Statistical Modelling - Assignment 1

s164180 Qahir Yousefi
s164217 Jonas Juhler-Nøttrup
s164224 Simon Tommerup

October 2018

Descriptive statistics

The following section will briefly explain the nature of the data set and present some summary statistics of the data set, as well as some graphical presentation of the data set.

Data preprocessing

The given data set consists of measurements of the average hourly production of windpower between midnight and 18.00 on the Tunø-Knob wind power plant located on Samsø over 304 days.

Also the data set contains two meteorological variables, namely the daily average wind speed and wind direction in the same time interval from midnight to 18.00.

Data normalization

The power plant has fixed maximum production capacity of 5000 kW and thus every power observation must be in the range from 0-5000 kw. These observations p have been normalized, i.e. every observation is scaled to be strictly between 0 and 1, by the equation

$$\hat{p} = \frac{p}{\text{maximum capacity}}$$

This ensures that modelling with the beta distribution makes sense as this is defined on the interval [0,1].

Summary statistics

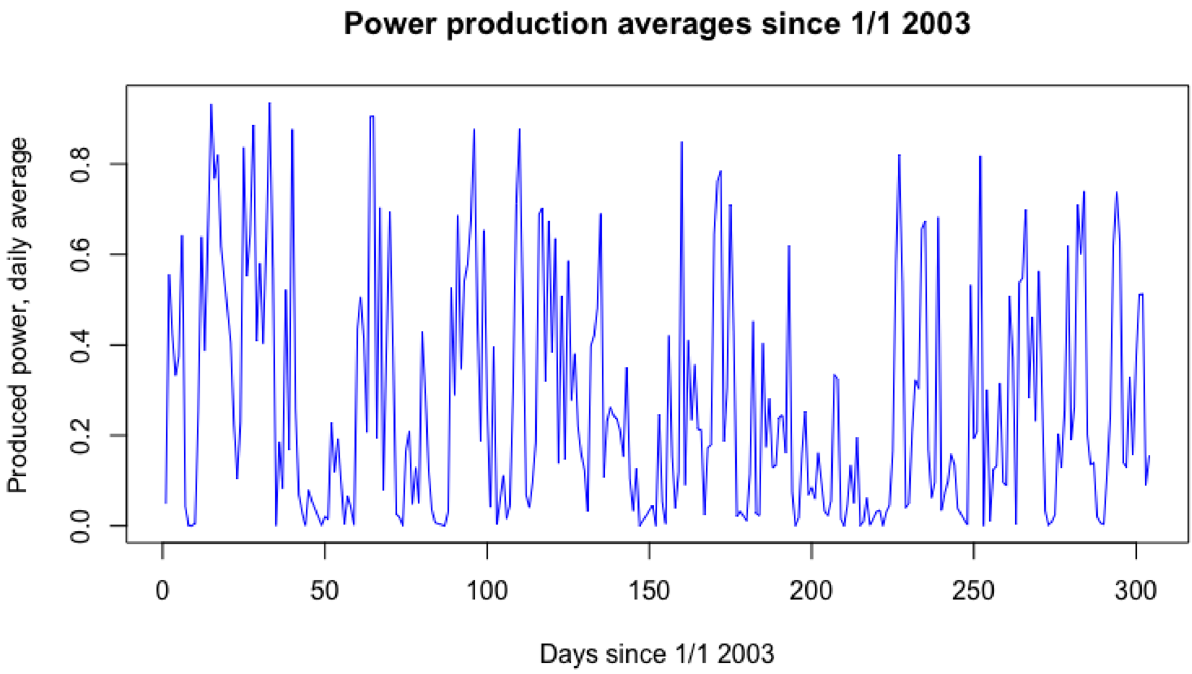
The following table contains the found summary statistics for the three variables power production, wind speed and wind direction:

parameter/variable	Power production,]0,1[	Wind speed	Wind direction
$\hat{\mu}$	0.2762392	9.111685	3.60239
$\hat{\sigma^2}$	0.06565834	20.78146	2.991617
$\hat{std}$	0.2562388	4.558669	1.729629

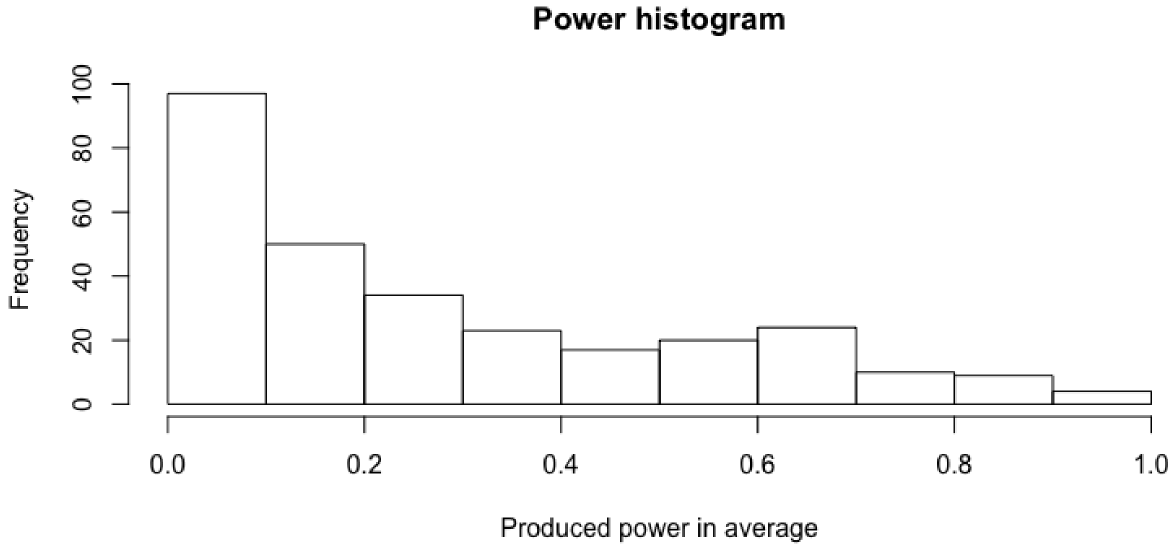
The table suggests that in average the power plant produces under one third of the maximum production capacity.

Graphical presentations

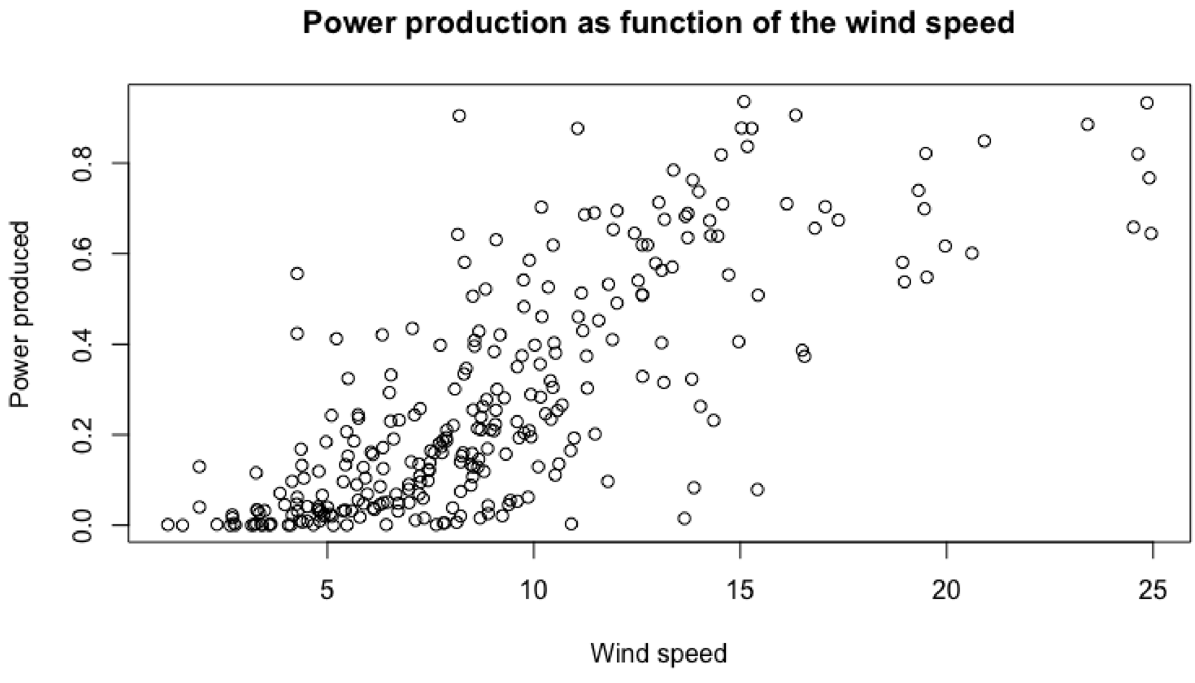
The following plot depicts the power production averages over the 304 days of measurements:



