Computer Exercise 4 - Forcasting of Wind power

Frederik Boe Hüttel, s151689,
Christan Glissov, s146996,
Jonas Gottlieb Svendsen, 152992

January 4, 2019

DTU Compute - Institute for Mathematics and Computer science
Advanced Time Series Analysis 02427 January 4, 2019
Henrik Madsen & Erik Lindstrøm

Technical University of Denmark

Contents

1 Introduction

In this assignment we attempt to model and estimate the power production from wind turbines using adaptive models. The data in this exercise is from Klim, a wind farm located near Fjerritslev in the Northwest of Jutland. A new weather forecast is made every 6 hours and it consists of estimates of the wind speed, direction, and temperature for every hour in the following 48 hours. Every hour the power production from the wind farm is recorded. The main focus in this assignment will be using the forecasted wind speed to get a 1, 2 and 3 hour prediction of the power production.

The data has already been made ready for use. However a significant amount of NaNs were found in the data. To avoid excessive amount of time to fix the data a subset of the data is simply used. The subset contains 8500 sample points. The data spans from `1999-12-21 06:00:00` to `2000-12-09 10:00:00`, so approximately a full year. In Figure 1 the 1-hour forecasted data and wind power is visualised.

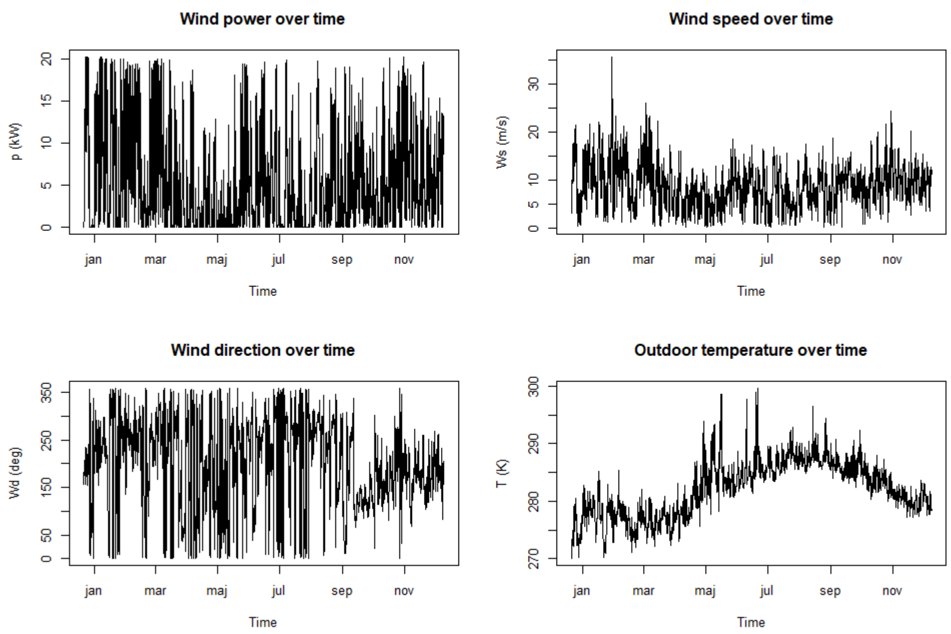

Figure 1: Plots of the different data, here for the 1-hour forecasted data over time and the wind power.

In Figure 2 the different wind speeds against the wind power can be observed. A clear non-linear power curve can be observed. Furthermore it is seen that not a lot of observations are present when wind speeds reach above 20 m/s, this is due to wind turbines will be locked when wind speeds are too high. An upper limit of the power production can also be observed, being around 20 `kW`.

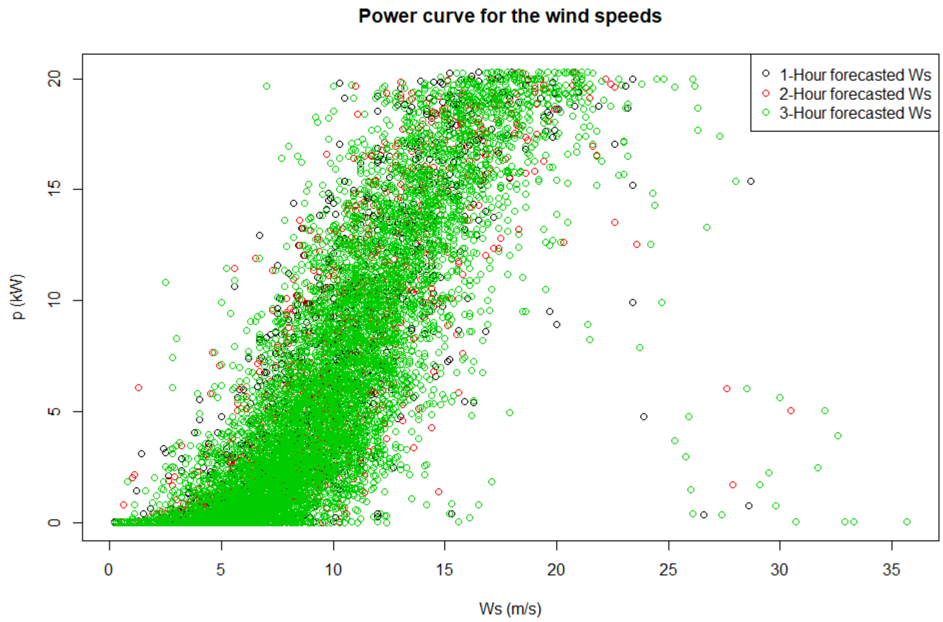

Figure 2: The 1, 2 and 3 hour forecasted wind speeds against the wind power.

A general assumption throughout the modelling is that the temperature and the wind directions does not have an impact on the wind power. This assumption might be considered naive - but was made in order to simplify the models. It is also assumed that the wind speed is the main contributor to the power generated by the wind turbines. Another unspoken assumption of the data is that the error of the estimated wind speed is Gaussian.

2 Models used for prediction

In order to predict the power production of wind farms given weather forecasts multiple models will be used and evaluated against each other. Each model used and how to verify them, will be shortly introduced here.

2.1 ARIMA - GARCH

Initially a simple ARIMA(p,q) model will be used to model the series

$$X_t = \mu + \sum_{i=1}^{p} \theta_i X_{t-i} + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} + \epsilon_t.$$

Also a ARX-model will be considered as the windspeed is assumed to have influence on the power produced - where the windpower is also dependend on the estimated wind speed.

$$Y_t = \sum_{i \in L_y} a_i \left(X_{t-m}\right) Y_{t-i} + \sum_{i \in L_u} b_i \left(X_{t-m}\right) U_{t-i} + \epsilon_t$$

From the Figure 1 the variance of the wind power appears to have some periods where it is volatile. A GARCH model will be included to model the heteroscedasticity of the variance if any is present [1]:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2.$$

. The orders of the ARIMA will be found by analysing the PACF and ACF of the data. For the GARCH the squared ACF and PACF of the residuals of the ARIMA will be analysed to determine a proper order. The software package `rugarch` will then be used to estimate the parameters of the ARIMA-GARCH model.

