ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MASTER THESIS ECONOMETRICS AND MANAGEMENT SCIENCE

BUSINESS ANALYTICS AND QUANTITATIVE MARKETING

Modelling the Effects of Wind Characteristics on Wind Gust Statistics in a GEV Distribution

---

Author:

Joanne Tjan, 413647

Supervisor:

dr. P. Wan

Second assessor:

Prof. dr. C. Zhou

April 25, 2022

Abstract

Many applications require accurate wind forecasts for example, in the renewable energy sector or improving the efficiency of wind power production. One approach to improve the reliability of the wind forecasts is to include additional characteristics about wind. This paper focuses on comparing several variable selection methods to determine which wind covariates are the most informative to include in the cGEV model. This model is based on a generalized extreme value distribution with censoring at the 50% quantile of the observations. The wind gusts observed from the Hamburg Weather Mast at five different height levels between 10 and 250 m are modelled with wind characteristics from the COSMO-REA6 dataset. We use five different variable selection methods, which are ridge regression, lasso, elastic net, adaptive lasso and adaptive elastic net. The most informative variables are the maximum wind gust diagnostic at 10 m and its variance, the barotropic mode, the mean of the horizontal wind speed at 700 pHa, and the surface pressure tendency which are selected in all variable selection methods. Results reveal that the adaptive lasso is the best performing method in terms of the continuous ranked probability score (CRPS) with an improvement of 7.8984% with respect to the baseline model. The adaptive lasso also performs well for the daily wind gusts, the top 5% strongest wind and overall. Lastly, this paper confirms that adding additional informative variables to the model results in an improvement of at least 23%.

Contents

1 Introduction **1**

2 Literature Review **3**

2.1 Statistical Wind Distributions 4

2.2 Variable Selection for Wind Covariates 5

3 Data **8**

3.1 Hamburg Weather Mast 8

3.2 COSMO-REA6 Regional Reanalysis 11

4 Methodology **14**

4.1 Censored Generalized Extreme Value Model 14

4.2 Overview of Variable Selection Methods 17

4.2.1 Ridge Regression 17

4.2.2 Lasso 18

4.2.3 Elastic Net 18

4.2.4 Adaptive Lasso 19

4.2.5 Adaptive Elastic Net 19

4.3 Performance Metrics 20

4.3.1 Cross-validation procedure: 21

5 Results **22**

5.1 Interpretation of the Covariates 25

5.2 Evaluation Results 26

6 Conclusion **29**

7 Limitations and Future Research **30**

Bibliography **32**

A Data analysis of observed wind gusts at the Hamburg Weather Mast **37**

1 Introduction

Wind power has been receiving increased attention over the last years because of the focus on the exploitation in renewable energy[46]. This is especially influenced by government policies for sustainable energy and the current state of global warming and climate change. In fact, wind energy is becoming one of the most reliable, significant and cleanest power sources of renewable energy and recently became an important source in electricity markets as well [39]. It provides as an environmentally friendly alternative for fossil fuels. However, wind energy is an intermittent power source and as a result, an efficient power system of wind energy depends on the ability to forecast future available wind power. Therefore, accurate and precise wind forecasts are required for stable wind energy.

Moreover, reliable wind forecasts are not only relevant for a viable and sustainable energy program but also important to identify extreme wind events in advance. Extreme winds are one of the main weather threats with dangerous consequences for both humans and economies. It often results in serious damage to infrastructures, such as direct damage to buildings in the city or indirect damage in the form of the risk of a loose object flying through the sky crashing on other construction or even on a citizen. In addition to that, the aftermath of the damages caused by severe wind events generally involves a large amount of money and time to clean up the damages. Besides, other risks of extreme wind speeds are the fact that offshore areas are more likely to experience stronger wind speeds such that the consequences are more severe in those areas. For example, offshore wind farms are more likely to experience higher damage [21].

Therefore, it is of great importance to have more accurate, reliable and precise wind forecasts. One of the main problems for wind forecasting is the fact that wind is often not continuous and steady which can cause irregular and strong variations of wind gusts [43]. Wind speeds occur intermittently, vary on an inter-annual, inter-decadal time scale and differ along their entire vertical extent. In other words, there are many factors in deciding which distribution is the most appropriate and suitable to use for modelling wind speed since these distributions are constantly changing over time [7].

One approach for wind prediction modelling is to combine the observed wind gusts with numerical weather predictions [25] and more specifically with meteorological reanalysis data which has recently gained popularity among academia and government. Reanalysis data is often used for monitoring climate change or describing the history of the atmosphere, land surface or ocean. In short, reanalysis datasets contain a lot of explanatory detailed atmospheric variables that are mostly used in meteorological and climatological studies. These datasets are generally obtained by means of data assimilation and put into a chosen NWP model. Then, these models

are used for providing weather forecasts based on current weather conditions. Newly received observations are continuously inserted into the model such that the old forecasts are corrected using the updated observations in order to give a more accurate outcome.

However, reanalysis produces large datasets, since these datasets contain decades worth of detailed information about the atmosphere or climate. Analyzing everything for predictive modelling takes a large computation time and space which often results in large computation incurred costs. As a solution, variable selection methods are implemented to remove the unnecessary variables and select the essential ones which have the most significant impact on wind forecasting [20]. The advantages of variable selection methods are that they not only reduce the computation time, space and data acquisition costs but also simplify the interpretation of the remaining variables. In addition to that, the efficiency and efficacy of the final model increase as the amount of variables reduces without affecting the accuracy and precision too much. Lastly, simpler models are usually easier to validate and are more easily and better understood. Therefore, variable selection methods are very popular and widely applied in different fields.

By increasing the reliability and precision of wind forecasts, these forecasts can be further enhanced to determine the wind more precisely, taking its intermittent nature into consideration. Moreover, wind farms can anticipate and schedule wind energy based on the intensity of the accurate wind forecasting as well as reduce the production costs to a large extent. In short, reliable wind forecasts are applied to the risk assessment of wind farms to improve the efficiency of wind power production [56]. Furthermore, a measure of the potential risk by means of reliable, precise forecasts offers better insights to emergency managers, air and trail traffic on the time of disruption, and on how to take certain actions to diminish destruction. Also, this measure offers helpful insights to offshore wind operators when considering opening a potential new wind farm [27, 44]. It explores the need for wind energy resources as well as the high cost of wind farm construction [31]. Lastly, this information might also be useful for public entities or companies in the energy sector if they are interested in investing in renewable energy.

We investigate and compare in this paper several variable selection methods for the censored generalized extreme value (cGEV) model for the wind gusts in Germany as we are interested to know whether the choice of a variable selection method will impact the predictive performance of the model. We define the objective and the focus of this paper with the following research question:

Which state-of-the-art variable selection methods can be used to optimally choose the most informative meteorological variables that provide a diagnostic of the observed wind gust at a height

of 10 m?

In order to answer this research question, we use a high-resolution regional reanalysis dataset for Europe called the COSMO-REA6. It consists of 150 different variables associated with the characteristics of the weather conditions from which we only choose 16 variables. This regional reanalysis covers the time period from 1995 to August 2019 and provides additional valuable beneficial information on the observed gusts at the height of 10 m. The cGEV model assumes generalized extreme value distribution (GEV) where observed gusts are censored above a certain threshold in order to avoid biases and focus on the extreme winds. Estimation is done by means of maximum likelihood estimation (MLE). Afterwards, various variable selection methods, such as lasso, elastic net, ridge regression are implemented and compared to each other. In addition to these traditional methods, we will also implement the adaptive version of the classic methods themselves. In short, these variable selection methods are chosen to include the most common methods and some relatively new ones.

This paper contributes to selecting the most adequate explanatory variables for wind gust observations by comparing different variable selection methods for the cGEV model. There is a considerable number of papers comparing several variable selection methods but relatively few focus on meteorological variables or give an overview of the variable selection methods for GEV models. This method allows us to investigate and quantify the effects of relevant covariates on the reliability of the wind speed distribution. When the significant variables are known, one could focus on obtaining the essential variables when time, space or money are limited.

This paper has the following structure: an overview of relevant topics and literature is given in Section 2. Section 3 describes the observed hourly wind gusts at the Hamburg Weather Mast and the COSMO-REA6 dataset with the corresponding data analysis. Next, the framework of the cGEV model is provided in Section 4 as well as a description of all variable selection methods and the verification of the post-processing model. Section 5 presents the results. Lastly, the conclusion and recommendation of this research are discussed in Section 6.

2 Literature Review

In this section, we give an outline of the relevant topics regarding this research. First, we give a brief summary of known aspects when modelling wind gust observations which give us a better understanding of handling observed wind gusts. Then, an overview is given about the main variable selection methods and their enhanced versions.

2.1 Statistical Wind Distributions

There have been several papers over the years that have investigated the purpose of wind in different settings, such as in designing any engineering structure [1], consequences of extreme wind events in urban areas [3] or the financial consequences of risks of high wind speeds [10].

Since the 1930s, wind has been comprehensively investigated and analyzed, and its corresponding probabilistic and statistical parameters have been thoroughly identified. [1] is one of the first papers handling wind gusts in which they generally describe the wind pressure data followed by a summary about wind pressure against a tall building. Since then, many papers have been published about deriving wind distributions depending on the wind direction, angle, speed or pressure.

In 1962, the first statistical concepts of wind engineering were introduced in [11] where they concluded that wind angle, speed, and pressure tend to follow a Gaussian process. However, [23] suggested that a non-Gaussian distribution might be more suitable for wind pressure when deriving its statistical distribution. In addition to that, they believed that the wind pressure data were not symmetric and tended to be skewed which was also shown in [41].

Compared to wind angle and wind pressure, wind speed is the most important factor in wind modelling since wind speed has been widely used in various fields, such as in the selection of suitable wind turbines [55], utilization of wind energy [45] or measuring high wind speeds in tropical cyclones [42]. In such applications, one common detail is that higher and larger variations in wind speed lead to very serious and often unfavourable results. Strong wind speeds may lead to the deactivation of wind farms along with huge financial losses or severe outcomes and material damages after a big tropical cyclone. As for the wind farms, a steady wind speed is required in order to maintain a stable electricity network and hence limited wind speed variation is preferred as well. Therefore, wind speed analysis and modelling are among the most important aspects of wind engineering as there are more risks involved in strong wind speeds than in the angle of the wind or strong wind pressures.

A similar conclusion can be drawn for wind speed, indicating that a Gaussian distribution seems to be unfitting due to the skewed data. The most common statistical distribution used for modelling wind speed is the Weibull distribution as it gives the best fit to the wind speed distribution compared to other distributions [28, 48, 50]. The most chosen estimation methods for the Weibull distribution are the maximum likelihood estimation, methods of moments and the least-squares methods.

However, even though the Weibull distribution is widely used among researchers, [7] concluded that the Weibull distribution is not always the most appropriate distribution to identify

the characteristics of wind speed in every scenario, e.g. in the case of a high rate of null wind speed. Other studies show that the Gumbel distribution might be reasonable to choose over the Weibull distribution as the Gumbel distribution estimates winds speed more accurately at the tail of the distribution and the lower levels [33] or conclude that the Gumbel distribution is more reliable to use to model extreme wind speeds than the Weibull distribution [30].

Another promising statistical wind speed distribution is the generalized extreme value distribution shown in [40] where they compared the Weibull method to the methods based on the extreme value theory. They concluded that using the Weibull method leads to incorrect estimates of the tails of the wind speed distributions and that the extreme value methods avoid these problems. Similarly, [14] developed and compared seven approaches, such as the gamma, log-normal and generalized extreme value distribution to derive a probabilistic analysis for German wind gusts. Among the given distributions, they revealed that the generalized extreme value distribution is the most suitable to estimate the statistical distribution of wind speed as it is the most reliable and theoretically consistent.

The generalized extreme value distribution (GEV) is the underlying extreme value distribution of the Block Maxima (BM) approach which is one of the main methods to model extreme events in the Extreme Value Theory (EVT). The BM approach divides the data into blocks of equal length and selects the maximum in each block. Then, the GEV distribution is fitted to the sample consisting of the maximum value of all blocks. The corresponding parameters of the GEV distribution are the location parameter μ, scale parameters $\sigma > 0$ and shape parameter ξ. The most common method to estimate these parameters is the maximum likelihood estimation providing stable estimates. The GEV distribution combines three different extreme value distributions into a generalized distribution; the Gumbel, Fréchet and reverse Weibull distribution also known as type I, II and III extreme value distributions depending on the value of ξ. This correspond with $\xi = 0$, $\xi > 0$ and $\xi < 0$ respectively. Based on this, the shape parameter ξ strongly influences the weight of the tail of the GEV distribution and decides whether the distribution has an upper or a lower bound.