The following histogram shows the frequencies of the observations of the average power production:



In this project, we're interested in making a model that than make predictions concerning the power production. One would suspect that wind speed is a possible predictor of the power production. The following scatter plot shows the average power production as a function of the wind speed:



Later on in this project a regression model will be fit to the function above.

Simple models

Different probability density models were fitted to the wind power and wind speed data. The maximum likelihood estimates were found using the general-purpose optimization function optim() provided from R stats-library. The maximum likelihood estimates for the parameters are listed in the following table:

PDF/variable	**Wind power**	**Wind speed**
Beta	$\hat{\alpha} = 0.5571045$, $\hat{\beta} = 1.4918276$	n/a
Gamma	$\hat{\alpha} = 0.6925961$, $\hat{\lambda} = 0.3988331$	$\hat{\alpha} = 4.156440$, $\hat{\lambda} = 2.191973$
Lognormal	$\hat{\mu} = 0.2763036$, $\hat{\sigma} = 0.2557868$	$\hat{\mu} = 9.108029$, $\hat{\sigma} = 4.550341$

The following approximate likelihood confidence intervals are calculated from the inverse of the Hessian matrix for the various likehood functions added to the squareroot of the exponential of the χ^2-distrubution with two degrees of freedom at the optimum points:

Model	Parameter 1 $(\hat{\alpha}, \hat{\mu})$ CI	Parameter 2 $(\hat{\beta}, \hat{\lambda})$ CI
Beta - Wind power	[-0.07410517, 1.18831418]	[-0.6050203, 3.5886755]
Gamma - Wind power	[-0.001557673, 1.386749896]	[-0.1659359, 0.9636020]
Lognormal - Windpower	[-0.02514402, 0.57775130]	[0.04263961, 0.46893403]
Gamma - Wind speed	[1.896032, 6.416849]	[0.8846222, 3.4993238]
Lognormal - Wind speed	[3.74540, 14.47066]	[2.880287, 6.220395]

Table 1: CI for parameters of different models

Graphically, it seems there are two candidate probability distributions for the wind power, namely the gamma and the beta distribution respectively:

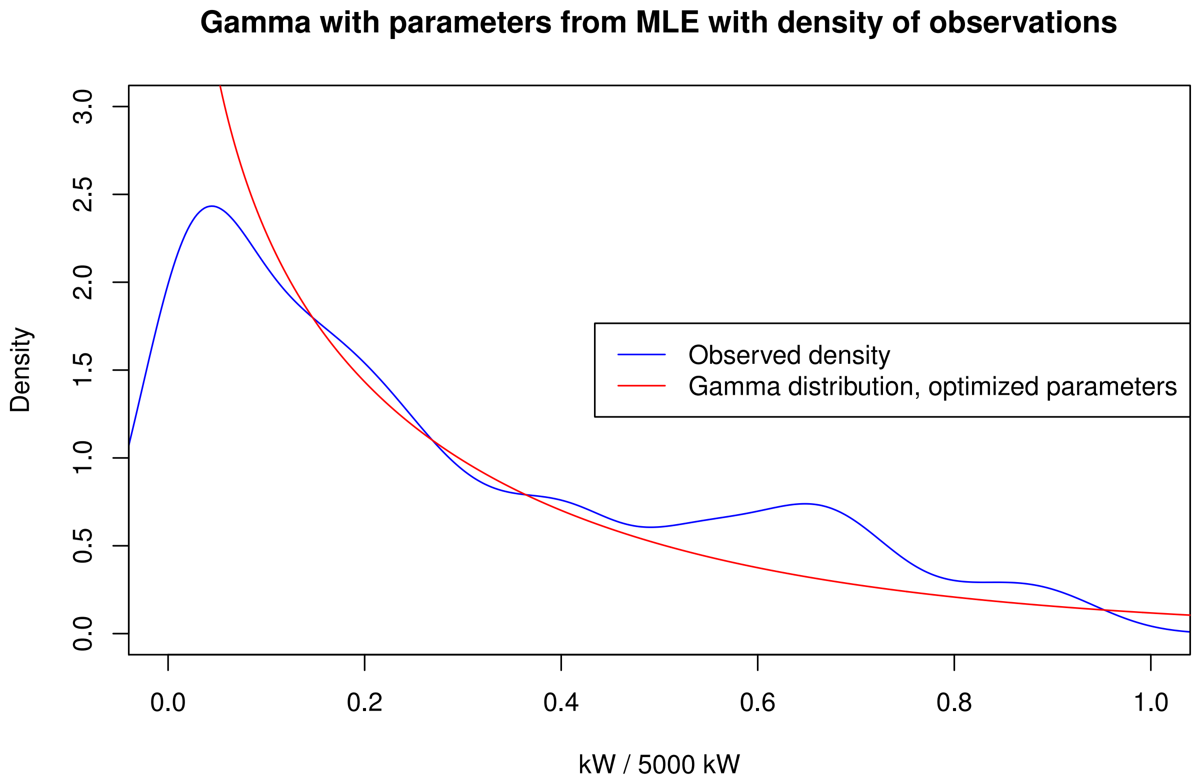


Figure 1: The density of the observed, normalized wind power alongside with the gamma distribution with optimal parameters