2.2 Adaptive Recursive Least Squares

The Adaptive Recursive Least Squares (ARLS) is a recursive extension to the ordinary least squares. The parameters of the ARLS are updated recursively based on new observations of a realisation. The parameters θ is allowed to be updated dynamically denoted $\hat{\boldsymbol{\theta}}$. The adaptive part of the model is reflected by adding a forgetting factor to the recursive least square (RLS) algorithm $0 < \lambda < 1$. The forgetting factor reduces the impact of past observations that increases the SSE by giving older observations less weight. Therefore a small λ will make the model less sensitive to past observations. This can be used as a tool to combat the influence of outliers in the model. Including the forgetting factor is based on *equation (11.27)* and *(11.28)* in [2]. This means that the parameters for the RLS algorithm $\mathbf{P}_t$ and $\mathbf{K}_t$ gets updated as, [2]:

$$\mathbf{P}_t = \frac{1}{\lambda(t)}\left(\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{X}_t\mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t}{\lambda(t) + \mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t}\right); \tag{1}$$

$$\mathbf{K}_t = \frac{\mathbf{P}_{t-1}\mathbf{X}_t}{\lambda(t) + \mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t}. \tag{2}$$

where

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{K}_t[e_t(\hat{\boldsymbol{\theta}}_{t-1})], \tag{3}$$

and

$$e_t(\hat{\boldsymbol{\theta}}_{t-1}) = Y_t - \mathbf{X}_t^T\hat{\boldsymbol{\theta}}_{t-1}. \tag{4}$$

Each weight is exponentially assigned to each previous value of the actual system, where $\lambda(t)$ are the exponential weights. The forgetting factor, λ, is found by minimising the sum of squared errors, SSE. One can make $\lambda(t)$ depend on the current observations, however this is not a good idea given there are a lot of fluctuations in the data, [1] and the forgetting factor is therefore assumed constant in this case.

2.3 Kernel Estimation

Kernel estimators can be used to solve regression problems. The kernel used is the Epanechnikov kernel, [1] (2.15) to fit the regression model.

$$K_h^{Epa}(u) = h^{-1}k^{Epa}\left(uh^{-1}\right) = \frac{3}{4h}\left(1 - \frac{u^2}{h^2}\right) I_{\{|u| \le h\}}$$

This will allow one to estimate weights of each observation, which can be used to make a fit using a weighted least squares. To find an optimal bandwidth, leave-one-out cross-validation of the MSE as suggested by [1] is used. The Epanechnikov is computational more efficient compared to the Gaussian kernel, due to the fact that some weights become less than equal to 0, and are filtered away for the weighted least square.

2.4 Stochastic Differential Equations

The final model will consist of Stochastic Differential Equations (SDE). From [3] a model using SDE can be set up without wind speed observations:

$$\begin{aligned} dX_t &= ((1 - e^{-X_t})(\rho_x \dot{p}_t + R) + \theta_x(p_t \mu_x - X_t))dt + \sigma_x X_t^{0.5} dW_t \\ dR_t &= -\theta_r R_t dt + \sigma_r dW_t \\ Y_{2,k} &= (0.5 - 0.5 tanh(\gamma_2(X_{t_k} - \gamma_3))) \frac{\zeta_3}{1 + e^{-\zeta_1(X_{t_k} - \zeta_2)}} + \epsilon_{2,k} \end{aligned} \tag{5}$$

The definitions of each parameter and variable can be seen in Table 1

γ_2	Transition to cut-out parameter
γ_3	Cut-out wind speed
$\zeta_{1,2,3}$	Power curve shape parameters
σ_r	Rate of diffusion for R
θ_r	Rate of reversion to mean for R
ρ_x	Factor normalizing the change in wind speed prediction, $\dot{p}_t$
μ_x	Factor normalizing the wind speed prediction, p_t
θ_x	Rate of reversion to mean for X
σ_x	Rate of diffusion for X
X_t	Actual wind speed at time t
R_t	Auxiliary stochastic variable determining X_t
Y_{2,t_k}	Observed wind power at time t_k
$\epsilon_{2,k}$	Wind power observation error

Table 1: Definition of each parameter and variable used in Equation 5

To estimate the SDE the software package for `R`, `CTSM-R` [4], will be used. Before this is possible it is necessary to rewrite Equation 5 to avoid eligibility errors from the `CTSM` package. First the diffusion term dW_t needs to be isolated. This is done by using a `Lamperti` transformation [5]. Given a SDE in the form of:

$$dX_t = f(X_t)\, dt + g(X_t)\, dB_t \tag{6}$$

Then the `Lamperti` transformed coordinate is given by

$$h(x) = \int^x \frac{1}{g(v)} dv \tag{7}$$

and by `Ito`'s lemma, the transformed SDE is

$$dY_t = \left[\frac{f\left(h^{-1}(Y_t)\right)}{g\left(h^{-1}(Y_t)\right)} - \frac{1}{2} g'\left(h^{-1}(Y_t)\right) \right] dt + dB_t \tag{8}$$

Doing this yields the following result for dX_t:

$$dX_t = (\frac{(1-e^{-X_t})(\rho_x \dot{p}_t + R) + \theta_x(p_t\mu_x - X_t)}{\sigma_x\sqrt{X_t}} - \frac{\sigma_x}{4\sqrt{X_t}})dt + dW_t \tag{9}$$

Furthermore `CTSM` does not support the function $tanh(x)$, [4]. Therefore the expression is rewritten to:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$$

The SDE is now ready to be estimated with the `CTSM` package.

2.5 Verifying Forecasts

To look at the accuracy of the models the root mean squared error (RMSE) and the mean absolute error (MAE) will be used:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}, \tag{10}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}(\left|Y_t - \hat{Y}_t\right|), \tag{11}$$

where Y_t is the observation, $\hat{Y}_t$ the prediction and n the number of observations. A perfect forecast will produce a score of 0.

To measure the skill the continuous ranked probability score (CRPS), will be used. In this assignment it is assumed that the forecast distribution is Gaussian, with mean μ and variance σ^2, then the CRPS can be found by[6]:

$$CRPS(\mathcal{N}(\mu,\sigma^2),y) = \sigma\left(\frac{y-\mu}{\sigma}\left[2\psi\left(\frac{y-\mu}{\sigma}\right)-1\right]+2\varphi\left(\frac{y-\mu}{\sigma}\right)-\frac{1}{\sqrt{\pi}}\right), \tag{12}$$

where $\psi(\bullet)$ and $\varphi(\bullet)$ are the CDF and PDF respectively of the standard Gaussian distribution [6].

3 Results

As mentioned earlier a focus on the wind speed predicting the wind power will be made. For modelling, the first 7500 observations will be used and for evaluating the last 1000 observations will be used from the subset of data used.

