Spatial modelling of extreme wind speed in the Netherlands

JIATAN JIANG
2, 2016

SUPERVISORS:
Prof. dr. ir. A. Stein (Alfred)
Dr. M. Oesting (Marco)

Spatial modelling of extreme wind speed in the Netherlands

JIATAN JIANG
Enschede, The Netherlands, 02.2016

SUPERVISORS:
Prof. dr. ir. A. Stein (Alfred)
Dr. M. Oesting (Marco)

THESIS ASSESSMENT BOARD:
Prof.dr.ir. M.G. Vosselman
Prof.dr. G. van der Steenhoven
Drs. J.P.G. (Wan) Bakx.

ABSTRACT

Extreme wind speeds, especially maximum wind speeds, have great impacts on human society and natural environments in both positive and negative aspects. It is one of the most important source of renewable energy, but it can also disrupt traffic, damage farming and destroy buildings. Extreme wind speed modelling, therefore, should be done to obtain the distribution, the regularities and probabilities of extreme wind speed events. In theory, there are well-developed methods to solve univariate extreme model in single location. But there would be difficult to model regional extreme events with multi-variate extreme value theories.

This thesis explore the use of the spatial extreme model, which is based on max-stable processes, for fitting extreme events in two aspects, the extreme spatial dependence structure and marginal distributions. The composite likelihood method is used to fit annual maximum wind speed in 44 years and for 35 stations in the Netherlands. The spatial extreme model obtained can be used to predict and analyse extreme wind speed events in the Netherlands. The results show that compared with the generalized extreme value (GEV) distribution in single meteorological station, the max-stable processes model has a lower degree of dispersion and smoother values of extreme wind speeds. This can be regarded as the modification of single location models since it do not model the spatial dependence for different stations. The results also shows the spatial dependence between different stations through the extremal coefficients. With station distances from 50km to 300km, the value extremal coefficient ranges from 1.6 to 1.8, which means the correlation becomes weaker with the increasing distance. Moreover, maps of return level in 25-year, 50-year, 100-year and 200-year return period in the Netherlands are given in the thesis, this can provide a reference for the construction projects. These maps also tell that there are correlation between the value of distance from stations to coastline and the distribution of extreme wind speeds. The extreme wind speed will increase from the south to the north and from east to the west parts of the Netherlands. The study conducted that missing data that are regularly occurring in wind speed data can be dealt with and that the max-stable process model was adequate to model extreme wind speed in the Netherlands.

Keywords: Spatial modelling, Extreme wind speed, Max-stable processes, Data and model selection, Composite likelihood

.

ACKNOWLEDGEMENTS

I would like to thank my first supervisor Prof. dr. ir. A. Stein (Alfred) in faculty of ITC for his immense knowledge and he always gave me guide in my research and thesis writing. He always steered me in the right direction and his office was always open for me.

I would like to thank my second supervisor Dr. M. Oesting (Marco) at University of Siegen for his patience whenever I ran into a trouble spot or had a question about my research or writing. I could not have imagined having a better advisor and mentor for my study.

I would also like to thank the experts who will attend my final defense: Prof.dr.ir. M.G. Vosselman, Prof.dr. G. van der Steenhoven and Drs. J.P.G. (Wan) Bakx.

Finally, I must show my very profound gratitude to my parents, L. Wang and other friends for supporting and encouraging me to finish my study, researching and writing this thesis. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation

Nowadays, renewable energy is becoming more and more important since it is an environmentally friendly power and its source is available everywhere. The Netherlands is a country with rich wind power and great experience in wind power using. According to the data by the Dutch government, by December 2013, 1,975 wind turbines were operational on land in the Netherlands with an aggregate capacity of 2,479 MW (Centraal Bureau voor de Statistiek, 2015). With plenty of wind turbines, safety should be taken into consideration since the integrity of turbines' structures can be threatened by extreme wind speeds (Palutikof, Brabson, Lister, & Adcock, 1999). Furthermore, extreme wind speeds have a great negative impact on human society and natural environments, for instance they may disrupt air, train and bus travel, damage the growth of crops and livestock and destroy man-made structure. So it is necessary to do research on extreme wind speeds, especially on its trends and laws of spatial distribution.