We opt for the generalized extreme value distribution in this research to derive the statistical distribution of the extreme wind speeds. We refer to [9] for more specification and details about the EVT.

2.2 Variable Selection for Wind Covariates

Wind speed prediction methods generally use univariate wind speed observations based on historic data where limited information is used for wind speed modelling and predicting. Hence,

we assume that wind speed prediction can be further improved if more information about the corresponding observations is added to the model for better and more accurate forecasts.

As a solution, an alternative approach is proposed by introducing covariates into the parameters of the wind speed distribution as suggested in [29]. Several papers have been studying how to incorporate covariates into the model for better accuracy [6] or to improve the goodness-of-fit of models [37]. These covariates usually contain additional valuable information, such as trends, physical characteristics, cycles etc. Incorporation of information has been done before in different fields and distributions, for example in [12] using a polynomial function on annual maximum precipitation data in the GEV distribution or wind speed modelling in eastern Canada in [36].

One potential data source for additional valuable information are reanalysis datasets. Reanalysis data are widely used in a variety of domains due to their reliability in containing detailed information about historic weather conditions [5]. Hence, reanalysis data are appropriate to use as covariates for meteorological and climate forecasting, such as done for wind speed modelling in [49]. The COSMO-REA6 data set was used as covariates which provided a diagnostic of observed wind gusts in Germany. Similar, [53] concluded that an introducing a covariate for heavy rainfall improves the modelling of extreme precipitation. Models including additional information are often called non-stationary models. Therefore, adding covariates to the model will improve the model performance to obtain more accurate and reliable forecasts.

Since reanalysis data consists of precise atmospheric variables, there is a strong correlation between these meteorological variables and the wind gust values. Therefore, we use this as input to improve the performance of the prediction models. However, having too many irrelevant variables affects the performance of the model negatively in terms of longer computation time and the expensive costs of data acquisition. Besides, the other advantage of using variables selection methods is to reduce the complexity of the prediction models without affecting the accuracy to maintain a parsimonious model. Hence, the purpose of variable selection is to increase the model prediction by identifying and selecting the most important and influential variables for the final simpler model.

Variable selection has been employed in many applications, such as for clinical predictive modelling in the medical field [47], bankruptcy forecasts in finance [51] or clustering marketing segmentation [32]. Traditional approaches for variable selection, such as forward and backward stepwise selection using p-values, AIC and BIC have been commonly used in general due to their simplicity. However, the traditional methods disregard the multi-collinearity problem among variables and are more likely to yield poor performance.

An alternative to the classic approaches is using modern variable selection methods, such as

methods based on machine learning algorithms or hybrid methods to overcome the challenges of multi-collinearity and overfitting. The interest in modern variable selection methods as a solution to address the challenges of the traditional approaches has been steadily increasing over the years. One of the most popular choices for variable selection is the least absolute shrinkage selection operator (lasso) [52]. Within the renewable energy literature, the lasso has often been used as it is known as a simple algorithm, yet gives great results, and is frequently able to outperform the standard approaches for variable selection. In addition to that, the lasso is applicable in different aspects of the renewable energy sector. For example, it has the highest increase in performance for solar radiation forecasting [18], is used for modelling electricity prices [57] and showed the best performance in precipitation occurrence [17]. Since then, many extensions of the lasso have been used and some of them show promising results, such as elastic net [59], group lasso [16], Bayesian lasso [38], adaptive lasso [58], etc. [35] compares ridge regression with the lasso and its extensions and concludes that all lasso-related methods had relatively high prediction accuracies compared to the ridge regression using simulated breeding values in a genome.

Nonetheless, the above-mentioned methods are proposed to address the challenges of the lasso but there is no dominant approach under all conditions that is better than the others since each method has its own specific strengths and weaknesses, and the model performance is highly data-dependent [13]. Despite lasso being a popular and reliable method for variable selection, there are very few papers about the variable selection available in the GEV model. For example, [34] proposed a method for variable selection in the GEV model based on the Akaike Information Criterion for wave height data. Similarly, other papers have shown that the most important predictors for the GEV distribution are found using machine learning algorithms, such as the fused lasso for the annual maximum precipitations [24], random forests for the annual streamflow data [54], or lasso to select the most valuable wind predictors [49]. To the best of our knowledge, there have been no papers available which give an overview of the variable selection methods in the GEV model, and especially in the GEV model for wind gusts data. Hence, the focus of this research is to incorporate the wind covariates into the cGEV model to derive a distribution for the hourly wind gusts using linear combinations. The interfering variables are removed with lasso-type methods, such as the elastic net, adaptive lasso, adaptive elastic net and the lasso itself, and the results are compared with the classic ridge regression.

3 Data

The data used in this paper consists of two parts: the hourly wind gust observations in Germany and the COSMO-REA6 dataset which is the reanalysis dataset from Europe. The variables in the COSMO-REA6 dataset are used as covariates, to determine how the covariates influence the expected speed of the wind gusts. We describe the characteristics and data analysis for each dataset as well as the data availability of the datasets.

3.1 Hamburg Weather Mast

The hourly wind gusts are measured at the Hamburg Weather Mast in Germany, where the mast is operated by the Meteorological Institute at the University of Hamburg since 1967. The mast is located in the Billwerder district and the geographical location of the mast is 53 ° 31 '09.0' 'N and 10 ° 06' 10.3 " E, and 53 ° 31 '11.7' 'N and 10 ° 06 '18.5' 'O at the highest and lowest height respectively. Five different height levels measure wind gusts at heights of 10, 50, 110, 175 and 250m. The raw wind data are obtained and the average of the wind data is taken every 3 seconds. Then, these values are used to calculate the hourly wind gusts by selecting the maximum value over these average observations per hour. The data is collected from 1 January 2004 until 31 December 2014, amounting to a total of 96432 observations for every height level. The descriptive statistics of the hourly observed gust observations are presented in Table 1 at different height levels. Values denoted as 99999 are unknown values and are therefore considered missing values. All wind gusts are noted in meters per seconds.

Table 1. Descriptive statistics of the observed hourly gust observations measured at the Hamburg Weather Mast from 1 January until 31 December 2014.

Descriptive statistics	Minimum	Maximum	Mean	Median	Standard Deviation	Observations
10M	0.390	28.070	6.122	5.790	3.1139	93025
50M	0.530	33.210	7.969	7.400	3.8168	95700
110M	0.460	35.980	9.143	8.650	4.0127	93167
175M	0.450	38.930	10.078	9.690	4.3390	92300
250M	0.410	40.580	10.850	10.540	4.7152	89106

Table 1 shows the descriptive statistics of the wind gusts, such as the minimum, maximum, mean, median, standard deviation and the total number of observations. The minimum value of the observed gust observations is fairly similar among the heights whereas this is not the case for the maximum value. The maximum value is increasing for every larger height level, as are the mean, median and standard deviation. This indicates that stronger wind gusts are

more likely to appear at a higher height and that wind gusts at higher height levels are more likely to deviate than at lower height levels. Based on the given graphs and differing mean and median, we conclude that the distribution of the observed wind gusts at each height level is lightly skewed. Lastly, we omit the missing values and show the final number of observations in the last column of Table 1. As previously mentioned, more observations are missing for higher heights implying that wind gusts are not always present at all height levels. Graphs of wind gust at 10 m and 50 m throughout the years are shown in Figure 1. For both height levels, boxplots for every hour and every month are shown in Figure 2 and Figure 3. The remaining graphs and boxplots at other height levels are given in Appendix A.

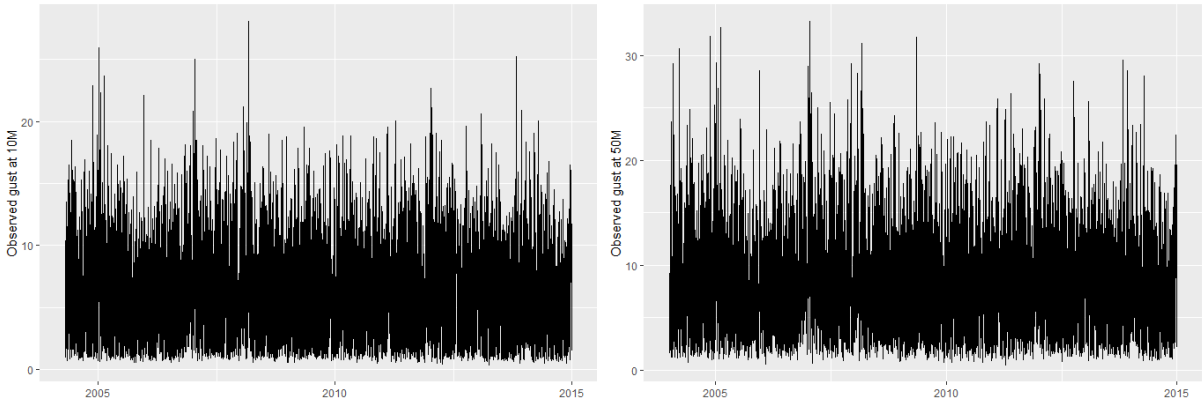

Figure 1. Observed wind gusts from 1 January 2004 until 31 December 2014 at 10 m (left) and 50 m (right).

We see in Figure 1 that seasonality is strongly present in the graphs at 10 m and 50 m and this effect can also be found in Appendix A at other height levels. The higher peaks mostly resemble the stronger wind gusts in the winter, as the top ten strongest wind gusts appear in the winter. This information is consistent with Figure 2 and Figure 3. The strongest wind that was observed occurred on 18 January 2007 when the wind speed was 25.05, 33.21, 35.98, 38.93 and 40.58 for 10 m, 50 m, 110 m, 175 m and 250 m respectively. This extreme wind event was caused by the windstorm Kyrill and the fastest recorded wind speeds in Germany are mainly caused by storms or cyclones, which was evident at all height levels.

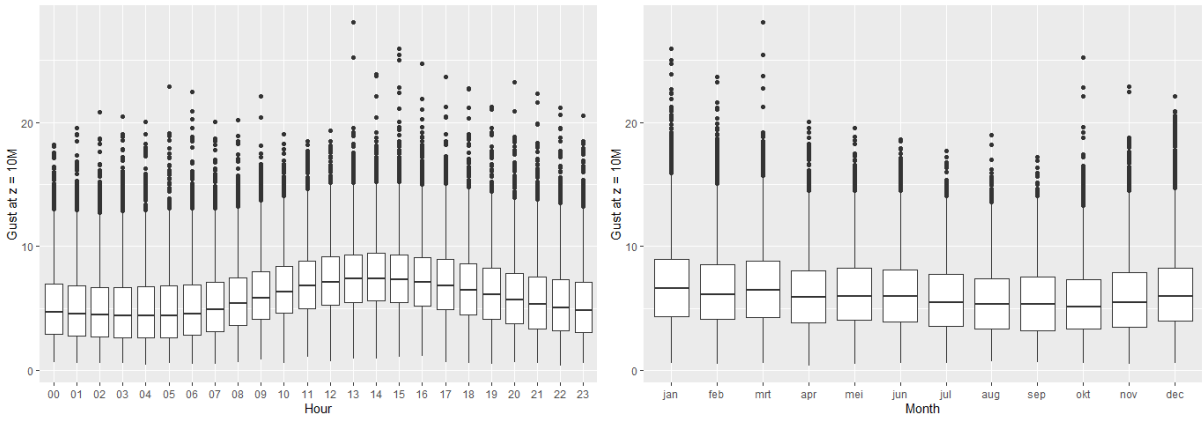

Figure 2. Boxplots of hourly (left) and montly (right) wind gusts at 10 m.

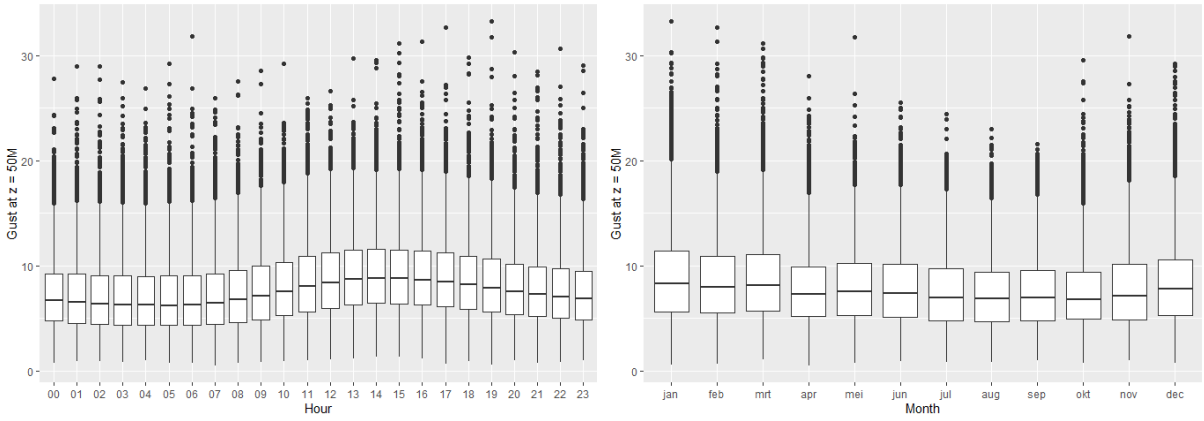

Figure 3. Boxplots of hourly (left) and monthly (right) observed wind gusts at 50M.

