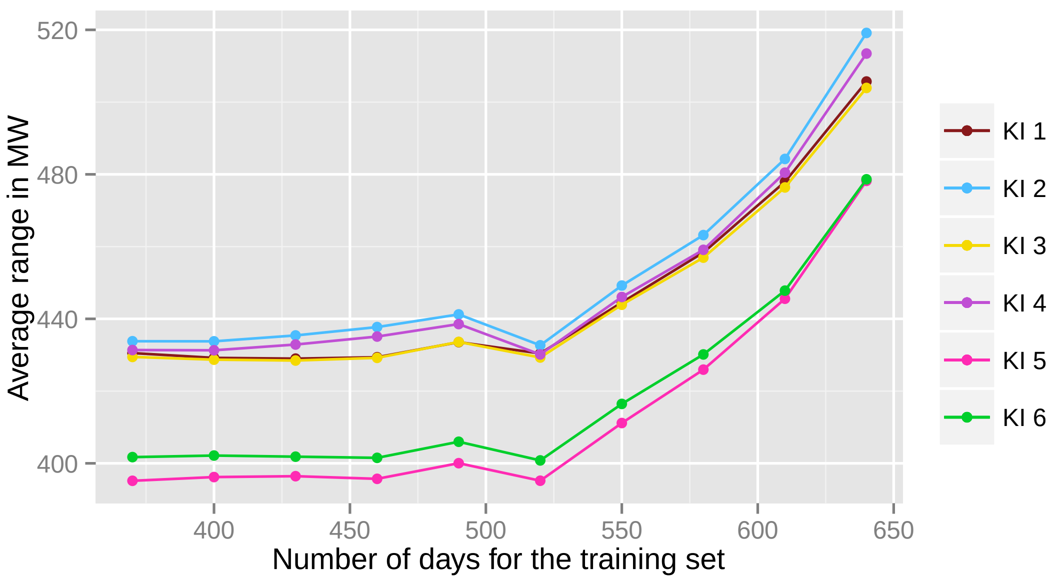

Figure 5.14: Cross validation results showing the average range of each confidence interval.

What is also of interest, is to see how the models perform according to each hour of the day. For that reason we compare the results of the confidence intervals (5.36), (5.34) with the according estimations of the (daily mean) square error as well as the interval

$$[\min_i\{\hat{p}_{itq} - 1.17 \cdot I_t \cdot \hat{\sigma}_{itq}, \max_i\{\hat{p}_{it} + 1.17 \cdot I_t \cdot \hat{\sigma}_{itq}\}] \tag{5.38}$$

where $\hat{\sigma}^2_{itq}$ is the estimated variance from the GARCH(2,1) model (refer to section 3.7.2). For this comparison the parameters of each model were estimated using a training data set of 370 days. The confidence intervals were then build for the test data set for the day ahead (t+1) given the information up to one day ago (t-1). The results of these intervals on an hourly basis are shown below.

Figure 5.15 shows how much percentage of the measured power in the test set is covered form the confidence intervals (5.38), (5.36) and (5.34). Here the results for the interval (5.38) are marked with a blue colour, the results from (5.36) in pink and the results form (5.34) in green. As it can be seen from all three models, there are differences for each time of the day. At all times the average coverage is above 90%,but the intervals seem to have a higher average coverage for the morning till afternoon hours in each case.

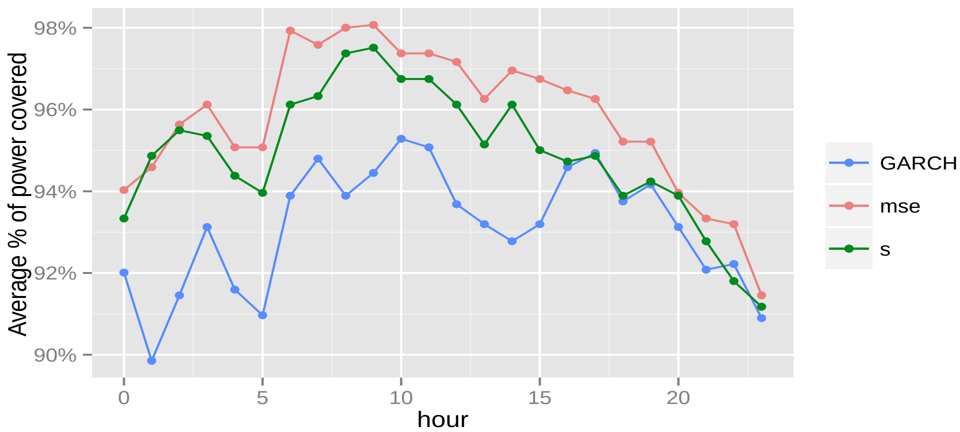

Figure 5.15: Average coverage of the confidence intervals (5.36) (in pink), (5.34) (in green) and (5.38) (in blue) for each hour of the day by using a training data set of 370 days.