Since extreme events have their particularities, classical statistical models have limitations and cannot be used in the analysis because of the following problems: firstly, the extreme wind speed event is an expression of wind speed variability, that means wind is a stochastic event, and climatic factors which cause wind speed variability are also stochastic and not easy to identify. So it is difficult to model stochastic extreme wind speed events. However, there are statistical regularities in extreme wind speed values, so sufficient and effective statistical models can be built to analyse the regularities and probabilities of extreme wind speed events. Secondly, for spatial data analysis, a key insight is that observations in space cannot typically be assumed to be mutually independent and that those observations that are close to each other are likely to be similar (Lee et al., 2013). Studies on a single location are not sufficient to interpret the intrinsic link between different locations throughout the region, which is called the spatial dependence or correlation structure of the regional extremes. So it is necessary to study and analyse the probability of regional extreme wind speed events and their distribution. For the vast area of land in Netherlands, it is important to study models for spatial extremes, to understand their correlation and predict the occurrence regularities and probabilities.

1.2. Study background

Over the past few decades, statistical theory on extremes has been developed fast and comprehensively on single locations. It has been widely used in environmental statistical analysis and extreme models can explain laws of random variation in environmental phenomena well. Some extreme models which are used in analysis of rainfall, wind speeds, temperature and sea-level can perfectly obtain the characteristics of the corresponding environmental phenomena on single locations and are theoretically dependent on the univariate extreme model, which has some limitations in the space modelling and analysis. Firstly, environmental changes are often area-related, but univariate extreme models cannot consider the dependence between stations. For example, if a strong wind affects a large area, wind speed data from weather stations in this area surely represent the regional correlation; and also when there is a flood, all cities and villages located around the river will be affected, but this regional correlation cannot be theoretically explained in univariate extreme models. So the multivariate extreme model has been developed. However, multi-variate extreme model can explain the question of regional correlation theoretically, but for practical application, since there are many dimensions (for wind speed data, each weather station is a dimension), the curse of dimensionality is always met in practice, and this makes the model fit badly. For example, in research of regional rainfall, if there are too many weather stations like 20 or 50, the multi-variate model cannot even be used, and the regional correlation of environmental phenomena cannot be explained as well.

Because of the limitations mentioned above, more spatial extreme models have been developed rapidly on both theoretical and practical aspects, the aims of these models and methods are to use analyse, quantify and explain the change of regional environmental phenomena like wind speed, precipitation and temperature. For example, Coles & Casson (1998), Casson & Coles (1999)and Sang & Gelfand (2009) use latent variables processes to model and analyse the spatial extremes, however, the results appear unreasonable because they cannot explain the risk behaviour of spatial extremes well; Heffernan & Tawn (2004)and Butler, Heffernan, Tawn, & Flather (2007) use a conditional model of extreme events to analyse extreme values, it explains the spatial dependence of different weather stations, but this model cannot be used in places where there are no weather stations.

Considering the limitations, a class of models called max-stable processes was introduced by de Haan (1984). Max-stable processes are a class of asymptotically-justified models that are capable of representing spatial dependence among extreme values, and also an extension of multivariate extreme value theory (Shaby, 2012). A first max-stable process model built from spectral representation is designed by Smith (1990). After that many researchers have studied max-stable processes widely, and different spatial max-stable process models have been proposed such as the Schlather model built by Schlather (2002), the Brown-Resnick model generalized by Kabluchko, Schlather, & de Haan (2009)and the extremal-t model built by Davison, Padoan, & Ribatet (2012). There are two advantages in using max-stable processes: firstly, this model can calculate the multi-variate distribution family without dimension curse; secondly, max-stable processes are generally a spatial model, it can be used to explain spatial problems which multi-variate models cannot solve, like spatial dependence structure and spatial data aggregation problems. So spatial models for extreme values which is based on max-stable processes can be used for more detailed methods for statistics and analysis and provide explanation on regional commonality and correlation.