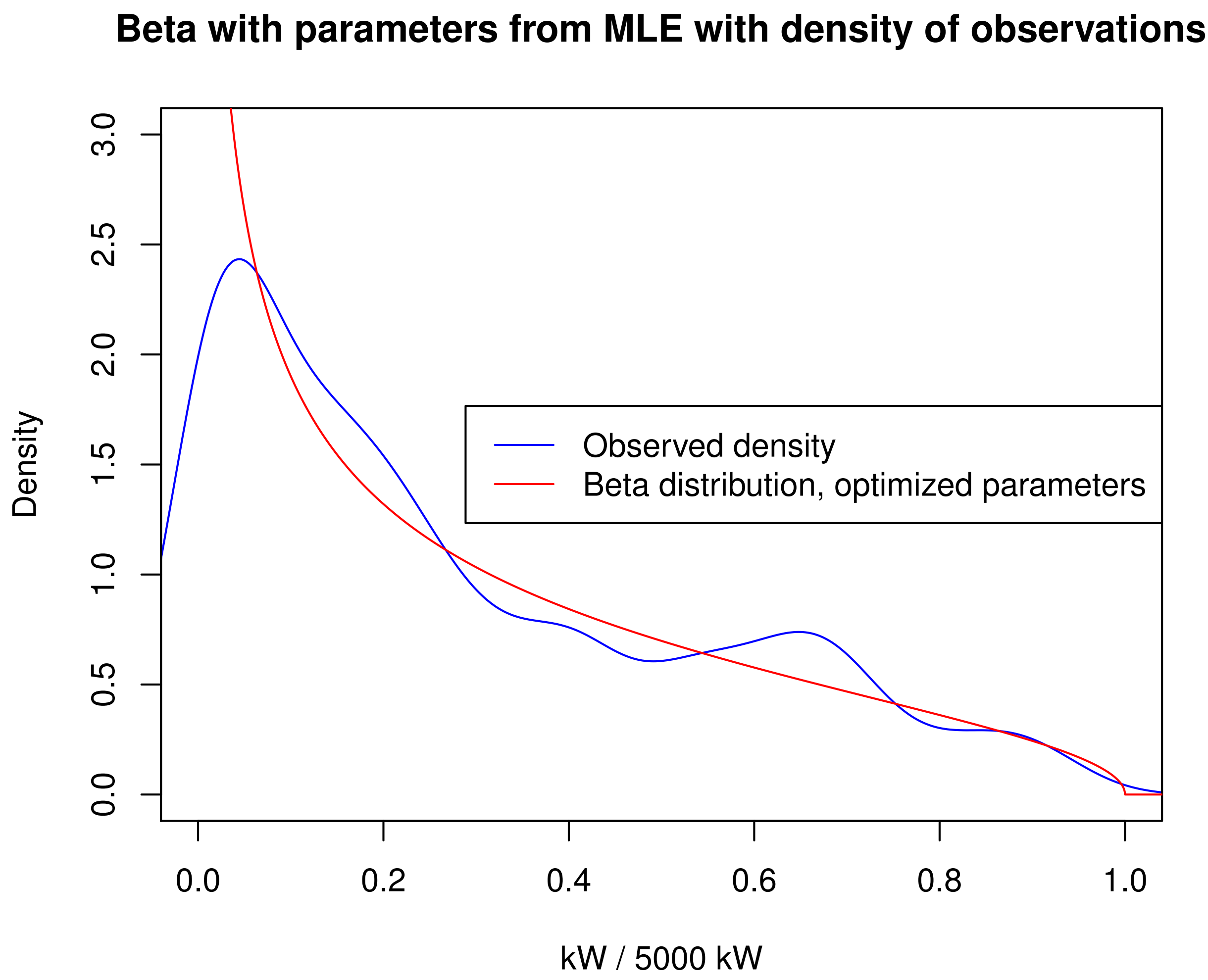


Figure 2: The density of the observed, normalized wind power alongside with the gamma distribution with optimal parameters

To make a decision between these two models, the AIC can be calculated from the formula:

$$AIC = -2 \cdot \log(L(\hat{\theta}) + 2 \cdot p$$

where $L(\hat{\theta}$ is the found maximum likelihood estimate when optimising the parameters and p denotes the number of parameters in the likelihood function. The AIC's found for the two candidate PDF's are listed in the following table:

AIC/model	**Gamma(0.6925961, 0.3988331)**	**Beta(0.5571045, 1.4918276)**
AIC	$-2 \cdot (97.38174) + 2 \cdot 2 = -190.7635$	$-2 \cdot (121.6618) + 2 \cdot 2 = -239.3236$

Table 2: Computed AIC's for gamma and beta distribution for modelling wind power

As the AIC for the Beta-distribution is less than for the Gamma-distribution, it can be concluded that for the wind power variable, the Beta-distribution is the most appropriate model. For the wind speed variable, only the gamma distribution and lognormal distribution is viable, because the wind speeds aren't normalized and the beta distribution is only defined on the interval (0,1). Again, we can make a graphical evaluation of the two models compared to the density of the observations:

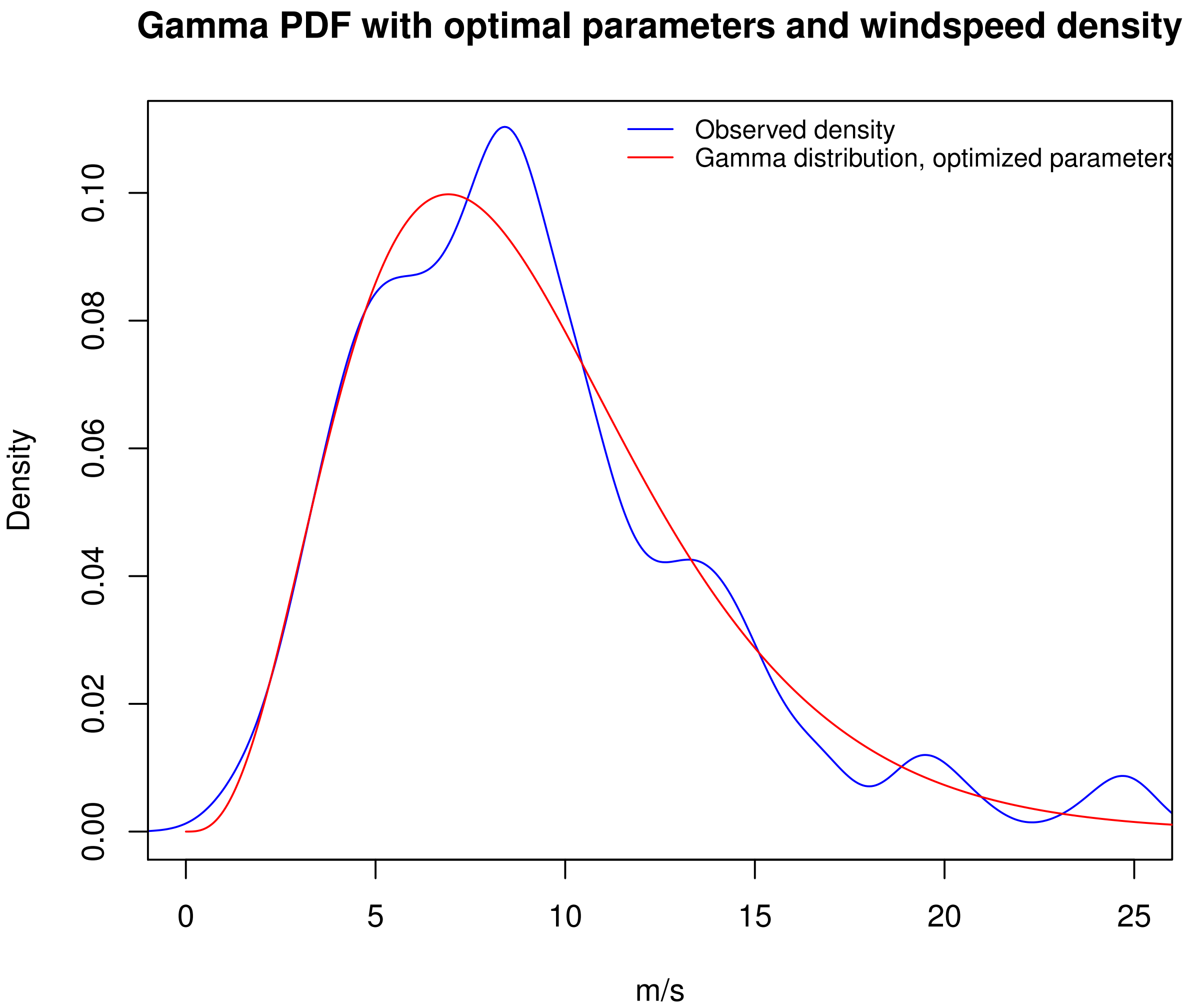


Figure 3: Gamma PDF with optimal parameters and observed density

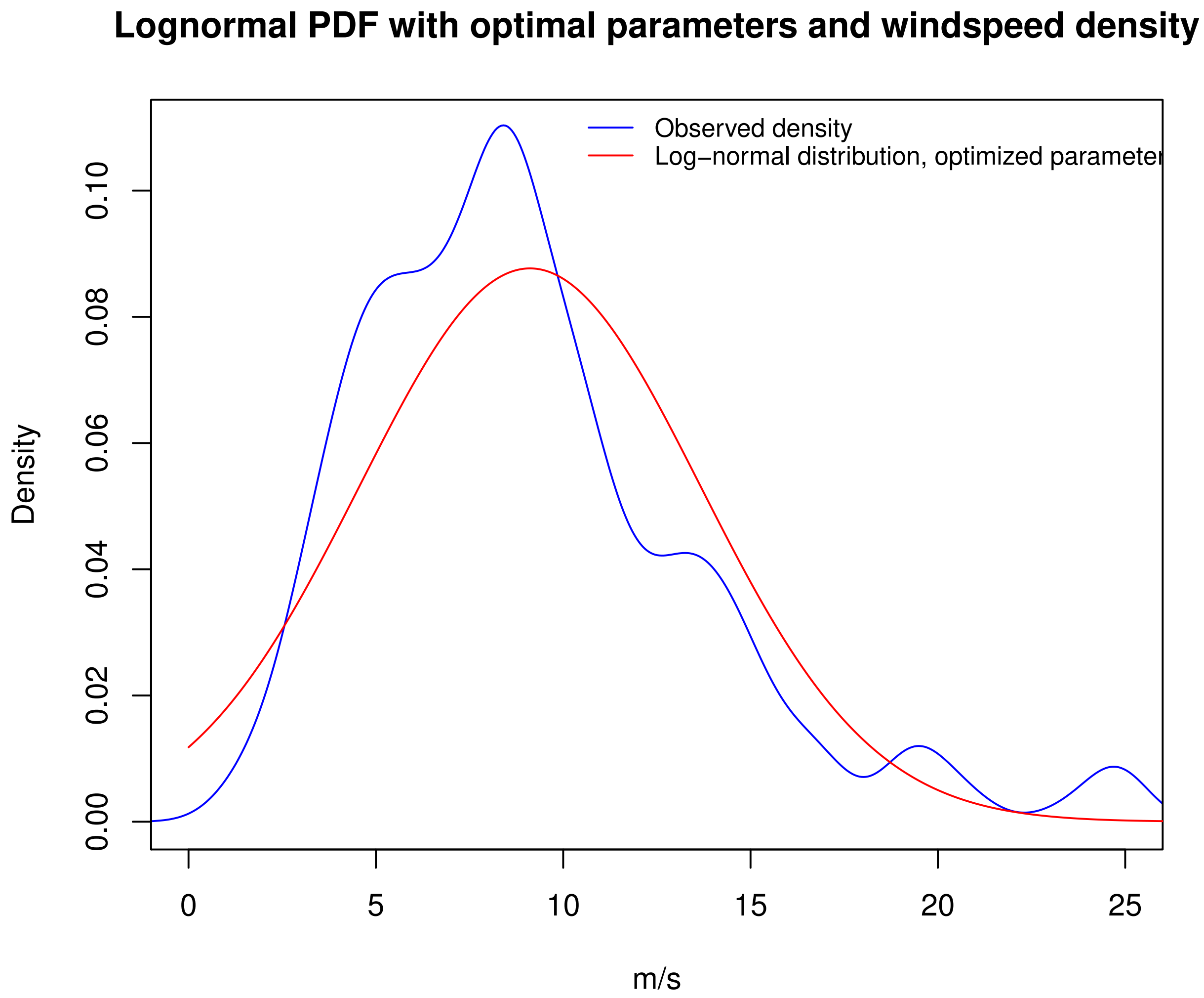


Figure 4: Log-normal PDF with optimal parameters and observed density

It seems the gamma distribution is a better fit to the observations than the log-normal distribution. The AIC's are calculated to be:

AIC/model	Gamma(4.156440, 2.191973)	Log-normal(9.113851, 4.550320)
AIC	$-2 \cdot (-815.3223) + 2 \cdot 2 = 1634.645$	$-2 \cdot (-845.0583) + 2 \cdot 2 = 1694.117$

Table 3: Computed AIC's for gamma and log-normal distribution for modelling wind speed

The AIC for the gamma distribution is less than the AIC for the log-normal distribution. In conclusion, the gamma distribution with the optimized parameter estimates is the most appropriate distribution to model the wind speed variable.

Transformations

The following transformations were applied to the wind power:

$$y^{(\lambda)} = \frac{1}{\lambda}\log\left(\frac{y^\lambda}{1-y^\lambda}\right), \quad \lambda > 0 \tag{1}$$

$$y^{(\lambda)} = 2\log\left(\frac{y^\lambda}{(1-y)^{1-\lambda}}\right), \quad \lambda \in [0,1] \tag{2}$$

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases} \tag{3}$$

It can be shown using the slides from week 4 that the profile likelihood of λ w.rt to $\boldsymbol{y}$ of such a transformation is given by:

$$l_P(\lambda; \boldsymbol{y}) = -\frac{n}{2}\log\left(\hat{\sigma}^2\right) + \sum \log\left|\frac{\partial y^{(\lambda)}}{\partial y}\right| \tag{4}$$

Where $\hat{\mu}$ and $\hat{\sigma}$ are given by:

$$\hat{\mu} = \frac{1}{n}\sum_i y_i^\lambda \tag{5}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i \left(y_i^{(\lambda)} - \hat{\mu}\right)^2 \tag{6}$$

A graph of the profile likelihood for each of these transformations was made and the profile likelihood was optimized numerically using the *optimize* function in *R*. The MLE estimates can be seen on table 4.

MLE of λ	95 % LI	Transformation
0.2620818	[0.1231391 0.4010245]	1
0.2523298	[0.1959213 0.3087382]	2
0.3467257	[0.2790137 0.4144377]	3

Table 4: MLE estimate of λ for different transformations.