On the other hand, figures 5.16, 5.17 and 5.18 show box-plots of the range of each of these three intervals for every hour of the day (here some outliers have been cut off the graphics in order to provide a better view). The interval that uses the estimation of the conditional variance from the GARCH modeling (figure 5.16) shows that the median of the interval's width increases with time. This is to be expected, as this procedure uses a forecast of 159 steps ahead, so the intervals become wider with the passing time. On the other hand, the two other intervals seem to be more stable in regard to time of the day.

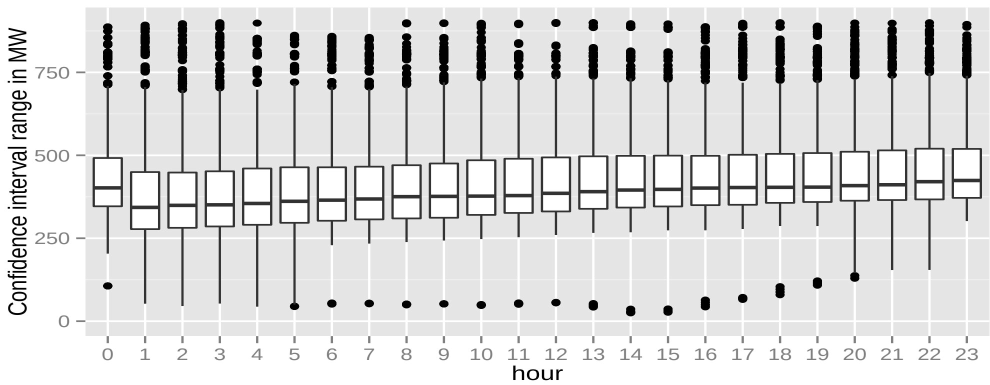

Figure 5.16: Hourly box-plots of the range of the confidence interval $[\,\min_i\{\hat{p}_{itq} - 1.17 \cdot I_t \cdot \hat{\sigma}_{itq}, \max_i\{\hat{p}_{it} + 1.17 \cdot I_t \cdot \hat{\sigma}_{itq}\}\,]$, where $\hat{\sigma}_{itq}$ is modeled according to section 3.7.2.

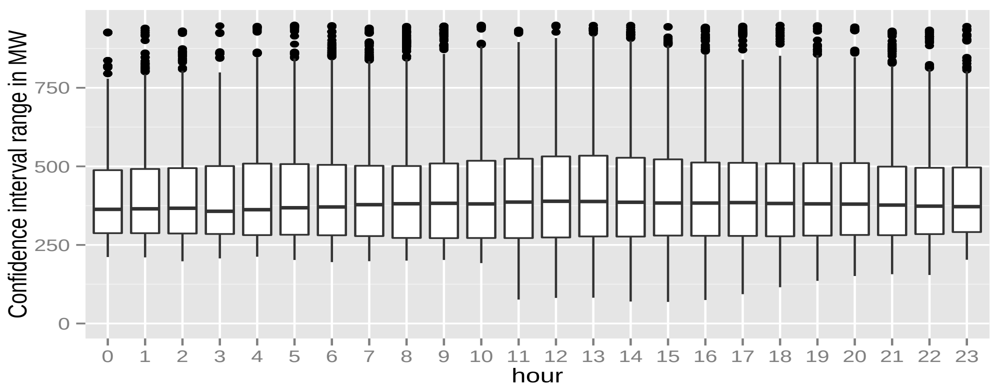

Figure 5.17: Hourly box-plots of the range of the confidence interval $[\min_i\{\hat{p}_{itq} - 1.41\sqrt{\widehat{mse}_{it}}\}, \max_i\{\hat{p}_{itq} + 1.41\sqrt{\widehat{mse}_{it}}\}]$, where $\widehat{mse}_{it}$ is modeled as in section 5.3.1