3.1 ARIMA-GARCH

The first model to estimate is the ARIMA-GARCH model. By analyzing the ACF and PACF an optimal mean model was found to be an ARX(2) model with 1-hour prediction of the wind speed as the exogenous variable. Also looking at the squared residuals of the ARX(2) model in order to determine the order of the GARCH model. From Figure 3 the squared residuals can be seen to have significant lags and it is clear that heteroscedasticity is present in the residuals, suggesting a GARCH component is needed.

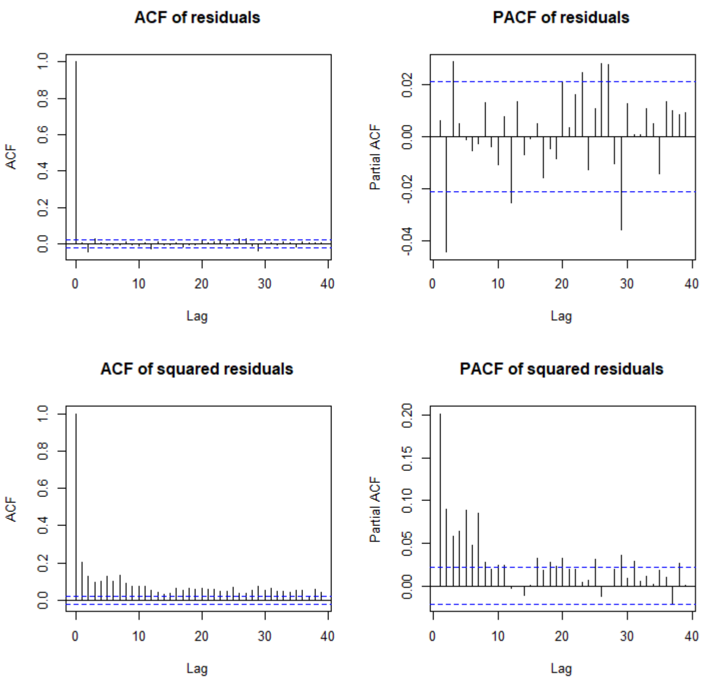

Figure 3: Analysis of residuals for ARX(2) model.

Therefore the conditional variance should be taken into account. This is done using GARCH models, which, based on the squared residuals, was estimated to be a GARCH(1,1). The best model was hence an ARX(2)-GARCH(1,1) model.

Table 2: Estimated parameters of the ARX(2)-GARCH(1,1) model using 1-hour wind speed forecasts.

μ	θ_1	θ_2	xregWS	ω	α_1	β_1
0.4036694	1.1686054	-0.2248519	0.0508663	0.0231765	0.2472544	0.7517456

Looking at Figure 4

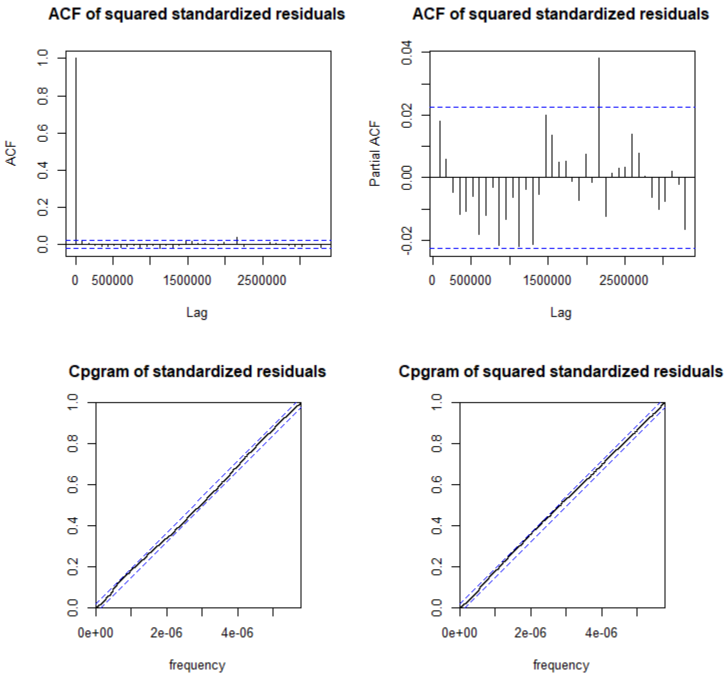

Figure 4: Analysis of residuals for ARX(2)-GARCH(1,1) model.

It can be seen that the residuals no longer have significant squared autocorrelation. The cumulative periodogram also looks very good.

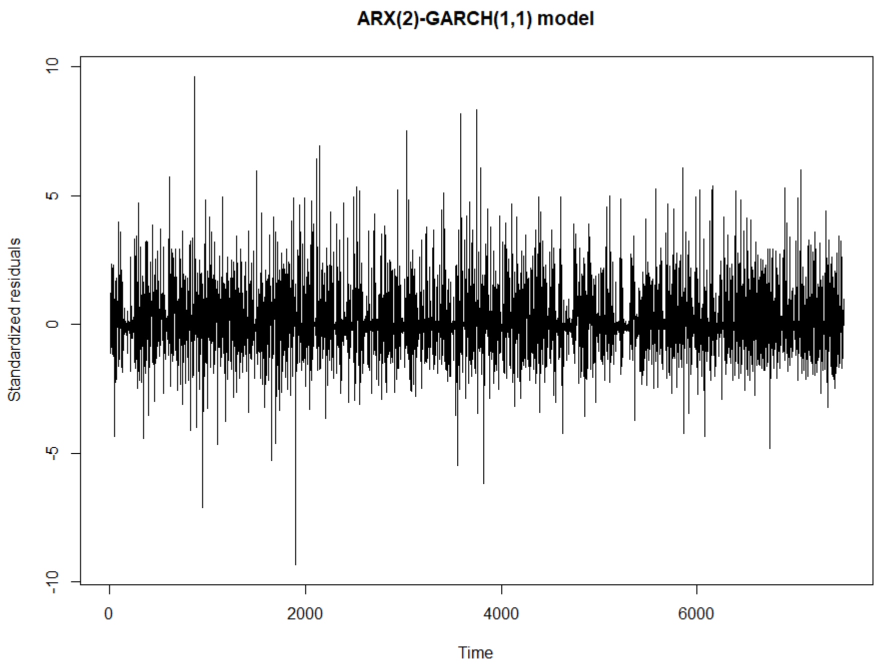

Figure 5: Residuals of ARX(2)-GARCH(1,1) model.