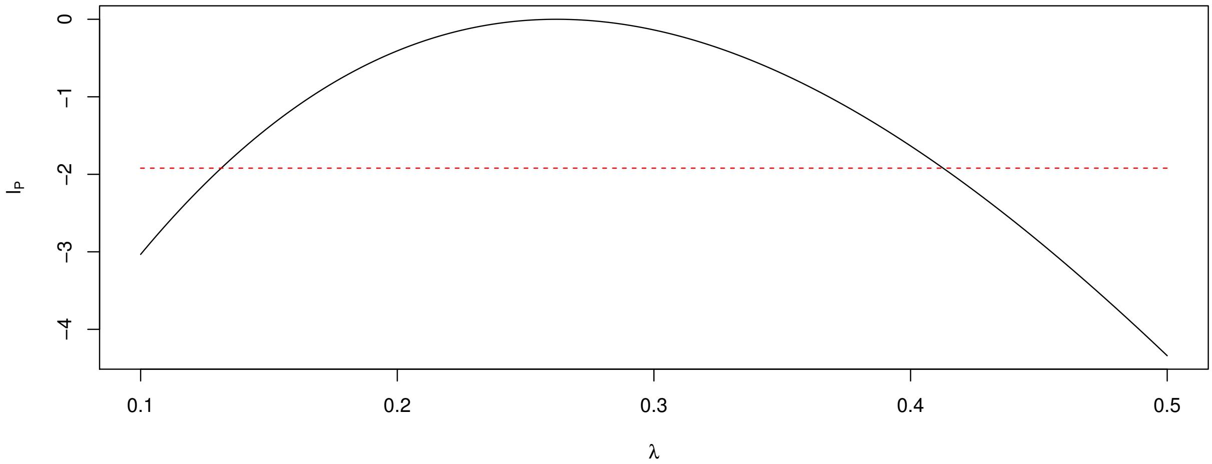


Figure 5: The red line indicates a 95% confidence interval for λ

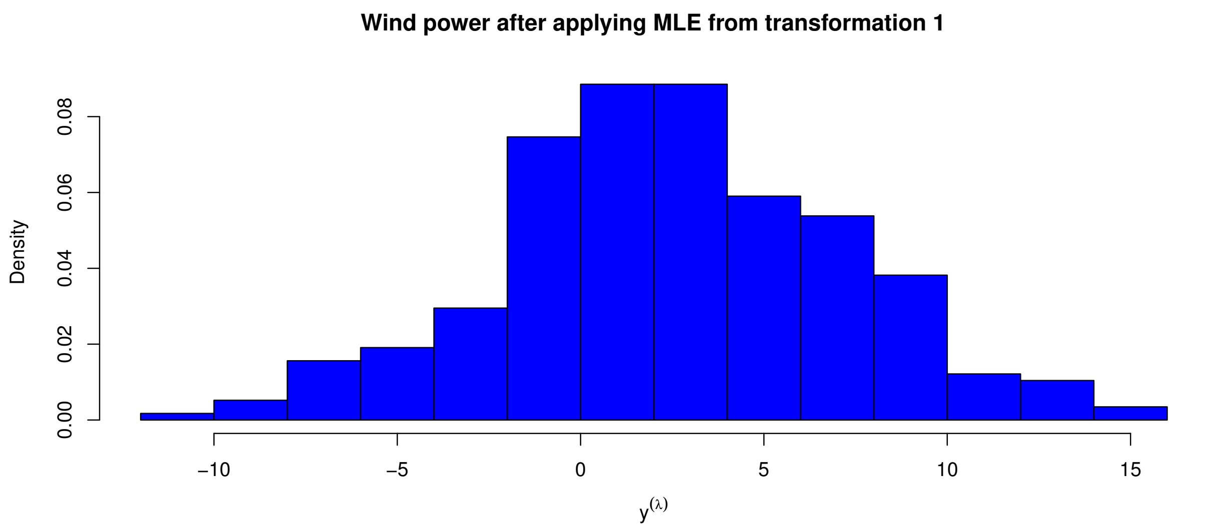


Figure 6: Histogram of wind power after applying transformation 1

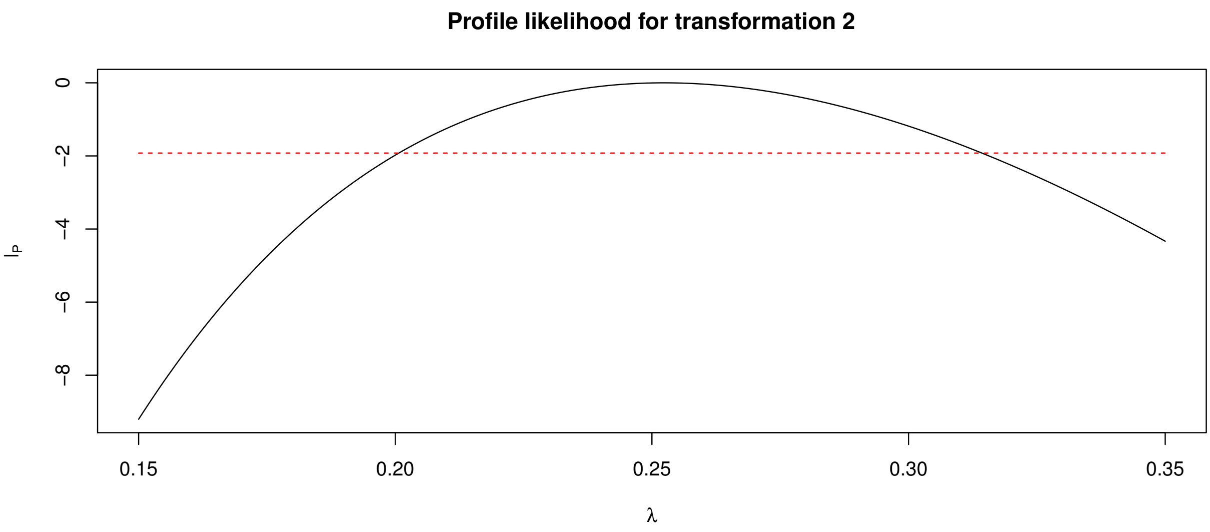


Figure 7: The red line indicates a 95% confidence interval for λ

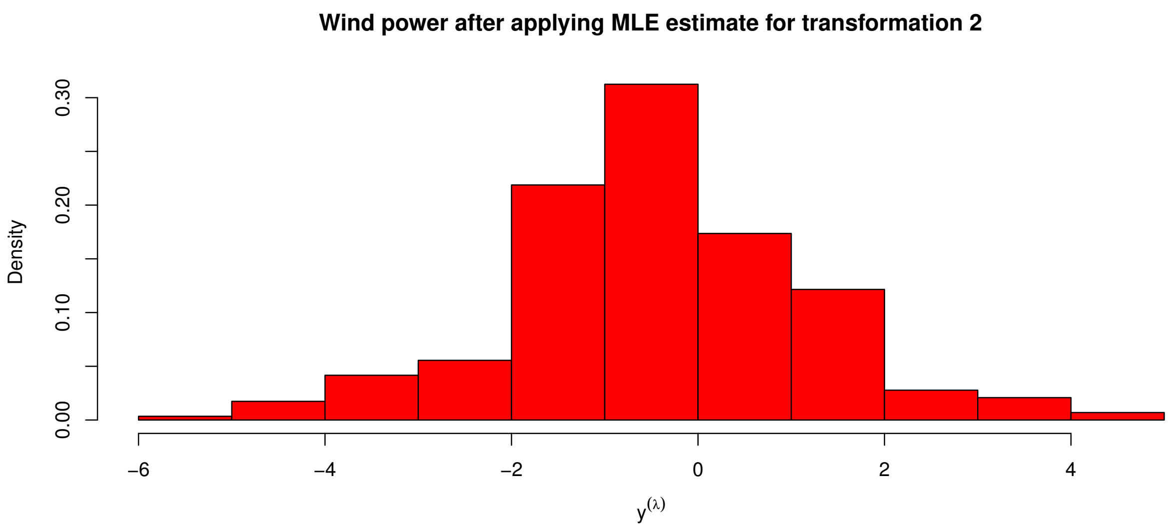


Figure 8: Histogram of wind power after applying transformation 2

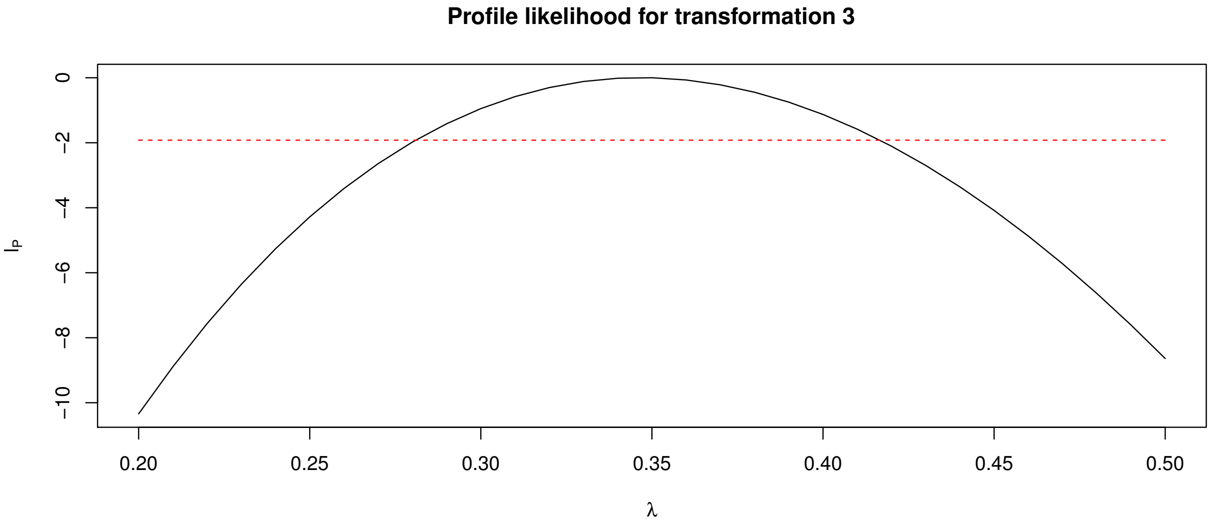


Figure 9: The red line indicates a 95% confidence interval for λ

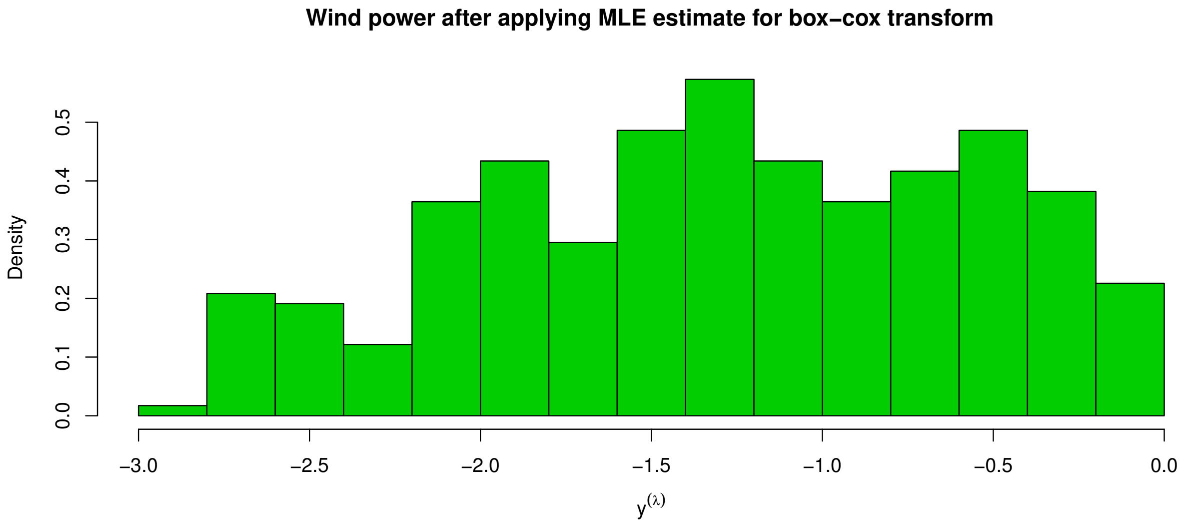


Figure 10: Histogram of wind power after applying transformation 3

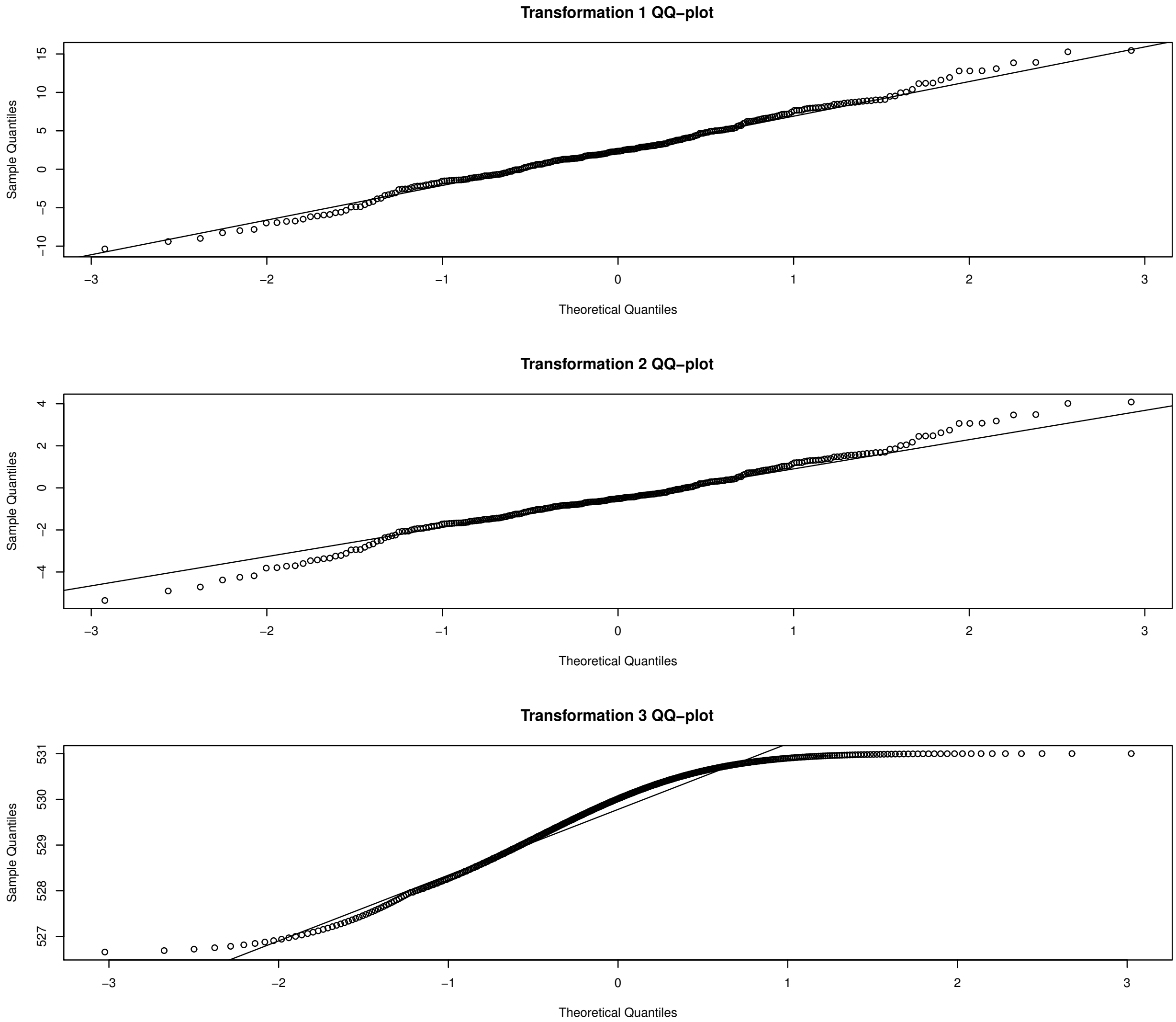


Figure 11: QQplot of different transformation

From figure 11 one can see that transformation 1 fits the theoretical quantiles the best. This one would be the most adequate for performing a normal regression model.