In Figure 2 and Figure 3 (left), the box plots for the hourly observed wind gusts are shown at heights of 10 m and 50 m. The median of the wind gusts is the largest around noon and in the afternoon and the lowest in the night and the early morning. This indicates that wind at noon is usually stronger than at nighttime. This pattern is apparent at 10 m and becomes less visible and wavy at higher heights (see Appendix A). Therefore, wind becomes more stable during the day at higher height levels. Moreover, outliers of wind gusts around noon usually have a smaller range than those at other hours. This applies to all height levels except for a height of 50 m at midnight, suggesting that wind gusts are faster and steadier at noon than in the late afternoon.

As for monthly wind gusts, they seem to be quite stable over the year where the wind becomes weaker during summer and stronger during winter. This is evident at all height levels but more apparent at higher height levels. The same pattern is visible in terms of outliers where more extreme wind gusts appear in the winter season than in the summer season. From this analysis, we conclude that in general stronger wind gusts appear in the winter in Germany. Therefore,

more wind energy becomes available during the winter season than in the summer season due to the higher probability of stronger wind speeds in the winter months.

3.2 COSMO-REA6 Regional Reanalysis

The COSMO-REA6 reanalysis is developed at the Hans-Ertel-Centre for Weather Research of Deutscher Wetterdienst (DWD) for Europe [2]. The COSMO-REA6 dataset was released to overcome the limitations of the former reanalysis to provide more improved variables for renewable energy-related applications. As previously stated, reanalysis data are obtained by implementing data assimilation into an NWP model. The NWP forecast model used for obtaining the COSMO-REA6 dataset is called the COnsortium for Small-Scale MOdelling limited-area model (COSMO-LAM). Based on this information, the COSMO-REA6 dataset yields worthwhile information as diagnostics on the observed wind gusts at the Hamburg Weather Mast.

The COSMO-REA6 regional reanalysis covers a period from 1995 until August 2019 but we only use the necessary data observed from 1 January 2004 until 31 December 2011 for this research. There is a total of approximately 150 different variables containing 2D and 3D variables. For the purpose of this research, 16 variables are preselected beforehand which might be valuable and potential covariates for the observed wind gusts. Hence, we use the COSMO-REA6 to provide us with diagnostics of the gust observations from the Hamburg Weather Mast. The following variables are considered from the COSMO-REA6 reanalysis as covariates: *VMAX_10M* (the wind gust diagnostic at 10 m), *TWATER* (total water content), *T_2M* (atmospheric temperature at 2m). Some variables are obtained by transforming the data, such as taking the difference between variables (LI; lifted index, d_tCAPE; tendency in convective available potential energy, *Vh_SHEAR*; vertical shear of horizontal wind between 6 and 1 km, d_tP; surface pressure tendency), the mean or standard deviation of the vertical and horizontal wind speed at 700 hPa denoted as Vh_700 and W_700 respectively or the variance VAR_tVMAX_10M. Lastly, other variables are retrieved by applying principal component analysis (PCA) on the wind series over 11 years (Vh_EOF1 and Vh_EOF2) or by including the annual cycle which is represented by a linear combination of the sine and cosine function (AC_COS and AC_SIN). A summary and a more detailed description of the preselected covariates are shown in Table 2.

Table 2. Summary and description of preselected covariates from the COSMO-REA6 reanalysis from [49].

Acronyms	Variable	Description	Unit
VMAX_10M	Wind gust diagnostic at 10 m	Grid value	Meters per second
VAR_t *VMAX_10M*	Temporal variance of VMAX_10M	Variance of five consecutive (±2h) grid values	Meters per second
Vh_EOF1	Barotropic mode of absolute horizontal wind at lowest layers	Principal component of first eigenvector of covariance matrix from wind time series (11 years) at lowest 300m (six layers)	
Vh_EOF2	Barotropic mode of absolute horizontal wind at lowest layers	Principal component of second eigenvector of covariance matrix from wind time series (11 years) at lowest 300m (six layers)	
$Mean_h$ *Vh_700*	Mean absolute horizontal wind at 700 hPa	Mean of 25 mast-surrounding grid values at layer 23	Meters per second
SD_h *Vh_700*	Standard deviation of aboslute horizontal wind at 700 hPa	Standard deviation of 25 mast-surrounding grid values at layer 23	Meters per second
$Mean_h$ *W_700*	Mean vertical wind at 700 hPa	Mean of 25 mast-surrounding grid values at layer 23	Meters per second
SD_h *W_700*	Standard deviation of vertical wind at 700 hPa	Standard deviation	Meters per second
d_tP	Surface pressure tendency	Mean difference between current and previous surface pressure from mast-surrounding grid values	Millibars
LI	Lifted index	Difference between the temperature at 500 hPa (layer 18) and the temperature of an adiabatically lifted surface air parcel	Celsius.
TWATER	Water content	Water content of the mast-including grid column	m^3
d_tCAPE	Convective Available Potential Energy tendency	Difference between current and previous CAPE of the mast-including grid column	Joules per kilogram of air
Vh_SHEAR	Horizontal wind shear	Difference between absolute horizontal wind in 6 km (layer 17) and 1 km (layer 30)	Meters per second
T_2M	Temperature at 2 m	Grid value	Celcius
AC_COS	Annual cosine cycle	Cosine oscillation with 1-year period	
AC_SIN	Annual cosine cycle	Sine oscillation with 1-year period	

Table 2 shows 16 potential covariates for the wind gusts observations from the Hamburg Weather Mast. Three of the variables are directly taken from the COSMO-REA6 (textitVMAX_10M, *TWATER* and *T_2M*) and the other 13 variables are obtained through feature engineering. Descriptive statistics of the preselected covariates are shown in Table 3. We apply the Jarque-Bera test for each covariate to check whether the covariates follow a normal distribution. The null hypothesis for all covariates are rejected indicating that the covariates are not normally distributed. Moreover, a matrix showing the correlation between the covariates is displayed in Figure 4. There is minimal correlation present in the covariates except for the variables directly taken from the COSMO-REA6 dataset. Other variables that seem to have a higher correlation with other variables are *LI* and *AC_COS*.

We presume the gust diagnostic variable *VMAX_10M* to be the most explanatory and informative among the covariates. It corresponds to the maximum turbulent and convective wind gust diagnostic at 10 m. This variable intends to estimate the potential and the maximum speed of a gust near a surface. We examine the differences between the wind gusts observed at the Hamburg Weather Mast and the wind diagnostic from the COSMO-REA6 reanalysis in Figure 5 using a graph and a histogram.

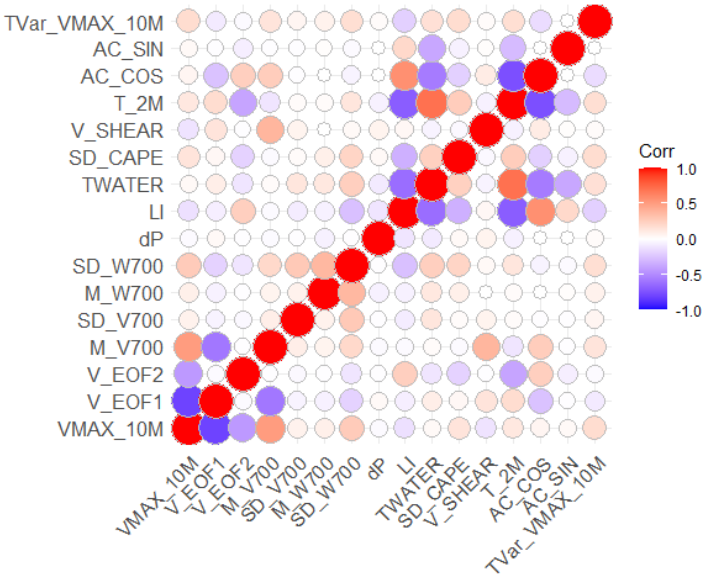

Figure 4. Correlation matrix of the covariatios from the COSMO-REA6 reanalysis.

Table 3. Descriptive statistics of the high-resolutional COSMO-REA6 reanalysis from 1 January until 31 December 2014.

Descriptive statistics	Minimum	Maximum	Mean	Median	Standard Deviation	Observations
VMAX_10M	0.3282	27.7230	7.1487	6.7036	3.3954	96378
VAR_t *VMAX_10M*	0.000	53.9022	1.7202	0.7807	2.6787	96398
Vh_EOF1	-42.0210	15.8885	0.0000	0.1927	7.2043	93456
Vh_EOF2	-5.5141	9.1749	0.0000	-0.0252	1.8454	93456
$Mean_h$ *Vh_700*	0.2386	49.6046	11.8841	11.0229	6.5520	93456
SD_h *Vh_700*	0.0206	5.4667	0.4300	0.3587	0.2999	93456
$Mean_h$ *W_700*	-0.6876	1.1303	0.0003	-0.0022	0.0371	93456
SD_h *W_700*	0.0014	1.0863	0.0294	0.0227	0.0284	93456
$d_t P$	-406.9631	403.1225	0.0058	0.4275	47.4581	96366
LI	-23.376	38.507	7.317	6.776	7.9018	93450
TWATER	1.135	56.680	16.339	15.166	7.9301	96377
$d_t CAPE$	0.000	1476.557	15.574	0	64.44102	96378
Vh_SHEAR	-19.2289	51.6371	6.8084	5.9730	8.5512	93456
T_2M	251.2	315.9	283.7	283.5	8.8252	59853
AC_COS	-1.0000	1.0000	0.0000	0.0000	0.7071	96432
AC_SIN	-1.0000	1.0000	0.0000	0.0000	0.7071	96432

The histogram and the graph of the differences between the observed wind gusts at 10 m and the variable *VMAX_10M* are shown in Figure 5. The mean and the standard deviation of the differences are -1.0259 and 1.7962 respectively. This indicates that there is a small negative bias as the COSMO-REA6 reanalysis overestimates the maximum wind gust diagnostic compared to

the observed wind gusts from the Hamburg Weather Mast. A similar conclusion can be drawn for the wind gusts in the summer months. The largest differences often appear in the summer months and therefore, we assume that the wind gusts are more stable in the winter months than in the summer months. Hence, the variable VAR_10M overemphasizes the strengths of the wind gusts and this often happens for the gusts in the summer months.

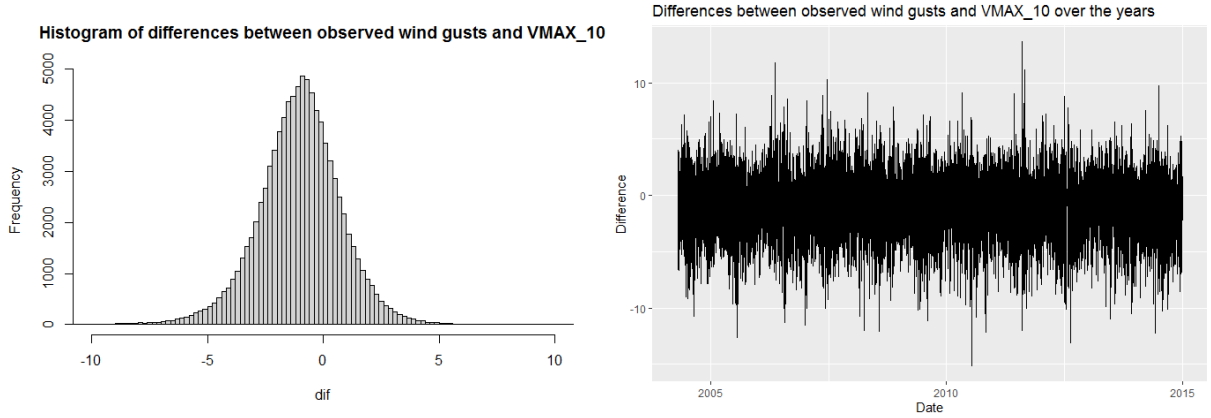

Figure 5. Histogram and graph of the differences between the observed wind gusts at the Hamburg Weather Mast and the COSMO-REA6 *VMAX_10M*.