The standardized residuals look random with no clear trends of heteroscedasticity. The model seems to be valid. A snippet of the model used for 1-hour prediction of wind power based on the test set and also for the fitted values can be seen in Figure 6

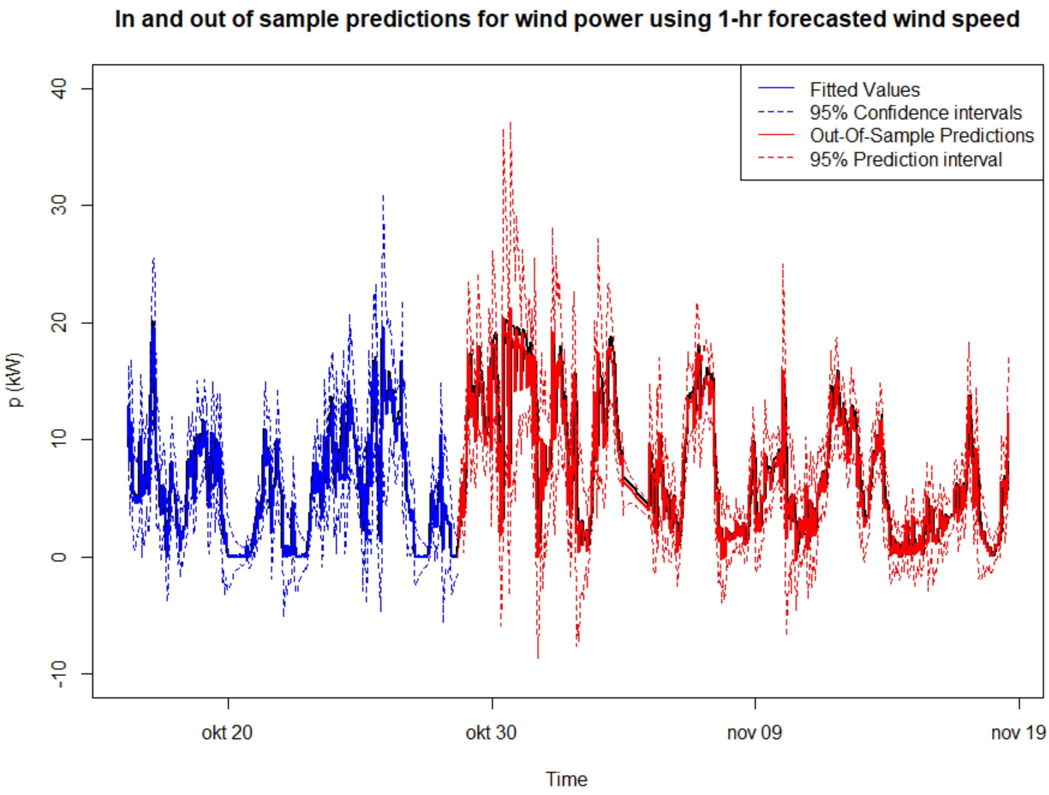

Figure 6: Observations (black), fitted values (blue) and out-of-sample 1 hour predictions (red) using a rolling forecast from the ARX(2)-GARCH(1,1) model. Not all data is used, which explains the shorter time period of the x-axis.

The out-of-sample predictions in Figure 6 are based on rolling the model through 1-hr predictions for the entire test set (1000 observations). After this the model is evaluated as explained in subsection 2.5. Doing this with each of the wind speed predictions, the following results are found:

Table 3: Verification estimates of ARX(2)-GARCH(1,1)

ARX(2)-GARCH(1,1)	RMSE	MAE	CRPS
WS 1-Hour Prediction	0.6024	0.4554	0.4734
WS 2-Hour Prediction	0.6024	0.4550	0.4734
WS 3-Hour Prediction	0.5977	0.4512	0.4720

From Table 3 one can see that the error doesn't change much for each wind speed prediction. The wind power output is highly dependent on the wind speed as was seen in Figure 2, the further away from the observed measurement we are, the higher is the likelihood that the wind speed has changed, which adds further uncertainty to the predictions. However the results in Table 3 indicates that in general the wind speed forecasts does not change a lot over the 3-hour periods.

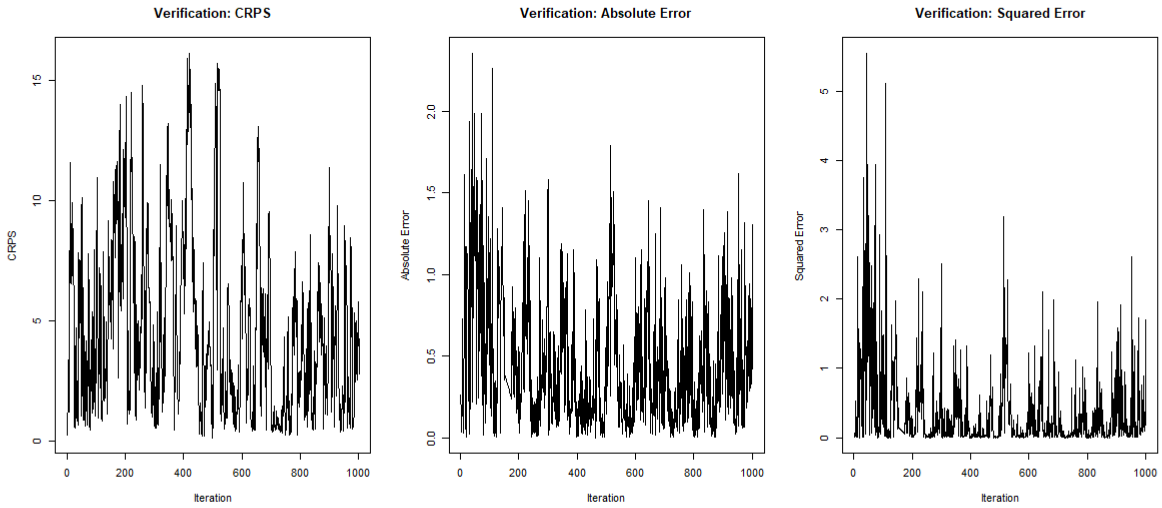

Figure 7: Different error measures visualised for the ARX(2)-GARCH(1,1) model.

Finally looking at Figure 7 it can be seen that the model is quite bad at estimating high wind power. Comparing Figure 6 to the error measurements, it is seen that high errors are given at periods of high wind power. Especially for RMSE which punishes large errors a lot more, due to the squared term. One reason for the errors might be that there is a lack of data when reaching high wind speeds, increasing the uncertainty. However especially at peaks of the wind power the model seem to under-estimate, indicating that some factors are not taken into account. This will be discussed later.

3.2 ARLS

The ARLS is an online adaptive method. It is highly adaptive due to its recursive nature and forgetting factor λ. First a model is made for 1-hour wind speed forecasts. When optimising the forgetting factor using the `optim` function in `R` and the sum of squared residuals. The optimal λ is found to be 0.04835. This gives a memory of:

$$T = \frac{1}{1 - 0.04835} = 1.05081 \tag{13}$$

This means the prediction will be highly reliant on recent observations. In Figure 8 the power curve fit of the ARLS can be seen.