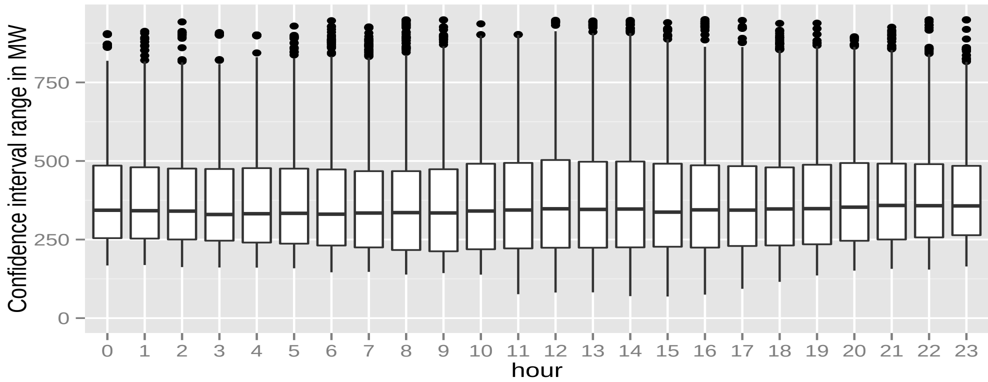

Figure 5.18: Hourly box-plots of the range of the confidence interval $[\min_i\{\hat{p}_{itq} - 1.16\sqrt{\hat{s}_{itq}}\}, \max_i\{\hat{p}_{itq} + 1.16\sqrt{\hat{s}_{itq}}\}]$, where $\hat{s}_{itq}$ is modeled as in section 5.4.2.

CHAPTER 6

Conclusion

The aim of this thesis was to analyse the errors of five different wind energy prediction models and to build a confidence interval for these errors. At first different statistical tests were done on the errors of each of the prediction models to look for different patterns. In general it was seen that the ALA models performed better (especially the second ALA model) and their errors were less volatile than those of the ECM models.

An attempt was made to analyse how the errors behave in the different wind sectors (the different directions the wind comes from). All models showed higher volatility in sectors 3, 5, 7 and 8, while the ECM models showed the highest volatility in sector 1. However the distribution of the sectors is not uniform as most data falls into sector 6 or it is not specified (sector 0), so these results are not quite significant.

The different times of the day also showed to have an effect on the errors. During the morning to midday hours the error distribution seems to be more stable and the forecasts more accurate. This information was also used to build the confidence intervals.

Using GARCH models, I tried to model the conditional variance of the errors. Fitting the errors in a GARCH(2,1) model and using this to build a confidence interval provided quite a good result. However in order to use this model in practice, a very high forecasting horizon needed to be used and this resulted in quite wide confidence intervals.

Other confidence intervals were modeled based on the daily mean absolute and square errors as well as the wind speed estimation and wind power estimation. While evaluating the performance of the different confidence intervals it could be seen that there is a trade off between the amount of data an interval covers and its average range. In this thesis we opted to build a confidence interval which covers 95% of the data. From the results presented in chapter 5, I would recommend to use the confidence interval

$$[\,\min_i\{\hat{p}_{itq} - 1.57 \cdot \hat{a}_{itq}\},\ \max_i\{\hat{p}_{itq} + 1.57 \cdot \hat{a}_{itq}\}\,] \tag{6.1}$$

where $\hat{a}_{itq}$ is modeled as shown in section 5.4.1. In praxis (refer to the cross validation results as shown in figure 5.13 and figure 5.14) this interval showed to cover less data compared to the others, however it still covered over 94% of the power data measured and it has the narrowest average range.

APPENDIX A

Abbreviations

Abbreviations used in this thesis :
ECM 1 ... refers to the first ECMWF model
ECM 2 ... refers to the second ECMWF model
ALA 1 ... refers to the first ALA model
ALA 2 ... refers to the second ALA model
ALA 3 ... refers to the third ALA model
APG ... Austrian Power Grid
MSE ... mean square error
MAE ... mean absolute error
AIC ... Akaike information criterion
OLS ... Ordinary least square
GLS ... General least square

APPENDIX B

Graphics

Appendix from chapter 2, section 2.2. Plots :