4 Methodology

In this section, we first explain the model for the observed wind gusts at the Hamburg Weather Mast based on the extreme value theory. Then, we demonstrate how the covariates obtained from the COSMO-REA6 reanalysis are incorporated into the cGEV model. The objective of this paper is to compare several variable selection methods in order to provide the optimal covariates for the observed wind gusts in Germany. Therefore, the chosen variable selection methods will be thoroughly described, i.e. lasso, elastic net and their adaptive version and the ridge regression. Lastly, a brief overview of several approaches to measure the performance of a variable selection method will be given. All implementation for the parameter estimation as well as for the variable selection methods is done in R software. The code for all variable selection methods and implementation is available on Github.

4.1 Censored Generalized Extreme Value Model

The model used in this research is the post-processing method for hourly wind gusts described in [49]. First, we denote the hourly gust observations as $Y(z,t)$ where the wind gust depends on the height level z and time t of the day for modelling. To model the wind gust observations,

we choose the generalized extreme value model and censor the wind gusts to focus more on the extreme wind events, and to avoid biases due to the non-asymptotic behaviour. The general idea of the censored generalized extreme value (cGEV) model is to censor $Y(z,t)$ if it is below a certain threshold c and to $Y(z,t)$ otherwise, summarized as

$$Y(z,t) = \begin{cases} c & \text{if } Y(z,t) < c, \\ Y(z,t) & \text{if } Y(z,t) \geq c, \end{cases} \tag{1}$$

where $Y(z,t)$ is the hourly wind gust at height $z \in \{10, 50, 110, 175, 250\}$ and time $t = 1, \ldots, 96432$. Since the observed wind gusts represent the maximum values over the averaged observations for every hour, the block maxima approach seems to be the most suitable in this case. For the block maxima approach, the generalized extreme distribution (GEV) is the most appropriate underlying distribution relying on the Fisher–Tippett–Gnedenko theorem. For more elaboration or details about the extreme value theory and the GEV distribution, we refer to [9]. Thus, the asymptotic cumulative density function (cdf) of the GEV distribution is given by

$$G(Y(z,t); \mu(z,t), \sigma(z,t), \xi) = \begin{cases} \exp\Big(- [1 + \xi(\frac{Y(z,t)-\mu(z,t)}{\sigma(z,t)})]^{-\frac{1}{\xi}} \Big) & \text{if } \xi \neq 0 \\ \exp\Big(- \exp[-(\frac{Y(z,t)-\mu(z,t)}{\sigma(z,t)})] \Big) & \text{if } \xi = 0, \end{cases} \tag{2}$$

where $\mu(z,t) \in (-\infty, \infty)$, $\sigma(z,t) \in (0, \infty)$ and $\xi \in (-\infty, \infty)$ on $\{Y(z,t) : 1 + \xi(Y(z,t) - \mu(z,t))/\sigma(z,t) > 0\}$. The parameters $Y(z,t), \mu(z,t), \sigma(z,t)$ and ξ represent the hourly wind gusts, location, scale and shape parameters respectively. In a similar way as $Y(z,t)$, the location and scale parameter $\mu(z,t)$ and $\sigma(z,t)$ rely on height z and time t. More information is given later given when we describe the non-stationary behaviour of the model. Based on previous literature, we choose the Gumbel distribution fixing the shape parameter $\xi = 0$ to model the extreme wind events. There are two reasons for this decision. First, the Gumbel distribution has no upper or lower limit as opposed to the Fréchet or Weibull distribution. Predictive probability for future wind gusts above or below these limits will be zero which leads to bad forecasting. Secondly, increasing the number of parameters often leads to more uncertainties during the maximum likelihood estimation especially when estimating the shape parameter ξ in a non-stationary setting. By fixing the shape parameter $\xi = 0$, it stabilizes the optimization routines. Thus given the reasons mentioned above, $G(Y(z,t); \mu(z,t), \sigma(z,t)) = \exp\Big(- \exp[-(\frac{Y(z,t)-\mu(z,t)}{\sigma(z,t)})] \Big)$ denotes the cdf of the hourly wind gusts in this case.

We attempt to explain the non-stationary behaviour of the cGEV model through covariates. Based on previous studies, the most common approach is a linear combination because of its straightforwardness and direct interpretation. Therefore, we assume a linear relationship between the observed wind gusts and the covariates and incorporate this additional information

into the model. In other words, the non-stationary behaviour is defined as linear combinations of the wind covariates into the cGEV parameters. In this model, we assume the covariates $C(t)$ are identical at every height level for time t, to focus more on the impact of the covariates throughout all variable selection methods. These linear relationships between the wind covariates and the cGEV parameters are shown in the following equations:

$$\mu(z,t) = \mu_0(z) + \sum_{l=1}^{L} \mu_l(z) C_l(t), \tag{3}$$

$$= \mu_0 + \sum_{l=1}^{L} \mu_l C_l(t), \tag{4}$$

$$= \mu(t), \tag{5}$$

and

$$\sigma(z,t) = \exp\Big(\sigma_0(z) + \sum_{l=1}^{L} \sigma_l(z) C_l(t)\Big), \tag{6}$$

$$= \exp\Big(\sigma_0 + \sum_{l=1}^{L} \sigma_l C_l(t)\Big), \tag{7}$$

$$= \sigma(t), \tag{8}$$

where $Y(z,t)$ denotes the hourly wind gusts varying in height and time, $C_l(t)$ the lth covariate which only varies in time, L the total number of covariates and the constant parameters μ_0 and σ_0. The exponential function is required to guarantee a positive $\sigma(t)$ for all time t.

The parameters of the CGEV model are obtained using a maximum likelihood estimation (MLE) because this method can incorporate the wind covariates into the CGEV parameters efficiently. For the MLE, we require the asymptotic probability density function (pdf) of the GEV distribution for maximisation and the pdf is defined as

$$g(Y(z,t); \mu(t), \sigma(t), \xi) = \frac{1}{\sigma(t)} t(Y(z,t))^{\xi+1} e^{-t(Y(z,t))}, \tag{9}$$

where

$$t(Y(z,t)) = \begin{cases} 1 + \xi(\frac{Y(z,t)-\mu(t)}{\sigma(t)})]^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ \exp[-(\frac{Y(z,t)-\mu(t)}{\sigma(t)})] & \text{if } \xi = 0. \end{cases} \tag{10}$$

As mentioned, we assume that $\xi = 0$ for better predictive probabilities and stabilization routines such that $t(Y(z,t) = \exp\Big[-\Big(\frac{Y(z,t)-\mu(t)}{\sigma(t)}\Big)\Big]$. We derive from Equation 9 the log-likelihood function for maximisation which is given by

$$\ell(\mu, \sigma, \xi | Y) = -\sum_{z \in Z} \sum_{t=0}^{T} \Big(\log \sigma(t) + (\xi+1)\Big(\frac{Y(z,t)-\mu(t)}{\sigma(t)}\Big) + \exp\Big(\frac{Y(z,t)-\mu(t)}{\sigma(t)}\Big)\Big), \tag{11}$$

$$\ell(\mu, \sigma | Y) = -\sum_{z \in Z} \sum_{t=0}^{T} \Big(\log \sigma(t) + \Big(\frac{Y(z,t)-\mu(t)}{\sigma(t)}\Big) + \exp\Big(\frac{Y(z,t)-\mu(t)}{(t)}\Big)\Big), \tag{12}$$

where $\mu(t)$ and $\sigma(t)$ are the linear combinations with wind covariates of the location and scale parameter respectively at time t for $t = 1, \ldots, 96432$ and the wind gusts $Y(z,t)$ for height z and time t. In order to maximise Equation 11, we choose the iterative Boyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

4.2 Overview of Variable Selection Methods

Given Table 3, there is currently a total of 16 variables in the COSMO-REA6 reanalysis which gives us additional information about the hourly wind gusts. We use in this research a combination of classic methods and their adaptive versions for comparison in which a penalty term is added to the likelihood function, such as the classic ridge regression, the least absolute shrinkage and selection operator (lasso), elastic net and the adaptive version of the lasso and elastic net methods. These variable selection methods are chosen given their prevalence in the economic and renewable energy literature. The adaptive ridge regression will not be used in this research as there is limited previous literature showing the advantage of adaptive ridge regression over the lasso or elastic net. An overview of the aforementioned methods is given below with the structure and characteristics of each method. For more details about the classic variable selection methods, we refer to [19]. Lastly, the penalty terms irrespectively of the number of hyperparameters in all variable selection, do not include the constant parameters μ_0 and σ_0.

4.2.1 Ridge Regression

The first variable selection method is the ridge regression proposed by [22] with an L_2 penalty term on the coefficients defined. The ridge regression estimator is given in the following function:

$$\ell_{ridge}(\mu, \sigma; Y) = \ell(\mu, \sigma|Y) - \lambda_2 n||\beta||_2^2, \tag{13}$$

$$\beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \qquad ||\beta||_2^2 = \sum_{p=1}^{P} \beta_j^2, \tag{14}$$

where μ and σ are the estimates of the wind covariates in the linear relationships between the covariates and $Y(z,t)$ is the hourly wind gust for all height z at time t. n denotes the total observations for all heights, λ_2 is the penalty squared loss term given the unrestricted log-likelihood value $\ell(\mu, \sigma|Y)$ and P is the total covariates. The L_2 penalty reduces the estimates toward zero, but never exactly zero. Because of that reason, ridge regression has been criticized for not being capable to perform variable selection but it remains a relatively safe method for selecting the optimal variables when the variables are highly correlated.

4.2.2 Lasso

Another shrinkage method is the least absolute shrinkage and selection operator (LASSO) proposed in [52]. This method uses an L_1 penalty term λ_1 also known as the lasso regularization parameter. The following lasso criterion is maximized:

$$\ell_{lasso}(\mu, \sigma|Y) = \ell(\mu, \sigma|Y) - \lambda_1 n||\beta||_1, \tag{15}$$

$$\beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \qquad ||\beta||_1 = \sum_{p=1}^{P} |\beta_j|, \tag{16}$$

where $l(\mu, \sigma|Y)$ is the log-likelihood value with corresponding estimates for the wind covariates μ and σ, λ_2 is the regularization term, n is the total number of observations and P is the total number of covariates. The penalty term in the lasso is more severe than the penalty term in the ridge regression as the lasso regularization term allows insignificant variables to shrink to zero. Because of this aspect, the lasso is one of the easiest models to interpret and estimate.

Furthermore, the regularization terms λ_1 and λ_2 in the ridge regression and the lasso control how strict the penalization is and handle the sparsity of the solution. The larger the penalty term λ_1 or λ_2 is, the more variables are forced to be close to zero or exactly zero. Both ridge regression (Equation 13) and lasso (Equation 15) criteria can be simplified to Equation 11 if λ_1 and λ_2 is equal to zero.

4.2.3 Elastic Net

The next variable selection method is the elastic net (EN) described in [59], where two regularization parameters λ_1 and λ_2 are included. Elastic net is proposed to improve the ridge regression and lasso via a combined penalty term by adding both L_1 and L_2 penalization. The purpose of the first penalty L_1 is to perform automatic variable selection whilst the second penalty L_2 is to improve prediction and handling of the possible collinearity. The elastic net criterion for maximisation is defined as follows

$$\ell_{EN}(\mu, \sigma|Y) = \ell(\mu, \sigma|Y) - \lambda_1 n||\beta_1|| - \lambda_2 n||\beta||_2^2, \tag{17}$$

$$\beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \qquad ||\beta||_1 = \sum_{p=1}^{P} |\beta_j|, \qquad ||\beta||_2^2 = \sum_{p=1}^{P} \beta_j^2, \tag{18}$$

where λ_1 and λ_2 denote the penalty terms for the L_1 and L_2 penalization, respectively, μ and σ denote the wind covariates estimates, $\ell(\mu, \sigma|Y)$ is the standard log-likelihood value given μ and σ, and P is the total number of wind covariates. Since the elastic net includes both L_1 and L_2 penalization, Equation 17 can be simplified to either the ridge regression of the lasso depending

on the value for λ_1 and λ_2. We define $\alpha = \frac{\lambda_2}{\lambda_1+\lambda_2}$ and if α is set to 0, the elastic net becomes the ridge regression. If α is close to one, the elastic net is almost identical to the lasso. The elastic net seems to significantly improve the prediction accuracy if the covariates are highly correlated. To further improve the prediction performance of the model, we multiply the estimates of the elastic net by $(1 + \frac{\lambda_2}{n})$.