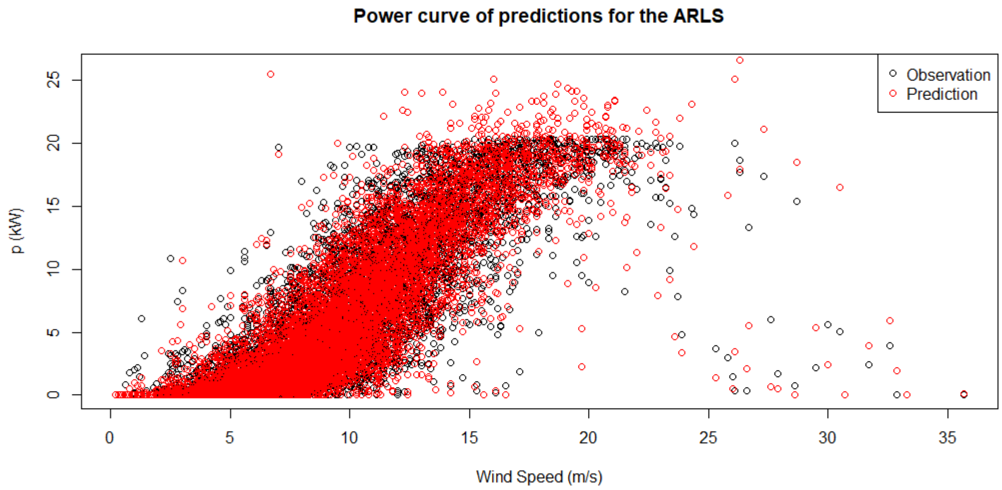

Figure 8: Power curve of the wind power against the 1-hour forecasted wind speed.

It is seen that ARLS seem to overestimate the wind power at times, making very large predictions. In Figure 9

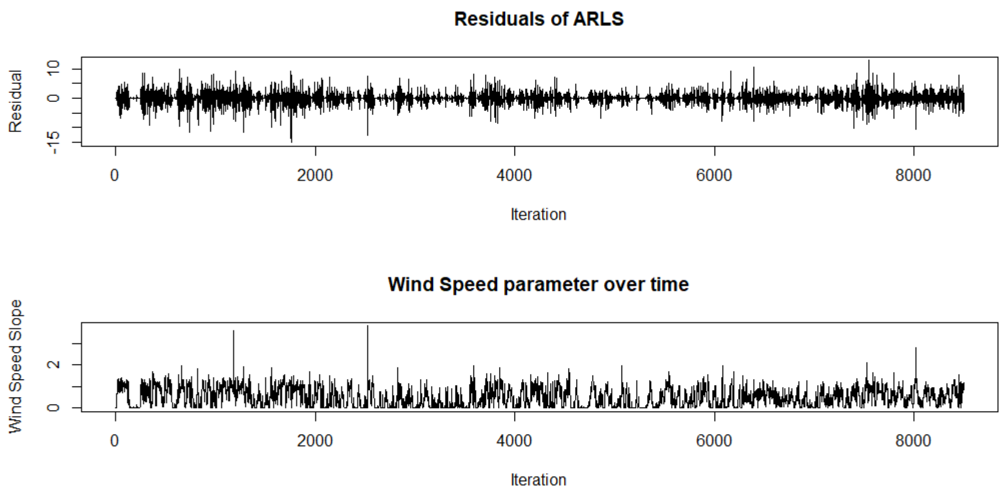

Figure 9: Residuals of the ARLS model and the 1-hour forecasted wind speed slope of the model.

From Figure 9 it is noticed that there are a few spikes in the estimated slope of the model. This is usually due to large residuals coming from outliers and only being depend on the most recent observation. This indicates that there are room for improvements in the model. Finally from Figure 10

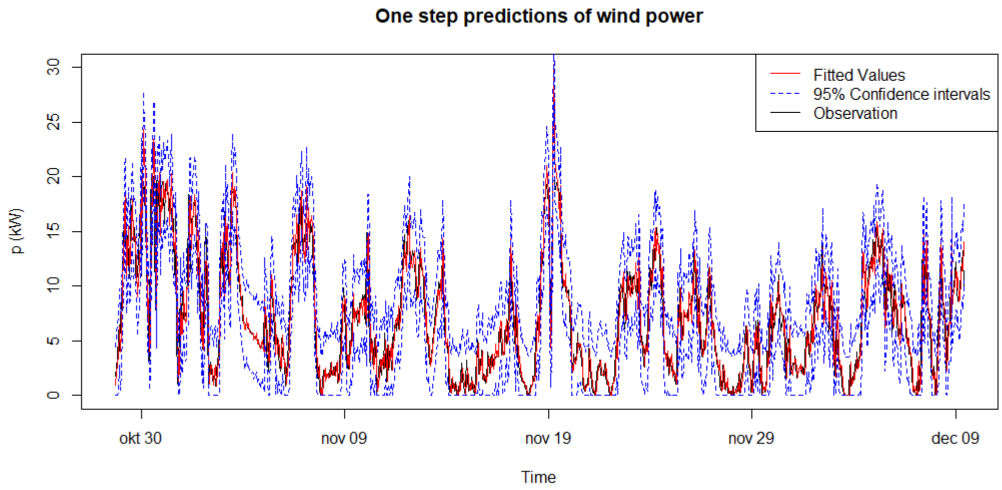

Figure 10: Wind power prediction over time of the ARLS model. Only the test data set is shown here, meaning the last 1000 observations are used.

It can be seen that the model consistently over-estimate the wind power at peaks. A clear indicator that something is missing from the model or that the uncertainty in the 1-hour forecasted wind speed is not being taken into account. The errors of the model is given in Table 4 and the forgetting factor for each model too.

Table 4: Verification estimates of ARLS and λ values

ARLS	RMSE	MAE	CRPS	λ
WS 1-Hour Prediction	1.98884	1.36911	1.03360	0.04835
WS 2-Hour Prediction	1.98657	1.37346	1.03683	0.104247
WS 3-Hour Prediction	1.96168	1.34768	1.02022	0.100529

The findings in Table 4 is consistent what was previously found. The errors do not change a lot whether one is using the 1, 2 or 3 hour forcasted wind speed predictions.

3.3 Kernel

A non-parametric model based on the Epanechnikov kernel (2.15) [1] is considered. The model will be using the forecasted wind speed. The optimal span width is found by using leave-one-out cross-validation. Due to the computational demanding nature of leave-one-out cross-validation only a subset of 1500 observations are used. This results in the plot seen in figure 11 and an optimal span of 0.3.