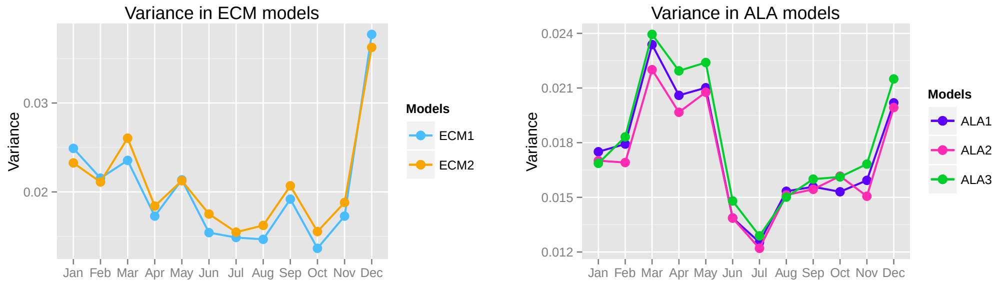

Figure B.1: Monthly mean square error

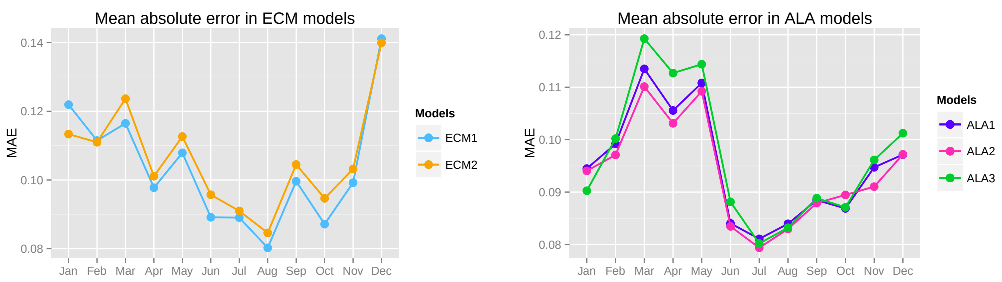

Figure B.2: Monthly mean absolute error

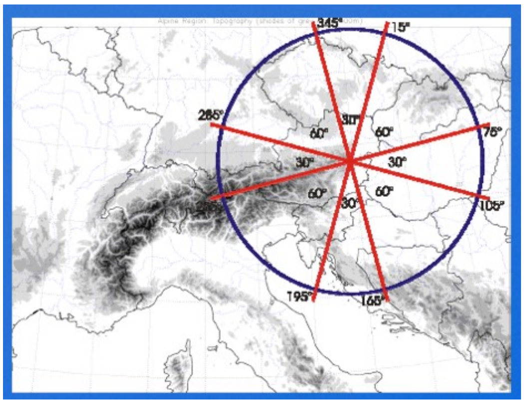

Figure B.3: The 8 wind front sectors [1]