This thesis builds a model for regional extreme wind speed events. The model is based on max-stable processes. The aim is to analyse the multi-station extreme wind speed data in the Netherlands. In this thesis we will use the composite likelihood method for modelling analysis and statistical inference. Furthermore, this thesis will analyse the spatial dependence structure and infer the characteristics of spatial dependence of extreme wind speeds in the Netherlands. There is also another paper which use the similar data to model spatial extremes in the Netherlands by Ribatet (2013). In that paper, annual wind speed data from 1971-2012 are used as data to find best model to describe the distribution and makes some simulation. However, the difference between Ribatet (2013) and this research is as follows. Firstly, as missing wind speed data will affect the model building, so further discussion and methods are used to choose research data and keep its quality. Secondly, since the research area are the European part of the Netherlands but not just single locations, so spatial dependence should be taken into consideration and thus the distance between different meteorological stations should be calculated, so variables of longitude and latitude should be transformed to other geographic coordinate to calculate the distance in this research. Thirdly, the results of Ribatet (2013) tells that the distribution of maximum wind speed has some connection with the coast distance of observed stations, so for further study, the coast distance of stations becomes one variable in our research to find the most fitted model.

The following parts are the introduction of thesis structure: the first chapter is the introduction of the thesis, it tells the motivation and background of the thesis. The second chapter has two parts, the first parts reviews basic theories in the thesis, including extreme theory, which focus on the generalized extreme value (GEV) distribution, max-stable processes (MSP), the composite likelihood method and the Takeuchi's information criterion (TIC). The second parts are the data and research area of the thesis. The third chapter is results, it elaborates the methods and results of data and model selection. With the most fitted model, spatial analysis are taken to learn the characteristics of extreme wind speed events in the Netherlands. The forth chapter is about discussion of the thesis. And the last chapter draws the conclusion of the research and summarizes the whole thesis.

1.3. Objectives

The objectives of this thesis are as follow,

a) Since the Netherlands has great opportunity to meet extreme wind events, it is necessary to model the extreme wind speed in the Netherlands and study its distribution, probability and regularity.
b) For distribution of extreme wind events, the model can show wind speed return level for specific return period, this can help prevent extreme wind disasters.
c) For correlation of extreme wind events between different locations, the model can show the intrinsic correlation between different locations in one wind events.
d) For extreme wind speed in single location and regional area, different model should be chosen. Comparison and analysis between different results from these two models can show which model is better fitting with the Netherlands. This will help improve the precision of the model.

1.4. Research questions

The research question of this thesis are as follows,

a) Which method are chosen to be the best method to model extreme wind speed events in the Netherlands?
b) Which variables should be considered into model of extreme wind speed events since there are many factors which can cause extreme wind speed events?
c) How to reflect the correlation between different meteorological stations in one extreme wind events?
d) Which method are chosen to select the extreme wind speed data since there are missing data?
e) Which methods are chosen for model-selection, and why?
f) What can we obtain from the results of the spatial analysis of extreme wind speed model in the Netherlands?

2. METHODS

2.1. Extreme value theory

This section gives an introduction on theoretical fundamentals of extreme value theory which will be used in the thesis, the first part is the introduction and theorem of univariate extreme value theory and generalized extreme value (GEV) distributions. The second part are the introduction and theorem of max-stable processes.

2.1.1. Univariate extreme value theory

The distribution of extreme events in single locations has been widely studied on modelling its tails and many applications have been presented (Chen et al., 2015; García-cueto & Santillán-soto, 2007). The basic theorem of method for modelling extreme events in single location is the existence of a domain of attraction for the maximum, that is: Let $X_1, \ldots, X_n$ be independent identically distributed (i.i.d.) observation of daily maximum wind speed. We suppose that are random and have the distribution function *F*. Then there exists two sequences (a_n) with n ∈ ℕ and (b_n) with n ∈ ℕ where $a_n > 0$, $b_n \in \mathbb{R}$, and a distribution function *G* such that

$$\lim_{n\to\infty} P\{\frac{\max_{1\le i\le n} X_i - b_n}{a_n} \le x\} = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x). \quad (2.1)$$