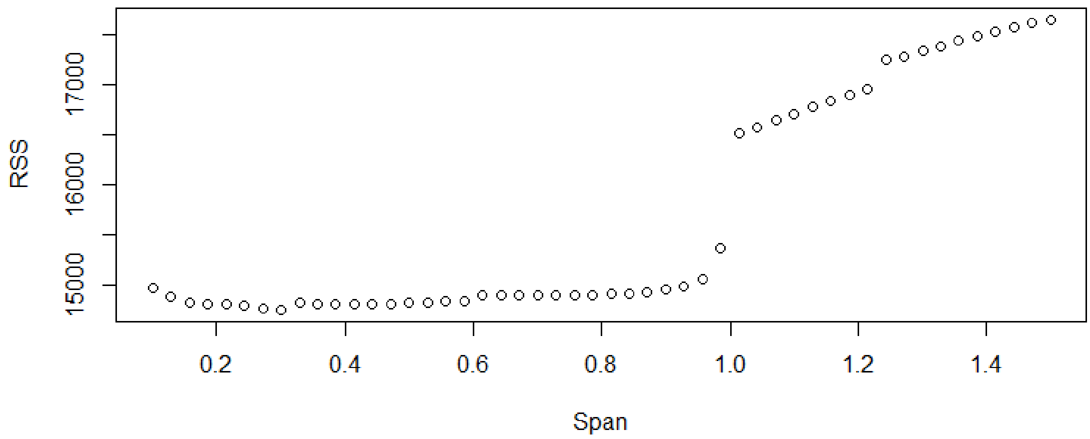

Figure 11: Plot of the optimal span width for the Epanechnikov kernel

Fitting the model on the training data set and using it to predict the test data results in the plots below. The first plot shows a zoomed in view of the fitted values with a 95% confidence interval and the prediction with a 95% prediction interval. The plot clearly shows that the model is not performing particularly well. The model follows the general trend but is almost always off by a relatively large margin.

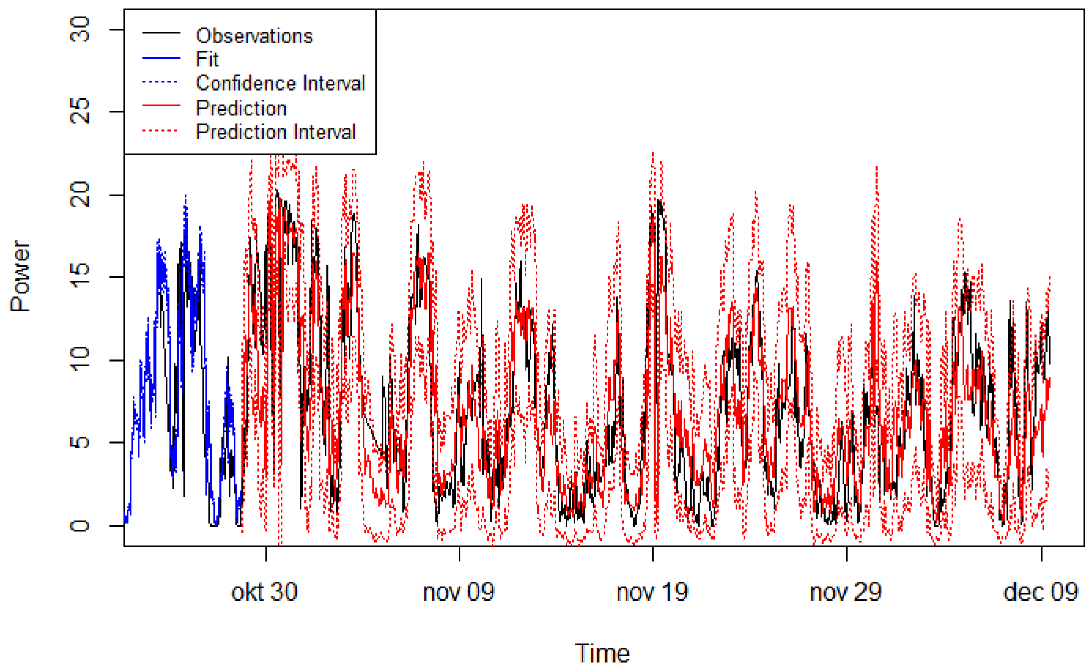

Figure 12: Plot of fit and prediction of wind power using the 1 hour wind speed prediction

In the two plots below the fitted power production and the prediction power production is plotted as a function of the 1-hour ahead wind speed. The plots makes it very obvious what causes the bad performance seen in figure 12. We see that the model has found an average power production for the given wind speed it uses this for fitting and predicting. This curve is the power curve for the power production. At such the model only finds the most general trend in the data but not any of the finer details.