Regression models

Model selection

The initial model is based on the data that has not been transformed and will therefore be a non-normal model. A rather complex structure is initially chosen since that will allow us to either iteratively remove non-significant parameters or expand our model if it is necessary. The choice of initial model is given by,

$$\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s + \beta_2 w_s^2 + \beta_3 w_s^3 + \beta_4 w_d, \tag{7}$$

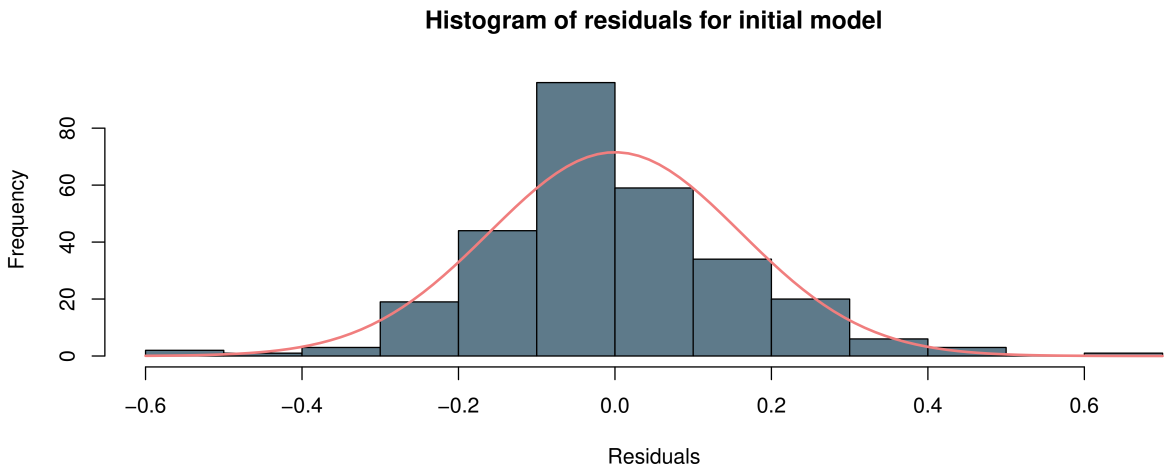


Figure 12: A histogram of the residuals in the initial model: $\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s + \beta_2 w_s^2 + \beta_3 w_s^3 + \beta_4 w_d$

where $\hat{y}$ are our estimates for the normalized power production, w_s is the wind speed, and w_d is the wind direction.