The distribution function *G* such that (2.1) is referred to as the generalized extreme value (GEV) (Bortot & Gaetan, 2013). It has the parametric form

$$G(x) = exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}, \quad (2.2)$$

for location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$. This distribution combines three extreme probability distributions, the Gumbel, Fréchet and Weibull distributions into one model. It makes generalized extreme value (GEV) distribution a continuous probability distribution in extreme events at single location. In extreme wind speed distribution, the location parameter μ represents the mean value of extreme wind speeds, the scale parameter σ represents the range of extreme wind speeds and the shape parameter ξ represents the distribution model among these three models.

For extreme wind speeds, we cannot use just the maximum value of one sample for analysis and modelling. Therefore a practical way is to separate samples into multiple blocks, extract the extreme value in each block and set the block maxima in a sequence. The sequence can be regarded as the independent and identically GEV distribution only if the amount of samples in each block is large enough (Faranda et al., 2011). We can set the original daily extreme wind speed data as $X_1, X_2, \ldots, X_{NK}$, and *Z* is the annual extreme wind speed over all data and time. The sequence of block maxima is shown as, $Z_i = \max_{N(i-1)+1\le j\le N*i} X_j$, $i = 1,2,\ldots,k$. In this formula, the Z_i is the maximum wind speed in year *i*, *k* is the number of blocks and parameter *N* is the number of samples in one block (number of days in each year) and function $N(i-1)+1 \le j \le N*i$ represents the first day and last day of each year. With the interpretation maximum wind speeds sequence Z_i follows the GEV distribution.

However, spatial data like wind speed or precipitation are collected at different places. Therefore the joint distribution of extreme values is of interest by researchers. Since univariate extreme value theory cannot represent the joint distribution, multivariate extreme value theory is needed, especially max-stable processes.

2.1.2. Max-stable processes

The aims of statistical modelling and analysis of extreme wind speed events are to fit all stations' data into the model, find the laws of marginal distribution and spatial dependence between different stations and predict extreme wind speeds in space and time from places where there is no weather station. To meet this requirement, max-stable processes, the extreme value theory of stochastic processes is introduced into the research. It can be regarded as the extension of multivariate extreme distribution. Since max-stable processes can model and quantify the spatial dependence of extreme events, they can be used in this thesis to study the distribution of extreme wind speed throughout the country, e.g. the Netherlands. The following parts will be the introduction of the definition, function, characterizing and fitting methods of max-stable processes.

A stochastic process Z is said to be max-stable if there exist continuous normalizing functions $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that

$$\mathrm{Z} \overset{d}{=} \max_{i=1,\ldots,n} \frac{Z_i - b_n}{a_n}, \quad \mathrm{n} \in \mathbb{N}, \tag{2.3}$$

where the Z_i are independent copies of Z (Ribatet, 2013). In this thesis, Z_i is the maximum wind speed in day *i*, *n* is the size of blocks and usually for 365 years. $\underline{\underline{d}}$ represents that the left and right parts in (2.3) has the same distribution, that means the distribution of process Z is fit for the distribution of formula (2.4) below.

In spatial modelling of extreme events, max-stable processes are widely used because they arise as the pointwise maximum taken over an infinite number of (appropriately rescaled) stochastic processes (Ribatet, 2013). More precisely, for a stochastic process *X,* which represent the extreme wind speed, if $\{ X_i : i \in \mathbb{N} \}$ is the collection of its independent copy and there exists normalizing functions $\{c_n > 0\}$ and $\{d_n \in \mathbb{R}\}$, the extreme wind speed X can be shown as

$$\max_{i=1,\ldots,n} \frac{X_i(s) - d_n(s)}{c_n(s)} \to Z(s), \quad n \to \infty, \quad s \in S, \tag{2.4}$$

Then *Z* is called max-stable process (de Haan, 1984). In the extreme wind speed events, the $X_i(s)$ represent a random variable of daily maximum wind speed in location *s* and day *i*, *n* is the size of blocks.