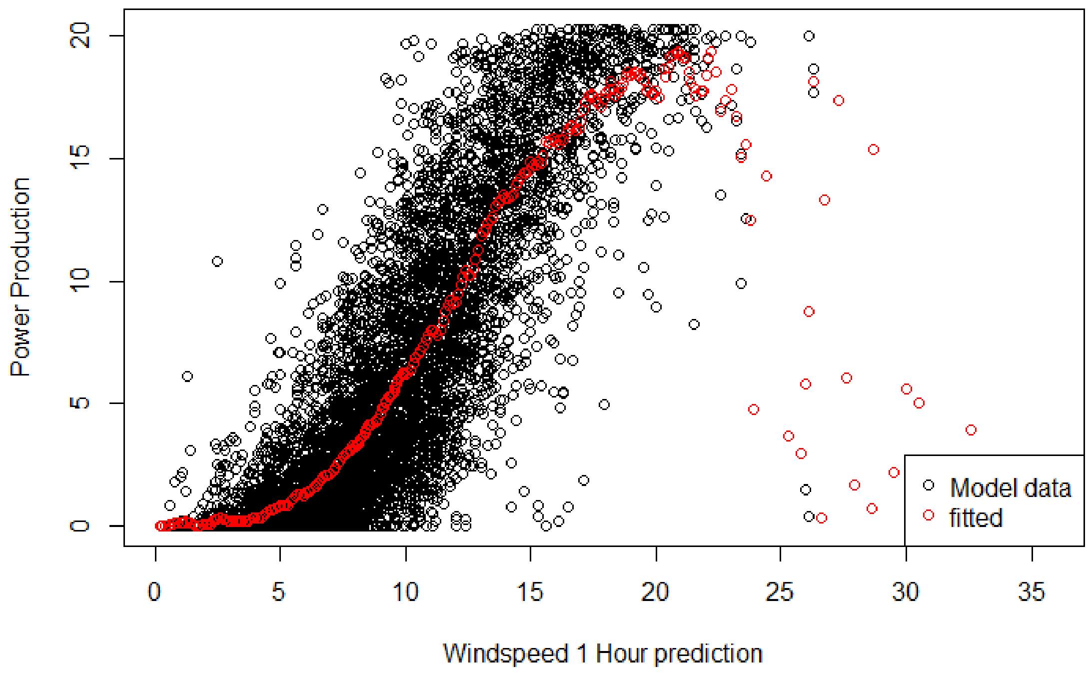

Figure 13: Plot of the fit of power production by the 1-hour ahead wind speed

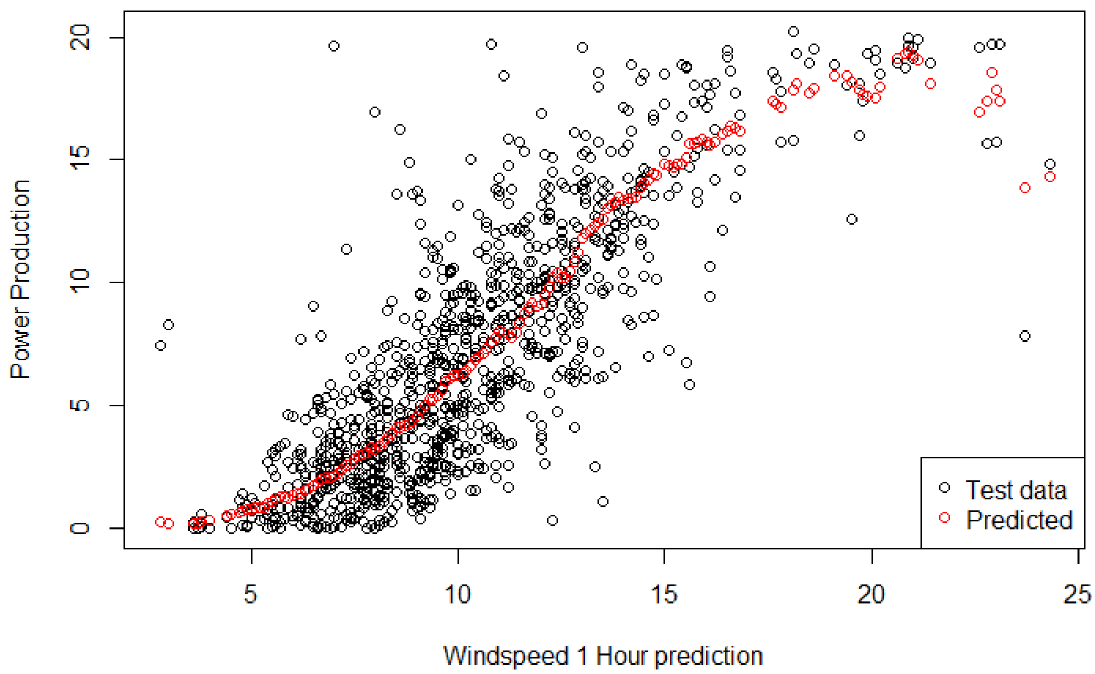

Figure 14: Plot of prediction of wind power as a function of wind speed

In the table below the verification estimates can be seen for the kernel model. It supports what the plots above showed, that the model is performing poorly and does so for all three wind speed predictions.

Table 5: Table of verification estimates for the Epanechnikov Kernel model

Epanechnikov Kernel	RMSE	MAE	CRPS
1-Hour Prediction	2.92	2.16	2.05
2-Hour Prediction	2.95	2.18	2.06
3-Hour Prediction	3.01	2.20	2.07

In the plot below the residuals for model using 1-Hour predictions of wind speed are seen. It is immediately apparent how large the residuals are and that huge spikes are visible. From the results shown from this model it is very apparent that the model is performing very poorly.

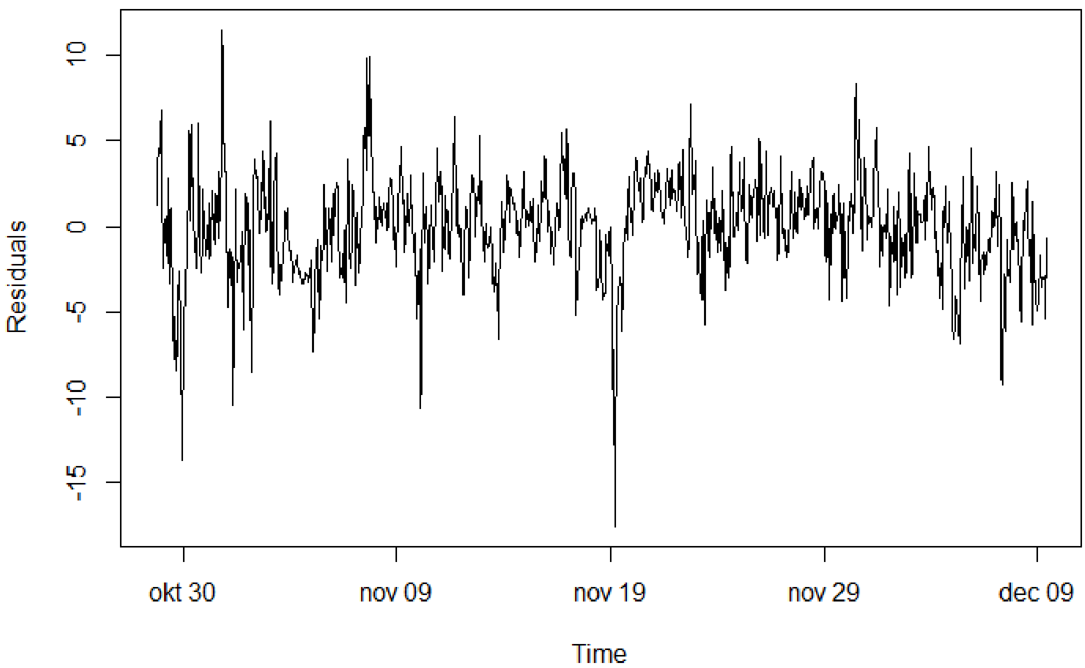

Figure 15: Plot of the residuals for the kernel model. For the model using 1-Hour predictions of wind speed.

3.4 SDE

The model introduced in subsection 2.4 and [3] is used with the `CTSM-R` package. To do this starting values have to be applied for each of the variables. One have to be careful of the initial guesses, otherwise floating point exceptions might occur or state covariance matrix might not be positive definite. Finally since the SDE in Equation 5 uses the time t as input the POSIX format has to be made numeric and turned into seconds, this is easily done in `R` by using the `as.numeric` function and dividing by 3600. The estimated parameters of the CTSM model can be seen in Table 6

Table 6: Summary statistics of the CTSM model.

	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
X_0	4.347194	0.520859	8.346205	0
R_0	0.985789	0.132979	7.413104	0
$e_{2,k}$	4.259091	0.019653	216.718735	0
γ_2	0.642911	0.018296	35.140196	0
γ_3	37.447698	0.732960	51.091024	0
μ	1.193384	0.006353	187.836308	0
ρ	0.121935	0.003653	33.377140	0
σ_r	0.930259	0.009283	100.216152	0
σ_x	0.353329	0.000940	376.018565	0
θ_r	3.079944	0.009086	338.973175	0
θ_x	0.778917	0.001033	754.336332	0
ζ_1	1.531993	0.240714	6.364377	0
ζ_2	4.856080	0.157011	30.928196	0
ζ_3	0.999467	0.000927	1078.155485	0

As seen all parameters are highly significant. It was noticed that a lot of the ζ and γ values were highly correlated to each other. Furthermore ζ_3 consistently seemed to hit its upper boundary. We were not able to fix these two problems, since the model was highly sensitive, especially to ζ_3, which seems to act as a scaling factor, making the model unable to converge to a solution if set too high. Due to the predictions being extremely bad, so bad, that including them in the assignment would not make sense, the SDE model and its results will be omitted. Clearly something went wrong during the modelling process of the SDE. The SDE model will briefly be discussed in this subsection 4.4.