4.2.4 Adaptive Lasso

The adaptive lasso presented in [58] is developed to overcome the drawbacks of the regular lasso. First is the lack of oracle property for the lasso, indicating that the lasso does not correctly identify the true model if the true underlying model is given in advance. By achieving the oracle properties of an estimator, the zero parameters will be exactly estimated at zero with probabilities converging to one. As a result, the lasso is in general not variable selection consistent. The second drawback is that the lasso becomes unstable with high-dimensional data and this could result in biased and inconsistent estimates for larger coefficients as well. Hence, the adaptive lasso is proposed as an improved method of the regular lasso including the oracle properties and is given by

$$\ell_{alasso}(\mu, \sigma|Y) = \ell(\mu, \sigma|Y) - \lambda_1 n||\omega\beta||_1, \tag{19}$$

$$\beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \qquad ||\omega\beta||_1 = \sum_{p=1}^{P} \omega_j|\beta_j|, \tag{20}$$

where μ and σ are vectors of estimates of the wind covariates for the location and scale parameter respectively, λ_1 is the penalty term, n is the total number of observations of all heights, P is the total number of covariates, and ω is a vector containing the adaptive data-driven weights where β holds all μ and σ. The weights vector ω is defined as

$$\omega = (|\hat{\beta}_j^{lasso}|)^{-\gamma}, \tag{21}$$

where β^{lasso} are the initial estimates which are yielded from the regular lasso and a corresponding positive constant γ.

4.2.5 Adaptive Elastic Net

Similar to the lasso, the elastic net does have a few key shortcomings: (1) the lack of oracle property where the zero parameters should be exactly zero with probability tending to one, and (2) instability with high-dimensional data. Therefore, the adaptive elastic net estimator described in [60] does satisfy the oracle properties and selects the relevant parameters while simultaneously

taking the possible high correlations between variables into account. The adaptive elastic net estimator is obtained by:

$$\ell_{aEN}(\mu, \sigma|Y) = \ell(\mu, \sigma|Y) - \lambda_2 n||\beta||_2^2 - \lambda_1 n||\omega\beta||_1, \tag{22}$$

$$\beta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \qquad ||\beta||_2^2 = \sum_{p=1}^{P} \beta_j^2, \qquad ||\omega\beta||_1 = \sum_{p=1}^{P} \omega_j|\beta_j|, \tag{23}$$

where μ and σ are the estimates of the wind covariates for the hourly wind gusts, λ_1 and λ_2 are the penalty terms for the L_1 and L_2 penalization, n is the total number of observations, P is the total number of covariates, ω is a vector with the data-driven weights, similar to the weights in the adaptive lasso. Equation 22 shows us that the adaptive elastic net is a mixture of the elastic net and the adaptive lasso. The stability of the method is enhanced by the adaptive lasso while the highly correlated variables are handled by the elastic net at the same time. Furthermore, the data-driven weights ω are constructed by

$$\omega_j = (|\beta_j^{EN}|)^{-\gamma}, \tag{24}$$

where β^{EN} are the initial estimates obtained from the elastic net and a positive constant γ.

4.3 Performance Metrics

The variable selection methods will be evaluated using various performance measures. The most appropriate measures for non-stationary model evaluation are comparing the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of the models. Both estimates are used for model selection and the formulas for the AIC and BIC value of a model are given below:

$$AIC = -2\ell(\hat{\mu}, \hat{\sigma}; Y) + 2P, \tag{25}$$

$$BIC = -2\ell(\hat{\mu}, \hat{\sigma}; Y) + Plog(n), \tag{26}$$

where $\ell(\hat{\mu}, \hat{\sigma}; Y)$ is the maximized log-likelihood value, $\hat{\mu}$ and $\hat{\sigma}$ are vectors of estimates of wind covariates, Y denotes the hourly wind gusts for all heights, n is total number of observations and P is the total number of covariates. The AIC determines the efficiency of the models and focuses more on the trade-off between the goodness-of-fit and the simplicity of the model, aiming to select the model with the optimal prediction performance. On the other hand, BIC aims to identify the model that is the closest to the true sparse model if the true estimates are in the candidate list. Moreover, the penalization for a free parameter is more severe for the BIC compared to the AIC. The most efficient model has either the lowest AIC or lowest BIC values, and we aim to have these values as low as possible when tuning the hyperparameters in the cross-validation procedure.

In addition to these information criteria, we use proper scoring rules for model verification for all variable selection methods as they assess the quality of the probabilistic forecast. We use the Continuous Ranked Probability Score (CRPS) because the CRPS is reasonably robust and is not devaluated by certain bad forecasts. The instantaneous classic CRPS is defined as

$$S_{CRPS}\Big(F, y\Big) = \int_{\infty}^{\infty} \Big[F(t) - H(t-y)\Big]^2 dt, \tag{27}$$

where F is the predictive distribution with observation y and the well-known Heaviside step function $H(t-y)$. Formulations of Equation 27 can be derived for other classic distributions and the formulation of the CRPS for the GEV distribution is described in [15]. Again, we recall the cdf $G_{GEV_{\xi=0}} = \exp\Big(-\exp[-(\frac{Y(z,t)-\mu(t)}{\sigma(t)})]\Big)$ and use this in the following expression to determine the CRPS of the cGEV as:

$$S_{CRPS}\Big(G_{GEV_{\xi=0}}, Y(z,t)\Big) = \mu(t) - Y(z,t) + \sigma(t)\Big[C - \log 2\Big] - 2\sigma(t)Ei\Big(\log G_{GEV_{\xi=0}}(y)\Big), \tag{28}$$

where $G_{GEV_{\xi=0}}$ is the cdf of the GEV distribution, $Y(z,t)$ denotes the hourly wind gusts at height z at time t, $\mu(t)$ and $\sigma(t)$ are the linear combinations with wind covariates for the location and scale parameters respectively at time t, and the Euler-Mascheroni constant $C \approx 0.5772$. Furthermore, $Ei(x)$ is the exponential integral and is defined as $\int_{-\infty}^{x} \frac{e^t}{t} dt$.

A similar conclusion can be drawn for the CRPS as for the AIC and BIC where a lower value for CRPS is preferable. We use the cGEV without covariates and constant parameters as our reference probabilistic forecast and baseline performance. Then, we measure for each variable selection method the percentage improvement from the baseline model. We divide the data into a training sample and a validation sample. The training sample is used to obtain the estimates of a given method. Then, the estimates are used to evaluate the predictive forecasts of the validation sample. In other words, the CRPS is chosen as the performance measure to determine the method that is able to provide the most accurate forecast.

4.3.1 Cross-validation procedure:

Cross-validation for the hyperparameters is required to ensure the validity of models since the choice of the hyperparameters affects the model performance strongly. For every variable selection method, at least one hyperparameter needs to be tuned for the penalty term. For example with a single hyperparameter, an increase in λ leads to a stronger penalization and fewer variables will be selected as a result. In case of two or more hyperparameters, λ_1 and λ_2 describe the preference between the L_1 and L_2 penalty and we denote this ratio as $\alpha = \frac{\lambda_2}{\lambda_1+\lambda_2}$. The higher α is, the higher the preference is for the L_2 penalization. To choose the optimal value for the hyperparameters in a variable selection method, a range of values is used and the one with the

best performance measure is chosen. We repeat this process for each variable selection method and choose the CRPS as the performance criterion.

5 Results

In this section, the results for all variable selection methods are presented and evaluated. Five different methods (ridge, lasso, elastic net, adaptive lasso and adaptive elastic net) are compared to the model without covariates denoted as the baseline model, and the model including all covariates without penalization denoted as the standard model. We use the AIC and BIC values, and the CRPS as performance measures to determine the best model.

First, we need to establish the threshold for censoring the wind gusts for the cGEV model. We choose the 50% quantile of the observation for each height level. The results of censoring are given in Table 4, showing the threshold per height level and the total number of censored observations.

Table 4. Censored observations for each height level.

Censored data	10M	50M	110M	175M	250M
Threshold in meters per second	5.79	7.40	8.65	9.69	10.54
Total number observations	93025	95700	93167	92300	89106
Total censored observations	46453	47769	46564	46108	44547

Table 5, Table 6 and Table 7 display the final coefficients and standard errors obtained from the baseline model, standard model and all variable selection methods, such as ridge regression, lasso, elastic net, adaptive lasso and adaptive elastic net. Significant wind covariates that resisted the penalization in the respective method are shown in bold. Lastly, the performance measures of the variable selection methods, the corresponding values for the hyperparameters and additional information are summarized in Table 8.

Table 5. Final estimates of the baseline model with only constant parameters.

Baseline model with constant parameters					
	Coefficient	*S.E.*	*Likelihood*	*AIC*	*BIC*
μ	**8.6151**	0.0055	-531,363.0	1,062,732	1,062,752
σ	**2.4125**	0.0011			

Table 6. Final estimates and corresponding standard errors of the standard model and the variable selection methods ridge regression, lasso and elastic net.

	Variable selection method							
	Standard model		*Ridge*		*Lasso*		*Elastic net*	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
μ_0	**8.8337**	0.0051	**8.5496**	0.0053	**8.51058**	0.00581	**8.54811**	0.00537
μ_{VMAX_10M}	**0.6365**	0.0152	**0.3181**	0.0038	**0.40181**	0.00874	**0.32791**	0.00397
$\mu_{VAR_tVMAX\_10M}$	-0.0100	0.0054	**0.0256**	0.0033	0.00002	0.00023	**0.00788**	0.00339
μ_{Vh_EOF1}	**-0.5246**	0.0139	**-0.3030**	0.0037	**-0.24316**	0.00836	**-0.30904**	0.00387
μ_{Vh_EOF2}	**-0.2142**	0.0084	**-0.1007**	0.0034	-0.00028	0.00025	**-0.08698**	0.00358
$\mu_{Mean_hVh\_700}$	**0.3670**	0.0076	**0.1844**	0.0036	**0.00083**	0.00030	**0.17556**	0.00377
$\mu_{SD_hVh\_700}$	**-0.0339**	0.0049	-0.0044	0.0032	0.00004	0.00023	-0.00018	0.00083
$\mu_{Mean_hW\_700}$	**-0.0303**	0.0053	-0.0015	0.0033	0.00004	0.00023	0.00011	0.00081
$\mu_{SD_hW\_700}$	**0.0214**	0.0061	**0.0361**	0.0034	0.00012	0.00024	**0.01774**	0.00345
μ_{d_tP}	**0.1717**	0.0053	**0.0581**	0.0033	0.00035	0.00026	**0.04508**	0.00346
μ_{LI}	**-0.0697**	0.0071	**-0.0135**	0.0035	0.00004	0.00023	-0.00025	0.00085
μ_{TWATER}	**-0.1277**	0.0076	**-0.0443**	0.0035	-0.00023	0.00025	**-0.03323**	0.00356
μ_{d_tCAPE}	**-0.0363**	0.0047	-0.0035	0.0031	0.00000	0.00023	-0.00013	0.00081
μ_{Vh_SHEAR}	**-0.1028**	0.0060	**-0.0503**	0.0033	-0.00009	0.00024	**-0.03078**	0.00345
μ_{T_2M}	**-0.1584**	0.0097	**-0.0185**	0.0036	-0.00016	0.00024	-0.00504	0.00352
μ_{AC_COS}	**-0.1414**	0.0082	**0.0146**	0.0036	0.00011	0.00024	0.00095	0.00108
μ_{AC_SIN}	**-0.0161**	0.0055	**0.0254**	0.0032	0.00025	0.00025	**0.01450**	0.00325
σ_0	**0.7366**	0.0016	**0.7573**	0.0017	**0.77042**	0.00173	**0.75829**	0.00170
σ_{VMAX_10M}	**0.0636**	0.0053	**0.0800**	0.0028	**0.03665**	0.00301	**0.07347**	0.00291
$\sigma_{VAR_tVMAX\_10M}$	**0.0193**	0.0018	**0.0143**	0.0016	**0.00066**	0.00028	**0.01022**	0.00159
σ_{Vh_EOF1}	**-0.1201**	0.0048	**-0.1223**	0.0026	**-0.16493**	0.00308	**-0.13049**	0.00273
σ_{Vh_EOF2}	**0.0305**	0.0028	**0.0170**	0.0020	0.00008	0.00024	**0.01291**	0.00199
$\sigma_{Mean_hVh\_700}$	**0.0978**	0.0027	**0.0908**	0.0021	**0.05062**	0.00204	**0.08631**	0.00214
$\sigma_{SD_hVh\_700}$	-0.0033	0.0017	-0.0030	0.0016	0.00001	0.00023	0.00006	0.00073
$\sigma_{Mean_hW\_700}$	**-0.0098**	0.0018	**-0.0093**	0.0016	-0.00002	0.00023	**-0.00487**	0.00161
$\sigma_{SD_hW\_700}$	**0.0114**	0.0020	**0.0126**	0.0018	0.00040	0.00026	**0.00489**	0.00172
σ_{d_tP}	**0.0327**	0.0017	**0.0336**	0.0016	**0.00086**	0.00030	**0.03141**	0.00160
σ_{LI}	**-0.0331**	0.0026	**-0.0198**	0.0021	-0.00038	0.00026	**-0.01142**	0.00209
σ_{TWATER}	**-0.0147**	0.0027	**-0.0114**	0.0022	-0.00012	0.00024	-0.00270	0.00215
σ_{d_tCAPE}	**-0.0041**	0.0018	**-0.0035**	0.0016	0.00010	0.00024	-0.00010	0.00074
σ_{Vh_SHEAR}	**-0.0182**	0.0021	**-0.0146**	0.0018	0.00004	0.00023	**-0.00855**	0.00179
σ_{T_2M}	**-0.0477**	0.0037	**-0.0284**	0.0026	-0.00023	0.00024	**-0.01743**	0.00239
σ_{AC_COS}	**-0.0268**	0.0029	**-0.0071**	0.0022	0.00018	0.00024	-0.00043	0.00085
σ_{AC_SIN}	**-0.0148**	0.0020	**-0.0039**	0.0017	0.00000	0.00023	-0.00064	0.00085

Table 7. Final estimates and corresponding standard errors of the adaptive lasso and the adaptive elastic net.