The parameters are estimated by solving the normal equations, which for a generalized linear model finds the Maximum Likelihood Estimates for the parameters $\boldsymbol{\beta}$. The MLE parameters to the objective function,

$$\arg\min_{\beta} \sum_{i=1}^{n} (y_i - \boldsymbol{x_i}\boldsymbol{\beta}))^2, \tag{8}$$

are therefore computed by,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{9}$$

Once the parameters for the given model has been estimated, certain model-statistics can be calculated in order to be able to decide whether the structure of the model should be changed. These model-statistics include the p-value, i.e. a test of whether the parameters are to 0 with a significance level of 5%, of each parameter, the AIC of the model, and a test of whether the residuals are normally distributed, since that is an assumption for the general linear model. Since the different regression models that are fitted to the data are also nested they can be compared by using a χ_q^2 distribution where q is the dimension of $\boldsymbol{\beta}$.

The p-values of the initial model are presented in table 5. The hypothesis that β_1 parameter is equal to 0 cannot be rejected and could therefore potentially be removed from the model. The same goes for β_4 but one parameter will be removed at a time. The AIC of the model is -225.4, which cannot be interpreted until another model has been produced. A histogram of the residuals is shown in figure 12, where it seems that they are quite well-behaved and approximately follows a normal distribution.

Parameter	Intercept	w_s	w_s^2	w_s^3	w_d
P-value	0.69029	0.60759	0.00188	0.00048	0.37203

Table 5: P-values for each parameter in the initial model: $\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s + \beta_2 w_s^2 + \beta_3 w_s^3 + \beta_4 w_d$

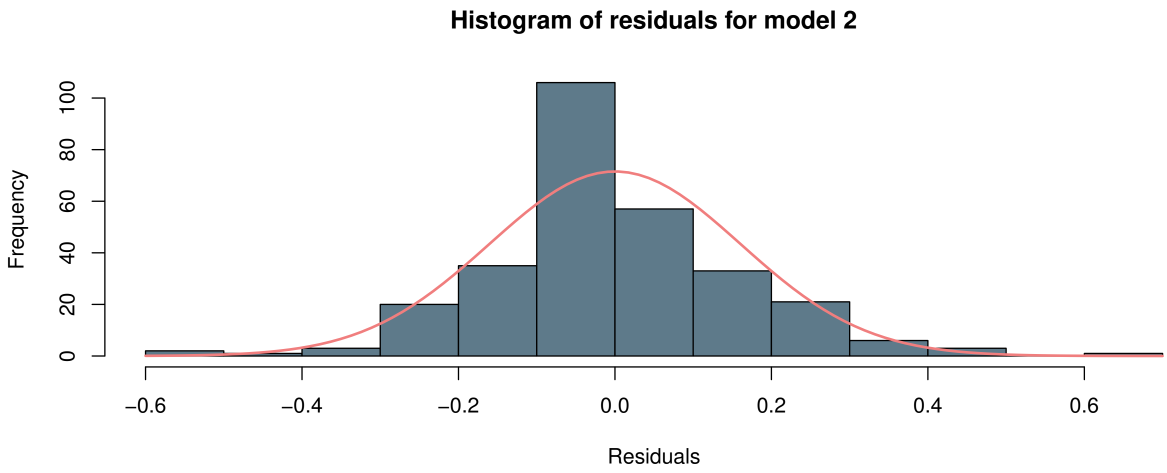


Figure 13: A histogram of the residuals in the second model: $\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3 + \beta_3 w_d$

In order to increase the performance of the model β_1, the w_s-parameter, will be removed and a new model of the form $\hat{y} = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3 + w_d$ will be fitted to the data. The p-values of this model are presented in table 6. In this model the p-values for the parameters of w_s^2 and w_s^3 are much more significant than previously which is a step in the right direction. The p-value for w_d infers that this parameter could be removed, which will be done in the next model-iteration. The residuals of this model are presented in figure 13 and do not seem to have changed much from the initial model. The AIC of the model is -227.13, which is a very slight improvement. However, as shown in the analysis of deviance table in 7 the difference between the 2 models is not significant.

Parameter	$Intercept$	w_s^2	w_s^3	w_d
P-value	0.831	$2 \cdot 10^{-16}$	$3.12 \cdot 10^{-16}$	0.375

Table 6: P-values for each parameter in the second model: $\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3 + \beta_3 w_d$

	df	$D(\cdot)$	$diff$	$p-value$
Initial model	$df_2 = 283$	$D(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\mathbf{2}}) = 7.3949$	$D(\hat{\boldsymbol{\mu}}_{\mathbf{2}}, \hat{\boldsymbol{\mu}}_{\mathbf{1}}) = -6.9 \cdot 10^{-3}$	$P(X > D(\hat{\boldsymbol{\mu}}_{\mathbf{2}}, \hat{\boldsymbol{\mu}}_{\mathbf{1}})) = 0.6072$
Second model	$df_1 = 284$	$D(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\mathbf{1}}) = 7.4018$		

Table 7: Comparison of initial model and second model by a χ_4^2 distribution.

As discussed previously the w_d-parameter will be removed in this model iteration, which means that the model to be fitted is $\hat{y} = f(w_s) = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3$. The p-values for this model are presented in table 8 and it can be concluded that the only parameter that is not significantly different from 0 is the intercept. The remaining parameters are highly significant. The residuals are not illustrated as they are again very similar to the previous model, which might be explained by the fact that the models do not seem to be significantly different which is highlighted by the the ANOVA χ_3^2-test, which has a p-value of 0.374. However, a small decrease in AIC was again achieved, with the AIC of this model being -228.33.

Parameter	*Intercept*	w_s^2	w_s^3
P-value	0.286	$< 2 \cdot 10^{-16}$	$3.98 \cdot 10^{-16}$

Table 8: P-values for each parameter in the third model: $\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3$

Final model

Out of the tested models the best performing one was the third model, i.e. the model of the form

$$\hat{y} = f(w_s, w_d) = \beta_0 + \beta_1 w_s^2 + \beta_2 w_s^3, \tag{10}$$

$$\hat{y} = f(w_s, w_d) = -0.021 + 0.00489 w_s^2 - 0.000147 w_s^3. \tag{11}$$

In order to asses the values of the parameters a confidence interval is calculated based on the profile likelihood of each parameter. Each parameter β_k is profiled out in the following way by defining $\delta = \beta_{\neq k}$

$$L_p(\beta_k) = max_\delta L(\beta, \delta) \tag{12}$$

The null hypothesis is $H_0 : \beta = \beta_0$ and the test statistic is

$$LR = 2(\log L_p(\hat{\beta}) - \log L_p(\beta_0)), \tag{13}$$

where $\hat{\beta}$ is the parameter value that maximizes $L_p(\beta)$. To test whether the null hypothesis can be rejected at a significance level of 0.05 the p-value is computed by comparing the test statistic to a t-distribution with one degree of freedom at the given significance level, i.e. $p = 2 \cdot P(T > |LR|)$. The confidence interval is then calculated by computing the values of β_0 for which the the null hypothesis cannot be rejected.