3.5 Comparison of results

For the 1,2 and 3 hour ahead prediction the ARX-GARCH model perfomed the best.

Model	RMSE	MAE	CPRS
ARX(2)-GARCH(1,1)	0.60	0.46	0.47
ARLS	1.99	1.37	1.03
Kernel	2.92	2.16	2.06

Table 7: 1 hour ahead predictions

Model	RMSE	MAE	CPRS
ARX(2)-GARCH(1,1)	0.60	0.46	0.47
ARLS	1.99	1.37	1.04
Kernel	2.95	2.18	2.06

Table 8: 2 hour ahead predictions

Model	RMSE	MAE	CPRS
ARX(2)-GARCH(1,1)	0.60	0.45	0.47
ARLS	1.96	1.34	1.02
Kernel	3.01	2.20	2.07

Table 9: 3 hour ahead predictions

It's assumed that the ARX-GARCH model is well suited for this problem. The relationship between the wind speed and the wind power, which the ARX-GARCH model captures by having the external regressors in the model. Another advantages of the ARX-GARCH model is the incorporation of the heteroscedasity, which allows for the time varying variance, which results in prediction varying in size.

4 Discussion

4.1 Adaptive

One of the drawbacks of the ARX-GARCH model is that the model is not as adaptive as the other models. The most adaptive model is the ARLS along with the SDE model (however SDE did not give any fruitful results), as both of these models allow for new observations to update the estimated parameters using the same principles of the Kalman gain, where the parameters are updated based on the 1-step prediction error. This would also imply that an Extended Kalman filter would be a good model for the predictions.

4.2 Assumptions

As mentioned earlier modelling the temperature and wind directions is assumed not to have an impact on the wind-power. However these assumptions are considered very naive. If the direction of the of wind is used to make directions such as

Wind-direction	Direction
0-90	1
90-180	2
180-270	3
270-360	4

Table 10: Group of direction based on wind direction

the following power curves can be observed.

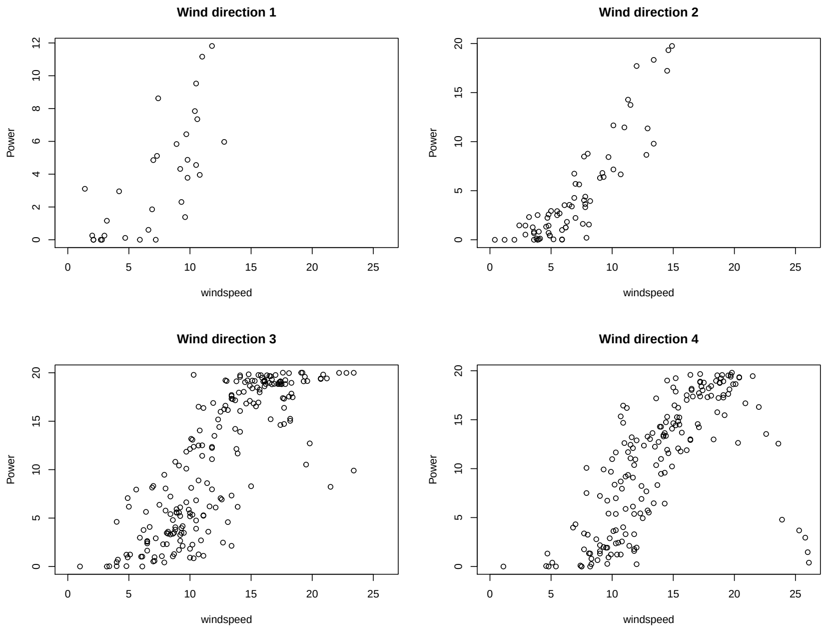

Figure 16: Power production based on estimated wind power for 4 different directions

This figure show that there are some differences in the direction the wind is coming from. Some directions to only have smaller wind speeds. This could be assumed to have an influence on the estimated power production. But on the other hand it could also just be due to the nature of the environment around the wind mills. Adding more factors might also make some of the models improve, such as improving estimation of high wind power, which seemed to be a weakness of especially the ARX-GARCH model. Either way the wind direction was not considered throughout this project, but should be further investigated. Just as outdoor temperature should, temperature does affect air densities and this might also show up in the residuals.

4.3 Uncertainty in dependent variables

Throughout this project the estimated wind speed is used in order to estimate the power production. Since the wind speed is an estimation of the actual wind speed at the desired time. This uncertainty can have a lot of impact on the estimated power output as the power curve is a non-linear function. Small variance in the estimate wind speed around the 7 to 15 m/s will make the prediction interval for power output to be slightly wider than the usual predictions interval. To model this uncertainty SDE was assumed to be a good model - since the observation equations allow for uncertainty in the observable values. Then the actually wind speed is incorporated in the states of the SDE along with the actual wind power.

Both the uncertainty and the assumptions mentioned above could allow for significant improvements of the models.

4.4 Future works

Future works would include percentile regression in order to make probabilistic forecasting of the wind power in order to determine what is the expected distribution of the forecasts.

Another thing would be to make the SDE model work. After all it was the best model to predict wind output in the following paper, [3], furthermore a case for the SDE model is also given in subsection 4.3. Convergence was not a problem when given correct initial values and boundaries. However the result was strange, perhaps due to the transformation of the observation equation Y or the boundary limits not set properly.

5 Conclusion

As was shown in subsection 3.5 the best model seems to be the ARX-GARCH model. However based on the assignment and the lack of adaptiveness of ARX-GARCH model, then the ARLS might be a suitable candidate for wind power prediction as well. It did not seem to matter much whether a 1, 2 or 3 hour wind speed forecast was used. However as was discussed in section 4 the models did not give an optimal prediction, this might be due to the over-simplicity of the models only relying on wind speeds. Especially at peaks of power outputs the models seem to fall short, one should therefore be vary of trusting the models. More complex models should be investigated for better inference and forecasting.

References

[1] J. H. Henrik Madsen, Modelling Non-Linear and Non-Stationary Time Series. 2006.

[2] H. Madsen, Time Series Analysis. Chapman & Hall/CRC, Taylor & Francis Group, 1 ed., 2008.

[3] J. E. B. Iversen, J. M. Morales González, J. K. Møller, P.-J. Trombe, and H. Madsen, "Leveraging stochastic differential equations for probabilistic forecasting of wind power using a dynamic power curve," Wind Energy, vol. 20, no. 1, pp. 33–44, 2017.

[4] P. B. J. K. H. M. Rune Juhl, Niels Rode Kristensen, "Ctsm-r user guide," 2013.

[5] U. H. Thygesen, Diffusion and Stochastic Differential Equations. 2018.

[6] D. S. Willks, Statistical Methods in the Atmospheric Sciences. Elsevier, Third edition ed., 2011. Chapter 8.4-8.5, ISBN 978-0-12-385022-5.