	Variable selection method			
	Adaptive lasso		Adaptive elastic net	
	Coefficient	S.E.	Coefficient	S.E.
μ_0	**8.55294**	0.00569	**8.55959**	0.00543
μ_{VMAX_10M}	**0.46200**	0.00862	**0.36708**	0.00465
$\mu_{VAR_tVMAX\_10M}$	0.00002	0.00017	0.00000	0.00043
μ_{Vh_EOF1}	**-0.26230**	0.00837	**-0.32405**	0.00454
μ_{Vh_EOF2}	-0.00018	0.00019	**-0.04668**	0.00398
$\mu_{Mean_hVh\_700}$	**0.00050**	0.00022	**0.14110**	0.00420
$\mu_{SD_hVh\_700}$	0.00000	0.00016	-0.00007	0.00042
$\mu_{Mean_hW\_700}$	0.00000	0.00015	0.00003	0.00041
$\mu_{SD_hW\_700}$	0.00007	0.00018	0.00035	0.00047
μ_{d_tP}	0.00015	0.00018	0.00094	0.00059
μ_{LI}	0.00004	0.00017	0.00009	0.00043
μ_{TWATER}	-0.00018	0.00019	-0.00073	0.00053
μ_{d_tCAPE}	-0.00005	0.00017	-0.00003	0.00041
μ_{Vh_SHEAR}	-0.00001	0.00017	-0.00042	0.00048
μ_{T_2M}	-0.00014	0.00018	-0.00049	0.00049
μ_{AC_COS}	0.00012	0.00018	0.00034	0.00046
μ_{AC_SIN}	0.00012	0.00018	0.00060	0.00051
σ_0	**0.76652**	0.00171	**0.76016**	0.00170
σ_{VMAX_10M}	**0.01029**	0.00295	**0.05282**	0.00258
$\sigma_{VAR_tVMAX\_10M}$	**0.00044**	0.00021	**0.00329**	0.00152
σ_{Vh_EOF1}	**-0.18798**	0.00305	**-0.15341**	0.00255
σ_{Vh_EOF2}	-0.00010	0.00018	0.00060	0.00050
$\sigma_{Mean_hVh\_700}$	**0.04417**	0.00203	**0.07902**	0.00193
$\sigma_{SD_hVh\_700}$	0.00003	0.00015	-0.00007	0.00041
$\sigma_{Mean_hW\_700}$	-0.00004	0.00017	-0.00043	0.00046
$\sigma_{SD_hW\_700}$	0.00032	0.00020	0.00056	0.00048
σ_{d_tP}	**0.00052**	0.00022	**0.02229**	0.00158
σ_{LI}	-0.00033	0.00020	-0.00076	0.00052
σ_{TWATER}	0.00001	0.00015	-0.00022	0.00043
σ_{d_tCAPE}	0.00018	0.00018	0.00006	0.00041
σ_{Vh_SHEAR}	0.00002	0.00016	-0.00041	0.00046
σ_{T_2M}	-0.00002	0.00017	-0.00068	0.00051
σ_{AC_COS}	0.00006	0.00017	-0.00001	0.00041
σ_{AC_SIN}	-0.00003	0.00017	-0.00007	0.00041

The computation time for each method was approximately one hour given the values of the hyperparameter(s). The duration of the corresponding cross-validation procedure depended on the number of hyperparameters, where methods like the ridge regression and lasso had a shorter cross-validation procedure than the adaptive elastic net. The cross-validation procedure for one hyperparameter took approximately one day to find the minimum value for the CRPS improvement. For the elastic net and the adaptive lasso, the cross-validation procedure took around four days to find the optimal value. The longest cross-validation procedure was for the adaptive elastic net which took around six days for three hyperparameters.

5.1 Interpretation of the Covariates

The number of selected variables is 30, 10, 23, 10 and 11 for the ridge regression, lasso, elastic, adaptive lasso and adaptive elastic net respectively. First, we discuss the covariates that are selected in all methods. Based on Table 6 and Table 7, the majority of the wind covariates except for "VAR_t $VMAX\_10M$" for the location parameter μ and "SD_h Vh_700" for the scale parameter μ are significant in the standard model. The most informative wind covariates for the location parameter μ are the wind gust diagnostic "$VMAX\_10M$", "$V_h\_EOF_1$" and the mean horizontal wind at 700 hPa "$Mean_h$ Vh_700" across all variable selection methods. "$VMAX\_10M$" seems, as expected, to have the largest positive coefficient which has the most impact on the wind gusts statistics as it includes additional information about the maximum wind gust near a surface at 10 m. It indicates what the maximum wind gust is, such that stronger winds are more likely to appear if "$VMAX\_10M$" is large. Similarly, the coefficient for "$Mean_h$ Vh_700", the averaged absolute horizontal, is positive where a higher mean indicates an increase in wind speed as well. The second-largest covariate is the "$V_h\_EOF_1$", which is the barotropic mode that captures most of the vertical variability of the wind velocity, and has a negative coefficient. This suggests that stronger wind fluctuations correspond to stronger wind gusts.

The impact of the chosen covariates is generally weaker for σ and more covariates are selected for σ than for μ across all methods, since σ measures the variability of the cGEV model, and is therefore, more sensitive to changes. For the scale parameter σ, the most informative covariates are also "$VMAX\_10M$", "$V_h\_EOF_1$", "$Mean_h$ Vh_700", "VAR_t $VMAX\_10M$" and "d_tP" where the last two covariates are not included in μ in all methods. We already presumed that "VAR_t $VMAX\_10M$" would be positive and selected for σ. A larger "VAR_t $VMAX\_10M$" ensures that the variance of the cGEV estimates increases significantly, and is a positive coefficient in all methods. The role of "d_tP" is to measure the surface pressure tendency at which a positive

coefficient indicates a larger difference in surface pressure, such as when a cold front is passing causing stronger gusty wind. This corresponds to the data analysis of the COSMO-REA6 dataset, where stronger irregular wind gusts are more likely to appear in the winter months. Based on this analysis, we conclude that parameters associated with stronger wind gusts lead to increased estimates in μ and σ.

Out of all methods, most of the covariates in the ridge regression resisted the L_2 penalization whereas the lasso and the adaptive lasso both only selected ten covariates. This confirms that the ridge regression is not able to properly perform variable selection, which is mentioned in Section 4. The coefficients and the standard errors of the insignificant covariates are extremely close to zero in the lasso-type methods compared to the standard errors of the irrelevant covariates in the ridge regression due to the severe lasso regularization. The lasso and the adaptive lasso both select the same wind covariates with different coefficients. This is because the adaptive lasso incorporates a data-driven weighted term for each covariate unlike the constant penalty term in the lasso.

The elastic net is a combination of the ridge regression and lasso with two regularization parameters λ_1 and λ_2. It performs well in selecting the important correlated covariates from the L_2 penalization part while simultaneously keeping the coefficients and standard errors of the unnecessary covariates as low as possible due to lasso penalization. The same results can be found for the adaptive elastic net with the penalty terms λ_1 including the data-driven weights of the adaptive lasso and λ_2. In terms of the number of selected covariates, the adaptive elastic net selects significantly fewer covariates which implies that the value of the hyperparameter for the L_2 penalization is larger for the adaptive elastic net.

5.2 Evaluation Results

Table 8 presents the results, such as the maximized log-likelihood value, the AIC and BIC value, the CRPS improvement with respect to the baseline model, the corresponding value(s) of hyperparameter(s) and the number of selected wind covariates for each method. As mentioned before, the CRPS is obtained as the percentage improvement to the baseline model using the coefficients in Table 5. We aim to have a CRPS, AIC and BIC value of a given method as low as possible.

Table 8. Performance measures of all variable selection methods.

	Variable selection method					
	Standard model	*Ridge*	*Lasso*	*Elastic net*	*Adaptive lasso*	*Adaptive elastic net*
Likelihood	-488,607.2	-501,864.63	-515,359.698	-503,358.108	-514,275.076	-505,841.742
AIC	977,278.3	1,003,789	1,030,739	1,006,762	1,028,570	1,011,705
BIC	977,607.2	1,004,098	1,030,842	1,006,999	1,028,673	1,011,819
CRPS with respect to the baseline model	-0.40076	-0.78577	-0.78637	-0.78847	**-0.78984**	-0.77437
λ_1			0.086	0.07	0.072	0.024
λ_2		0.12		0.106		0.067
γ					0.088	0.013
Number of selected wind covariates	32	30	10	23	10	11

First, it is observed in Table 8 that including wind covariates into the model clearly yields better CRPS across all variable selection methods by at least 4 percentage points. In addition to that, all AIC and BIC values are significantly lower than the AIC and BIC value of the baseline model. Hence, this indicates that introducing additional information about the wind gusts into the model has a beneficial effect on the quality of the forecasts. It also increases the efficiency of the model while staying close to the true sparse model, as opposed to the baseline model. Furthermore, all variable selection methods seem to perform slightly better than the standard model except for the adaptive elastic net with respect to the CRPS.

Given Table 8, it shows that the adaptive lasso outperforms the other variable selection methods in terms of the CRPS. Regarding the number of significant covariates, the adaptive lasso also selects the fewest covariates to include in the model. However, it is lacking in terms of efficiency as it has the highest value of AIC and BIC among the methods. With regard to the AIC and BIC values, we observe that models with the lowest AIC and BIC values select a higher number of wind covariates. Still, we conclude that the adaptive lasso is the best performing method as it has the best CRPS with particularly fewer covariates.

We discussed earlier in Section 4 that a single hyperparameter controls the severity of the penalization. Table 8 shows that the hyperparameter has the value of 0.12 and 0.086 for the ridge regression and lasso respectively. Even though the hyperparameter for the ridge regression is larger, the lasso is still better capable of selecting relatively fewer covariates and yielding a better CRPS improvement than the ridge regression. In the case of two hyperparameters, α is approximately 60.23% and 73.63% for the elastic net and the adaptive elastic net respectively, emphasizing the importance of the L_2 penalization and thus, the significance of correlated covariates.

One of the most informative covariates is "$V_h\_EOF_1$", which is the barotropic wind mode. The interpretation of "$V_h\_EOF_1$" is that stronger wind gusts have more fluctuations and more variability in height. Generally, the wind gusts are not constant at each height level. For each height level, one may have to take a separate approach to modelling. To further investigate the vertical variability of the wind gusts and to observe the differences in the quality of forecasting, we perform the variable selection methods for each height level separately. For this purpose, we divide the validation sample into three categories: (1) the whole validation sample, (2) the top 5% of the strongest wind in the validation sample and (3) the validation sample without the top 5% strongest wind, to observe the differences with and without the stronger wind events. In other words, we distinguish the wind variability for each height level and determine the performance of the variable selection methods depending on the sample size and extreme wind gusts. We summarise the results of the wind variability in Table 9.

Table 9. CRPS improvement with respect to the baseline model obtained from all variable selection method per height level and different validation sample.