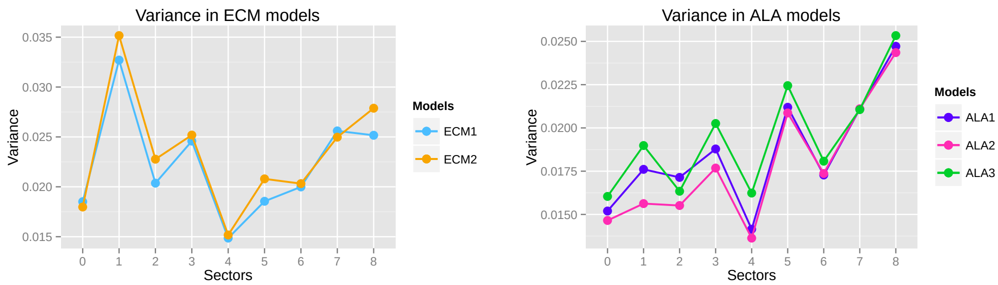

Figure B.4: Error variance among sectors

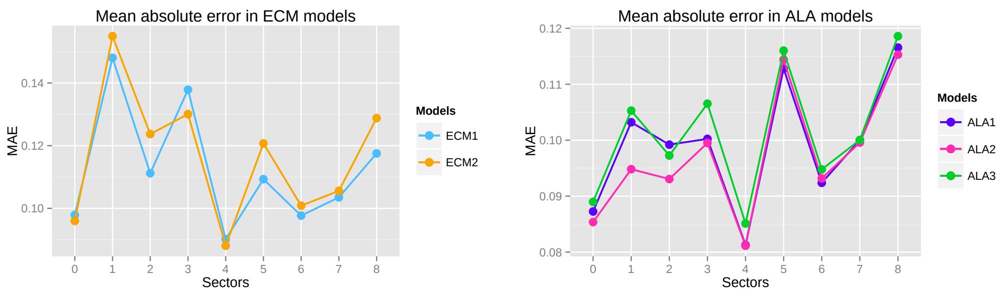

Figure B.5: Mean absolute error among sectors

[1] Graphic obtained from Austrian Power Grid

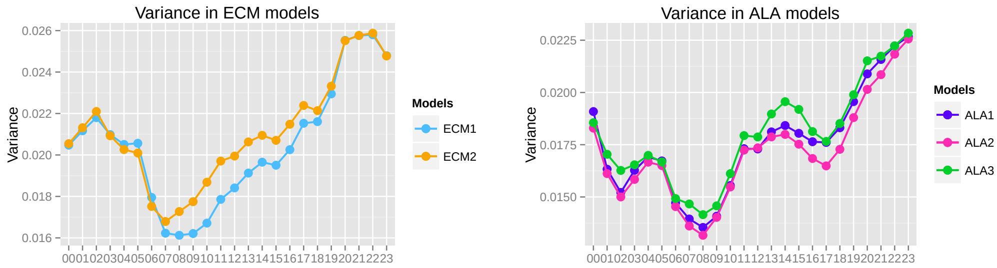

Figure B.6: Error variance according to daytime

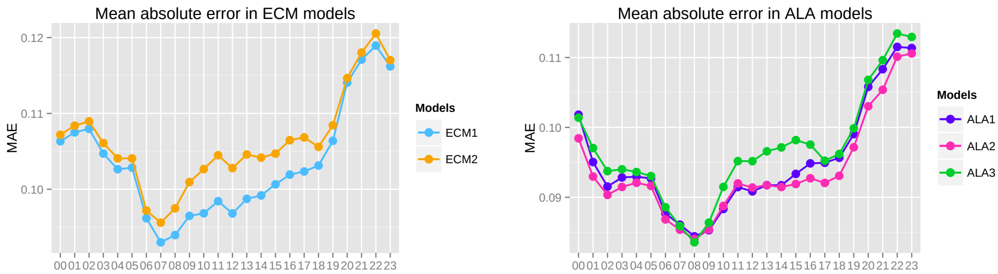

Figure B.7: Mean absolute error according to daytime

Bibliography

[1] Qiwei Yao Da Huang, Hansheng Wang. Estimating garch models: when to use what? *Econometrics Journal*, 11:27–38, 2008.

[2] Alexios Ghalanos. *rugarch: Univariate GARCH models.*, 2014. R package version 1.3-4.

[3] Alexios Ghalanos. *Introduction to the rugharch package*, 2015. R package version 1.3-1.

[4] M Michael, S MARC, and Y WAN. Statistical wind power forecasting for us wind farms. In *the 17th Conference on Probability and Statistics in the Atmospheric Sciences/2004 American Meteorological Society Annual Meeting Seattle, Washington, January11-15*, 2004.

[5] C Monteiro, R Bessa, V Miranda, A Botterud, J Wang, G Conzelmann, et al. Wind power forecasting: state-of-the-art 2009. Technical report, Argonne National Laboratory (ANL), 2009.

[6] Jesper Hybel Pedersen. Arma(1,1) - garch(1,1): Estimation and forecast using rugarch 1.2-2. 2013.

[7] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2016. R package version 3.1-128.

[8] Rob Reider. Volatility forecasting i: Garch models. 2009.

[9] David Ruppert. *Statistics and data analysis for financial engineering.* Springer, 2011.

[10] Wolfgang Scherrer. *Ökonometrie der Finanzmärkte.* Institut für Wirtschaftsmathematik, 2011.

[11] Richard Simard, Pierre L'Ecuyer, et al. Computing the two-sided kolmogorov-smirnov distribution. *Journal of Statistical Software*, 39(11):1–18, 2011.

[12] Christos Stathopoulos, Akrivi Kaperoni, George Galanis, and George Kallos. Wind power prediction based on numerical and statistical models. *Journal of Wind Engineering and Industrial Aerodynamics*, 112:25–38, 2013.

[13] James W Taylor and Jooyoung Jeon. Forecasting wind power quantiles using conditional kernel estimation. *Renewable Energy*, 80:370–379, 2015.

[14] Cort J Willmott, Scott M Robeson, and Kenji Matsuura. A refined index of model performance. *International Journal of Climatology*, 32(13):2088–2094, 2012.

[15] IG WINDKRAFT. Windkraft in zahlen. *https://www.igwindkraft.at*, 2015.

[16] Yuan-Kang Wu and Jing-Shan Hong. A literature review of wind forecasting technology in the world. In *Power Tech, 2007 IEEE Lausanne*, pages 504–509. IEEE, 2007.

[17] Yiyang Yang. Parameter estimation of garch model. 2012.