The standard error of each parameter and the 95%-confidence interval is presented in table 9.

Parameter	Value	Standard error	Lower 95% CI	Upper 95% CI
β_0	$-2.10 \cdot 10^{-2}$	$1.97 \cdot 10^{-2}$	$-5.95 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$
β_1	$4.89 \cdot 10^{-3}$	$3.86 \cdot 10^{-4}$	$4.13 \cdot 10^{-3}$	$5.65 \cdot 10^{-3}$
β_2	$-1.47 \cdot 10^{-4}$	$1.72 \cdot 10^{-5}$	$-1.80 \cdot 10^{-4}$	$-0.14 \cdot 10^{-4}$

Table 9: Values, standard error and confidence intervals for all parameters in the final generalized linear model.

The parameters of the final model include a couple of series expansions based on the explanatory variables. The wind direction was found to not have any significant effect on the performance of the model and was therefore excluded. However, higher order terms of the wind speed were included, i.e. w_s^2 and w_s^3 which allows the model to fit a third degree polynomial to the data. In figure 14 the fitted values of the final model are presented in a plot with power production w.r.t. wind speed. The fitted values does not depend linearly on wind speed and it seems that the reason why it performs better is that the effect that wind speed has on the power production stagnates at a value of about 20 m/s. Hence a linear fit would not capture the actual dependency between the explanatory variable and the target data. A linear model has not been tested but presumably the residuals of such a model would not be normally distributed to a satisfactory degree. The confidence interval on the plot defines the area where the "underlying" mean of the power production is estimated to be. The prediction interval defines the area where the model estimates that all future observations will lie with a probability of 95%.

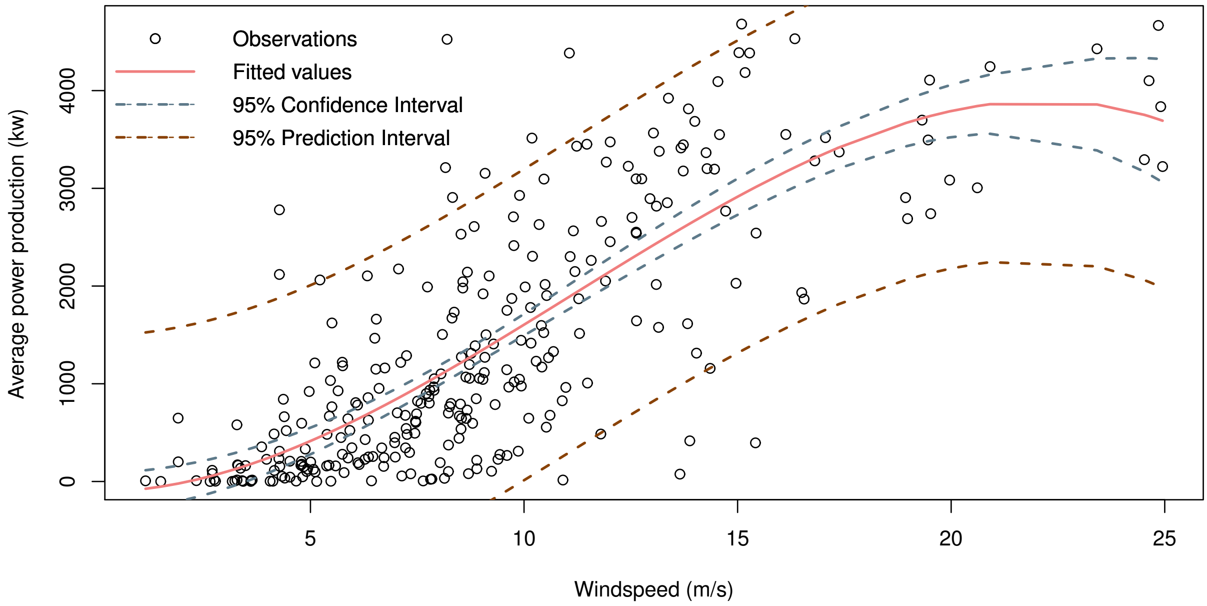


Figure 14: The fitted values of the final model plotted alongside all observations. The red lines constitute the lower and upper bound of the 95%-confidence interval.

Normal model

For the transformed data where transform 1 was performed a normal regression model was also made. Using the exact same procedure as in the prior section "Regression model" the following final model was made:

$$\hat{y}^{(\lambda)} = f(w_s, w_d) = \beta_0 + \beta_1 w_s + \beta_2 w_s^2 + \beta_3 w_d^2, \tag{14}$$

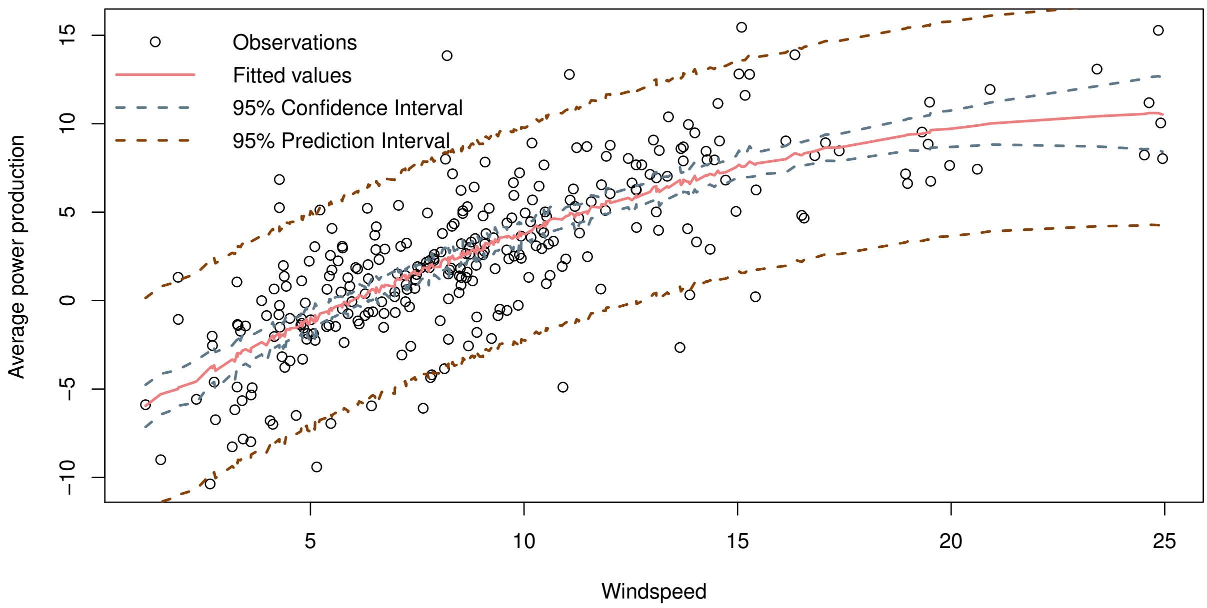


Figure 15: Final normal model with prediction and confidence interval

A table with the uncertainty of the parameters:

Parameter	Value	Standard error	Lower 95% CI	Upper 95% CI
β_0	-7.3285	0.7622	-8.8289	-5.8282
β_1	1.4142	0.1384	1.1418	1.6866
β_2	-0.0275	0.0057	-0.0387	-0.0163
β_3	-0.0105	0.0165	-0.0430	0.0219

Table 10: Values, standard error and confidence intervals for all parameters in the final normal linear model.