		Variable selection method					
		Standard model	*Ridge*	*Lasso*	*Elastic net*	*Adaptive lasso*	*Adaptive elastic net*
Whole sample	10M	0.1211	-0.0157	-0.0308	-0.0153	-0.0165	-0.0104
	50M	0.2236	0.0306	-0.0008	0.0275	0.0155	0.0285
	110M	0.2800	0.0714	0.0354	0.0689	0.0488	0.0712
	175M	0.3074	0.1617	0.1193	0.1560	0.1236	0.1526
	250M	0.2737	0.1982	0.1669	0.1952	0.1645	0.1911
Sample without top 5%	10M	0.1273	-0.0163	-0.0343	-0.0179	-0.0181	-0.0139
	50M	0.2194	0.0271	-0.0010	0.0246	0.0172	0.0285
	110M	0.3254	0.0958	0.0531	0.0924	0.0692	0.0942
	175M	0.3458	0.1780	0.1330	0.1743	0.1385	0.1717
	250M	0.2774	0.2005	0.1630	0.1965	0.1594	0.1896
Top 5% strongest wind gusts	10M	-0.4565	-0.3481	-0.3030	-0.3458	-0.3334	-0.3437
	50M	-0.5400	-0.3669	-0.2936	-0.3608	-0.3117	-0.3557
	110M	-0.5169	-0.3521	-0.2805	-0.3472	-0.2961	-0.3440
	175M	-0.4585	-0.3147	-0.2471	-0.3096	-0.2596	-0.3062
	250M	-0.4347	-0.3025	-0.2354	-0.2973	-0.2465	-0.2933

Table 9 shows that the CRPS values depend on the choice of the variable selection method, the validation sample size and the height level. Also, the CRPS decreases as the height level increases in general. Adding more wind covariates to the model is particularly beneficial to the lower height levels. As for the whole validation sample, we observe that the baseline model with constant parameters is the best choice, except for the sample at 10 m, in all variable selection methods in terms of the CRPS. Among all methods, the lasso and the adaptive lasso seem to outperform the other methods by having the best CRPS improvements at each height level.

The results of the whole validation and the validation sample without the strongest winds are almost identical, and a similar conclusion can be drawn for the validation sample without the strongest wind gusts. The baseline model provides better and more accurate forecasts than the ones including the wind covariates except for observations at 10 m where there is a slight improvement in CRPS of approximately 1-3% for all methods.

For the top 5% of strongest wind, all variable selection methods including the standard model outperform the baseline model by at least 23 percentage points. The biggest improvement of the CRPS occurs when all wind covariates are included in the model, at all height levels shown in the standard model. Yet, the standard model gives the worst forecasts for the whole sample and the validation sample without the strongest wind events. The more wind covariates are selected, the larger the improvement is in the CRPS for all height levels in all methods. Two things are notable in Table 9 for the sample with extreme wind gusts. First, the results of the adaptive elastic net are slightly worse than the results of the ridge regression and the elastic net but the adaptive elastic net selects significantly fewer wind covariates. If one wants to obtain similar results with less computation time and a smaller number of selected covariates, the adaptive elastic net might be the appropriate method. Secondly, the selected covariates are the same for the lasso and the adaptive lasso, as we mentioned before. The results show, however, that the CRPS for the adaptive lasso is clearly better due to the individual data-driven weights. In this case, the adaptive lasso might be the optimal solution over the lasso. Following these results, we conclude that the adaptive lasso and the adaptive elastic net are the best performing methods in terms of CRPS with respect to the validation sample size. They provide the lowest CRPS in the sample size of the strongest wind gusts as well as having one of the best CRPS for the other sample sizes. However, the adaptive lasso has already proven to provide the best results overall in Table 8, significantly better results than the adaptive elastic net. Therefore, we conclude in both scenarios that the adaptive lasso is the best performing method.

6 Conclusion

This research focuses on comparing various state-of-the-art variable selection methods for the censored generalized extreme value (cGEV) model, in order to find the most informative covariates, which have a significant impact on the wind gust statistics in Germany. Also, it provides an overview of the variable selection methods for non-stationary GEV models. The cGEV model is based on a generalized extreme value distribution with censoring. The threshold for censoring is at the 50% quantile of the observations at each height level and the Gumbel-type GEV distribution is used in the cGEV model. For this research, two different datasets have been

used: (1) the wind gusts observed at the Hamburg Weather Mast used to derive the wind gusts statistics and (2) the COSMO-REA6 reanalysis utilized as a proxy for the wind characteristics. The wind characteristics are incorporated into the location and scale parameter using a linear combination. In this research, five different regularization variable selection methods are used, which are a combination of common methods and relatively new ones. The five variable selection methods are ridge regression, lasso, elastic net, adaptive lasso and adaptive elastic net. Estimates of the methods are obtained using a maximum likelihood estimation and the variable selection methods are evaluated in terms of AIC and BIC values, and the CRPS for comparison purposes.

The results of this research reveal that the most informative covariates are the wind gust diagnostic at 10 m "$VMAX\_10M$" and its variance "$VAR_t\ VMAX\_10M$", the barotropic mode "$V_h\_EOF_1$", the mean of the horizontal wind speed at 700 pHa "$Mean_h\ Vh\_700$", and the surface pressure tendency "d_tP", which were selected by all variable selection methods. Among the variable selection methods, the adaptive lasso is the best performing method with a CRPS improvement of 7.8984% for the whole validation sample. The CRPS of the adaptive lasso is better than the CRPS of the lasso, in which the same ten wind covariates are selected due to the individual data-driven weights in the adaptive lasso. Furthermore, this result is slightly better than the second-best CRPS improvement, which is 7.8847% of the elastic net with significant fewer variables for the adaptive lasso.

We further investigate the wind variability across all height levels. The results reveal that adding wind covariates to the model is particularly valuable for wind gusts at 10 m. Furthermore, it also shows that the CRPS is improved by at least 23 percentage points if the wind covariates are included for the top 5% strongest wind forecasts. Based on the results, the adaptive lasso performs well overall, for the daily wind gusts and the top 5% strongest wind gusts at each height level in terms of CRPS.

7 Limitations and Future Research

We recommend five areas in which this research could be further improved on. First, it should be noted that the hyperparameters for all variable selection methods are obtained through a cross-validation procedure with a step size of 0.001. To further refine this research, one should consider a cross-validated grid-search and the validation technique k-fold for finer results.

Second, five variable selection methods are chosen for comparison in this research where two methods are adaptive. One suggestion is to investigate other variable selection methods, for example, other lasso-extensions methods, such as the weighted lasso, group lasso, bayesian

lasso, etc. or variable selection using neural network models [8].

Third, we assume a linear relationship between the wind covariates and the cGEV parameters in this research. It would be interesting to explore this relationship further by considering other types of functions, for example, using a conditional density network (CDN) based on a neural network [4].

Fourth, the results of this research are obtained by presuming that the values of the wind covariates are equal at every height level. It was also revealed that adding covariates was not always beneficial at each height level. Therefore, the cGEV model could be further improved by incorporating different height levels of the wind gusts to determine the influence of the wind covariates per height level, allowing for more flexibility.

Lastly, the GEV distribution was used to model wind gusts. Other variations of the GEV distribution could be further explored, for example, the Burr-Generalized Extreme Value mixture distribution [26] and the wind covariates could be included using a linear combination.

Bibliography

[1] Bailey, A.: Wind pressures on buildings. Selected Engineering Papers 1(139) (1933)

[2] Bollmeyer, C., Keller, J., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Pscheidt, I., et al.: Towards a high-resolution regional reanalysis for the european cordex domain. Quarterly Journal of the Royal Meteorological Society 141(686), 1–15 (2015)

[3] Brewick, P., Divel, L., Butler, K., Bashor, R., Kareem, A.: Consequence of urban aerodynamics and debris impact in extreme wind events. In: Proceedings of the 11th Americas conference on wind engineering, San Juan, Puerto Rico. p. 17 (2009)

[4] Cannon, A.J.: A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. Hydrological Processes: An International Journal 24(6), 673–685 (2010)

[5] Cannon, D.J., Brayshaw, D.J., Methven, J., Coker, P.J., Lenaghan, D.: Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in great britain. Renewable Energy 75, 767–778 (2015)

[6] Cao, Q., Ewing, B.T., Thompson, M.A.: Forecasting wind speed with recurrent neural networks. European Journal of Operational Research 221(1), 148–154 (2012)

[7] Carta, J.A., Ramirez, P., Velazquez, S.: A review of wind speed probability distributions used in wind energy analysis: Case studies in the canary islands. Renewable and sustainable energy reviews 13(5), 933–955 (2009)

[8] Castellano, G., Fanelli, A.M.: Variable selection using neural-network models. Neurocomputing 31(1-4), 1–13 (2000)

[9] Coles, S., Casson, E.: Extreme value modelling of hurricane wind speeds. Structural Safety 20(3), 283–296 (1998)

[10] D'Amico, G., Petroni, F., Prattico, F.: Wind speed prediction for wind farm applications by extreme value theory and copulas. Journal of Wind Engineering and Industrial Aerodynamics 145, 229–236 (2015)

[11] Davenport, A.G.: The application of statistical concepts to the wind loading of structures. Proceedings of the Institution of Civil Engineers 19(4), 449–472 (1961)

[12] El Adlouni, S., Ouarda, T.B., Zhang, X., Roy, R., Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research 43(3) (2007)

[13] Emmert-Streib, F., Dehmer, M.: High-dimensional lasso-based computational regression models: regularization, shrinkage, and selection. Machine Learning and Knowledge Extraction 1(1), 359–383 (2019)

[14] Friederichs, P., Göber, M., Bentzien, S., Lenz, A., Krampitz, R.: A probabilistic analysis of wind gusts using extreme value statistics. Meteorologische Zeitschrift 18(6), 615 (2009)

[15] Friederichs, P., Thorarinsdottir, T.L.: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. Environmetrics 23(7), 579–594 (2012)

[16] Friedman, J., Hastie, T., Tibshirani, R.: A note on the group lasso and a sparse group lasso. arXiv preprint arXiv:1001.0736 (2010)

[17] Gao, L., Schulz, K., Bernhardt, M.: Statistical downscaling of era-interim forecast precipitation data in complex terrain using lasso algorithm. Advances in Meteorology 2014 (2014)

[18] García-Hinde, O., Terrén-Serrano, G., Hombrados-Herrera, M., Gómez-Verdejo, V., Jiménez-Fernández, S., Casanova-Mateo, C., Sanz-Justo, J., Martínez-Ramón, M., Salcedo-Sanz, S.: Evaluation of dimensionality reduction methods applied to numerical weather models for solar radiation forecasting. Engineering Applications of Artificial Intelligence 69, 157–167 (2018)

[19] Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The elements of statistical learning: data mining, inference, and prediction, vol. 2. Springer (2009)

[20] Heinze, G., Wallisch, C., Dunkler, D.: Variable selection–a review and recommendations for the practicing statistician. Biometrical journal 60(3), 431–449 (2018)

[21] Henderson, A.R., Morgan, C., Smith, B., Sørensen, H.C., Barthelmie, R.J., Boesmans, B.: Offshore wind energy in europe—a review of the state-of-the-art. Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology 6(1), 35–52 (2003)

[22] Hoerl, A.E., Kennard, R.W.: Ridge regression: applications to nonorthogonal problems. Technometrics 12(1), 69–82 (1970)

[23] Holmes, J.: Wind action on glass and brown's integral. Engineering Structures 7(4), 226–230 (1985)

[24] Jeon, J.J., Sung, J.H., Chung, E.S.: Abrupt change point detection of annual maximum precipitation using fused lasso. Journal of Hydrology 538, 831–841 (2016)

[25] Joensen, A., Nielsen, T., Madsen, H.: Non-parametric statistical methods for wind power prediction. In: EWEC-CONFERENCE-. pp. 788–791. BOOKSHOP FOR SCIENTIFIC PUBLICATIONS (1997)

[26] Jung, C., Schindler, D., Laible, J.: National and global wind resource assessment under six wind turbine installation scenarios. Energy conversion and management 156, 403–415 (2018)

[27] Junginger, M., Faaij, A., Turkenburg, W.C.: Cost reduction prospects for offshore wind farms. Wind engineering 28(1), 97–118 (2004)

[28] Justus, C., Hargraves, W., Mikhail, A., Graber, D.: Methods for estimating wind speed frequency distributions. Journal of applied meteorology 17(3), 350–353 (1978)

[29] Kestens, E., Teugels, J.L.: Challenges in modelling stochasticity in wind. Environmetrics: The official journal of the International Environmetrics Society 13(8), 821–830 (2002)

[30] Lee, B.H., Ahn, D.J., Kim, H.G., Ha, Y.C.: An estimation of the extreme wind speed using the korea wind map. Renewable energy 42, 4–10 (2012)

[31] Li, G., Shi, J.: Application of bayesian model averaging in modeling long-term wind speed distributions. Renewable Energy 35(6), 1192–1202 (2010)

[32] Liu, H.H., Ong, C.S.: Variable selection in clustering for marketing segmentation using genetic algorithms. Expert Systems with Applications 34(1), 502–510 (2008)

[33] Martin, D., Zhang, W., Chan, J., Lindley, J.: A comparison of gumbel and weibull statistical models to estimate wind speed for wind power generation. In: 2014 Australasian Universities Power Engineering Conference (AUPEC). pp. 1–6. IEEE (2014)

[34] Mínguez, R., Méndez, F., Izaguirre, C., Menéndez, M., Losada, I.J.: Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables. Environmental Modelling & Software 25(12), 1592–1607 (2010)

[35] Ogutu, J.O., Schulz-Streeck, T., Piepho, H.P.: Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In: BMC proceedings. vol. 6, pp. 1–6. Springer (2012)

[36] Ouarda, T.B., Charron, C.: Non-stationary statistical modelling of wind speed: A case study in eastern canada. Energy Conversion and Management 236, 114028 (2021)

[37] Ouarda, T.B., Charron, C., Kumar, K.N., Phanikumar, D.V., Molini, A., Basha, G.: Non-stationary warm spell frequency analysis integrating climate variability and change with application to the middle east. Climate Dynamics 53(9), 5329–5347 (2019)

[38] Park, T., Casella, G.: The bayesian lasso. Journal of the American Statistical Association 103(482), 681–686 (2008)

[39] Patel, M.R., Beik, O.: Wind and solar power systems: design, analysis, and operation. CRC press (2021)

[40] Perrin, O., Rootzén, H., Taesler, R.: A discussion of statistical methods used to estimate extreme wind speeds. Theoretical and applied climatology 85(3), 203–215 (2006)

[41] Peterka, J.A., Cermak, J.E.: Wind pressures on buildings-probability densities. Journal of the structural division 101(6), 1255–1267 (1975)

[42] Powell, M.D., Vickery, P.J., Reinhold, T.A.: Reduced drag coefficient for high wind speeds in tropical cyclones. Nature 422(6929), 279–283 (2003)

[43] Rahimi, E., Rabiee, A., Aghaei, J., Muttaqi, K.M., Nezhad, A.E.: On the management of wind power intermittency. Renewable and Sustainable Energy Reviews 28, 643–653 (2013)

[44] Rose, S., Jaramillo, P., Small, M.J., Grossmann, I., Apt, J.: Quantifying the hurricane risk to offshore wind turbines. Proceedings of the National Academy of Sciences 109(9), 3247–3252 (2012)

[45] Safari, B.: Modeling wind speed and wind power distributions in rwanda. Renewable and Sustainable Energy Reviews 15(2), 925–935 (2011)

[46] Şahin, A.D.: Progress and recent trends in wind energy. Progress in energy and combustion science 30(5), 501–543 (2004)

[47] Sanchez-Pinto, L.N., Venable, L.R., Fahrenbach, J., Churpek, M.M.: Comparison of variable selection methods for clinical predictive modeling. International journal of medical informatics 116, 10–17 (2018)

[48] Seguro, J., Lambert, T.: Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis. Journal of wind engineering and industrial aerodynamics 85(1), 75–84 (2000)

[49] Steinheuer, J., Friederichs, P.: Vertical profiles of wind gust statistics from a regional reanalysis using multivariate extreme value theory. Nonlinear Processes in Geophysics 27(2), 239–252 (2020)

[50] Stevens, M., Smulders, P.: The estimation of the parameters of the weibull wind speed distribution for wind energy utilization purposes. Wind engineering pp. 132–145 (1979)

[51] Tian, S., Yu, Y., Guo, H.: Variable selection and corporate bankruptcy forecasts. Journal of Banking & Finance 52, 89–100 (2015)

[52] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1), 267–288 (1996)

[53] Tramblay, Y., Neppel, L., Carreau, J., Sanchez-Gomez, E.: Extreme value modelling of daily areal rainfall over mediterranean catchments in a changing climate. Hydrological Processes 26(25), 3934–3944 (2012)

[54] Tyralis, H., Papacharalampous, G., Tantanee, S.: How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset. Journal of Hydrology 574, 628–645 (2019)

[55] Wang, J., Hu, J., Ma, K.: Wind speed probability distribution estimation and wind energy assessment. Renewable and sustainable energy Reviews 60, 881–899 (2016)

[56] Zhou, Y., Wu, W., Liu, G.: Assessment of onshore wind energy resource and wind-generated electricity potential in jiangsu, china. Energy Procedia 5, 418–422 (2011)

[57] Ziel, F., Steinert, R., Husmann, S.: Efficient modeling and forecasting of electricity spot prices. Energy Economics 47, 98–111 (2015)

[58] Zou, H.: The adaptive lasso and its oracle properties. Journal of the American statistical association 101(476), 1418–1429 (2006)

[59] Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology) 67(2), 301–320 (2005)

[60] Zou, H., Zhang, H.H.: On the adaptive elastic-net with a diverging number of parameters. Annals of statistics 37(4), 1733 (2009)

A Data analysis of observed wind gusts at the Hamburg Weather Mast

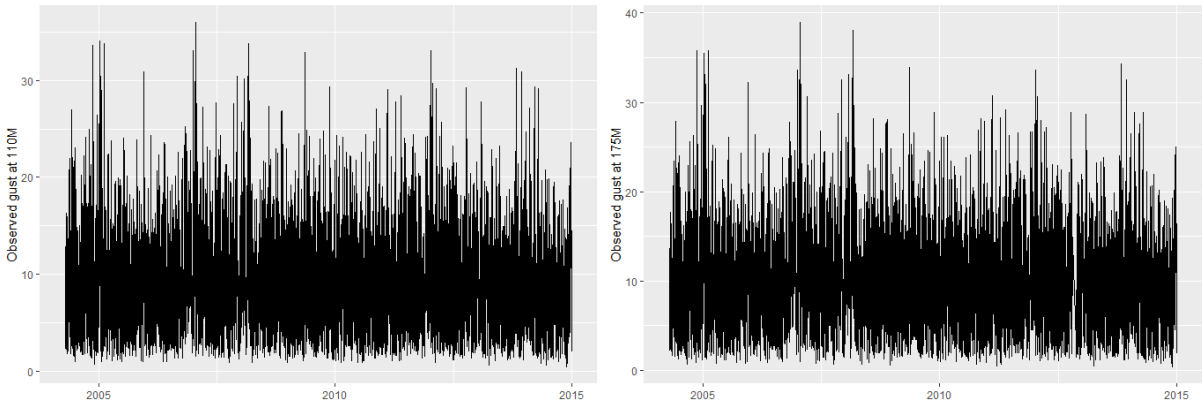

Figure 6. Observed wind gusts from 1 January 2004 until 31 December 2014 at 110M (left) and 175M (right).

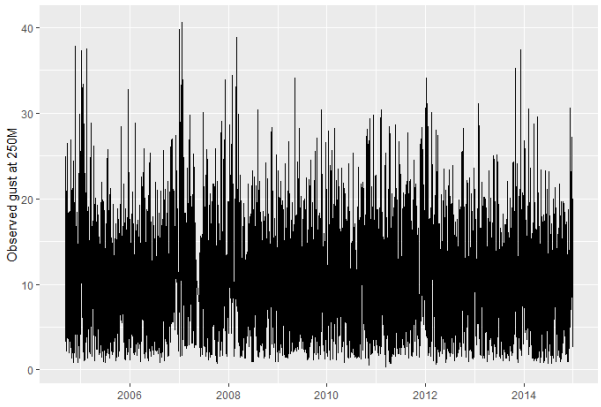

Figure 7. Observed wind gusts from 1 January until 31 December 2014 at 250M.

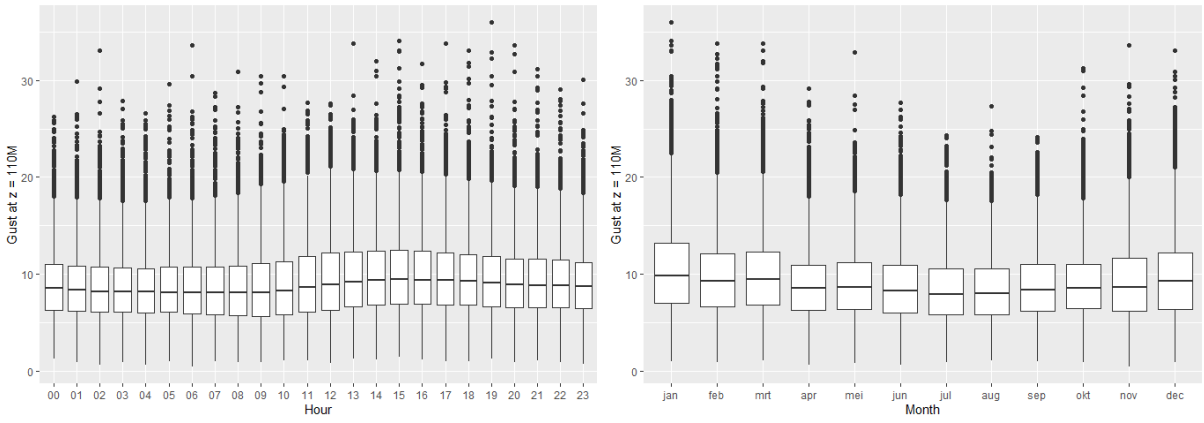

Figure 8. Boxplots of hourly (left) and monthly (right) observed wind gusts at 110M.

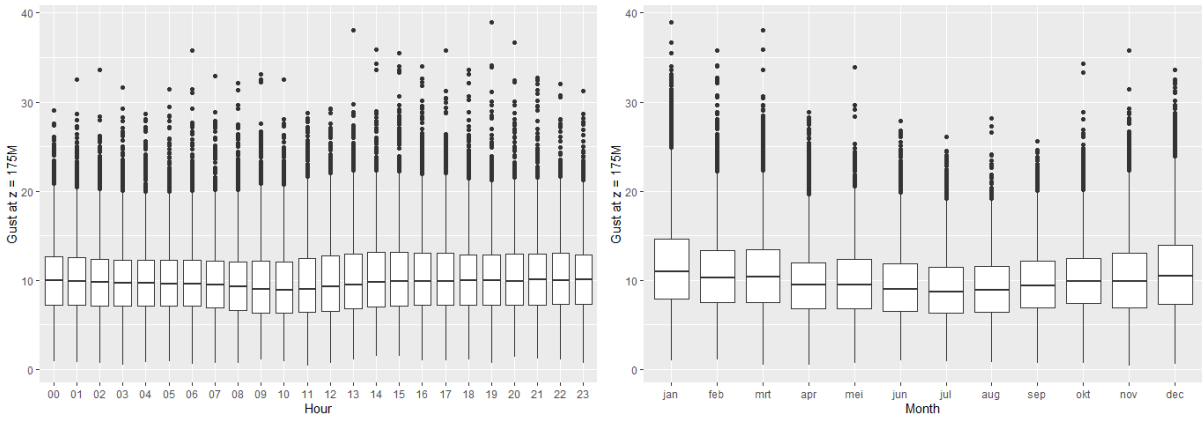

Figure 9. Boxplots of hourly (left) and monthly (right) observed wind gusts at 175M.

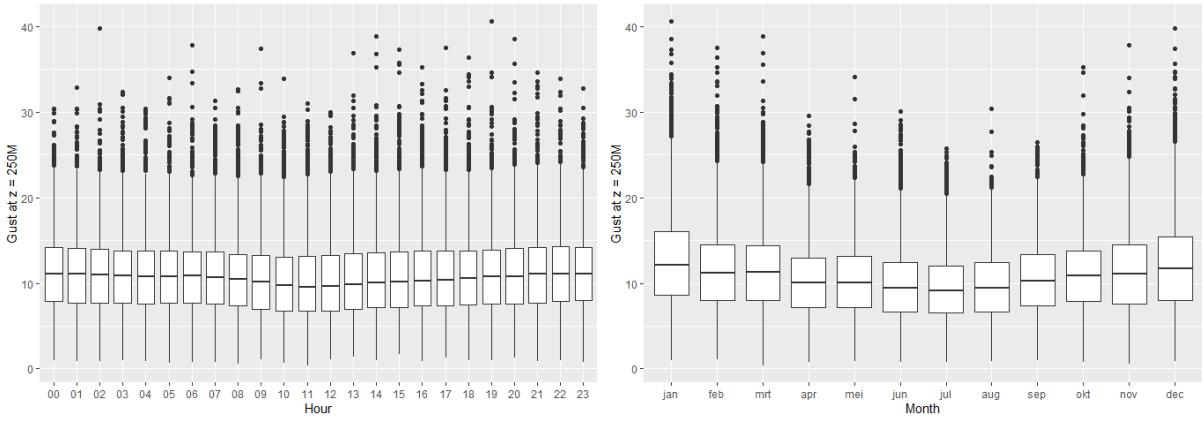

Figure 10. Boxplots of hourly (left) and monthly (right) observed wind gusts at 250M.