In extreme wind speed events, the marginal distribution is the generalized extreme value (GEV) distribution with unknown parameters. It is however difficult to model the marginal distribution and the spatial dependence at the same time. So for theoretical purposes, data transformation can be done to make the data following a simple distribution. The unit Fréchet distribution is usually chosen as the simple distribution for wind speed. This helps to study the spatial dependence of the data. It can be assumed that the limiting process Z is simple, that is, with unit Fréchet margins (De Haan & Ferreira, 2006), i.e.

$$Pr\{Z(s) \leq z\} = exp\left(-\frac{1}{z}\right), \quad z > 0, \quad s \in S, \tag{2.5}$$

we call this the simple max-table processes. This means in extreme wind speed events, all marginal distributions at individual observation in stations will follow the unit Fréchet distribution with same parameters. Therefore the probability of different stations are all the same. This theoretical assumption does benefits on the study of spatial dependence, but for concrete applications, the marginal distribution of extreme events varying in space (Ribatet, 2013). So other methods should be taken in this thesis.

To consider the representation of max-stable processes, researchers paid attention on it and proposed a method called spectral representation. Let $\{\zeta_i : i \in \mathbb{N}\}$ be the points of a Poisson process on $(0, \infty)$ with intensity of $\zeta^{-2}d\zeta$ (Ribatet, 2013). There exists a non-negative stochastic process Y with continuous sample paths such that $\mathbb{E}\{Y(s)\}=1$ for all $x \in \mathbb{R}^2$ and for which Z has the same distribution as

$$\max_{i\geq 1} \zeta_i Y_i(s), \quad s \in S, \tag{2.6}$$

Where Y_i are independent copies of Y. In practical application, the spectral representation of a max-stable process has its explanation. In the extreme wind speed event, the ζ_i is the magnitude of extreme wind in location *s*, stronger wind has higher magnitude. $Y_i(s)$ represents the spatial profile and stochastic distribution of this extreme wind events. Therefore $\zeta_i Y_i(s)$ represents the daily wind speed event, $\max_{i\geq 1}$ is the function that extract the maximum daily wind speeds. Our attention in this thesis will focus on the characterization (2.6).

The finite dimensional distribution of a max-stable process can be obtained from the spectral characterization (2.6). For $z_1, \ldots, z_k > 0$ and the location $s_1, \ldots, s_k$, the distribution equation

$$Pr\{Z(s_1) \leq z_1, \ldots, Z(s_k) \leq z_k\} = exp\left[-\mathbb{E}\left\{\max_{j=1,\ldots,k} \frac{Y(s_j)}{z_j}\right\}\right]. \tag{2.7}$$

In (2.7), the $Y(s_j)$ represents the spatial profile of extreme wind at location s_j, and z_j is individual copy of max-stable process *Z*, also the distribution of extreme wind events.

Based on (2.3) and (2.4), max-stable processes can be characterized by several models, and they have been proposed or applied in the last few decades, like Brown & Resnick (1977), Smith (1990), Schlather (2002) and Kabluchko, Schlather, de Haan (2009). There are three models that we choose in this research, the Schlather model, the Brown–Resnick model and the extremal–t model. The difference between the three models are as follows (Ribatet, 2013):

a. The Schlather model: according to the formula (2.6), we set

$$Y_i(x) = \sqrt{2\pi}\, max\{0, \varepsilon_i(x)\}, \tag{2.8}$$

where ε_i are independent copies of a standard Gaussian process with correlation function ρ, which are chosen as powered exponential correlation function family $\rho(h) = exp\{-(h/\lambda)^k\}$.

b. The Brown-Resnick model: in formula (2.6), we set

$$Y_i(x) = exp\{\varepsilon_i(x) - \sigma^2(x)/2\}, \tag{2.9}$$

where ε_i are independent copies of a centered Gaussian process with stationary increments, and we choose $\gamma(h) = (h/\lambda)^k$ as semi-variogram model.

c. The Extremal-*t* model: according to formula (2.6), we set

$$Y_i(s) = c_v max\{0, \varepsilon_i(s)\}^v, \quad c_v = \sqrt{\pi}2^{-\frac{v-2}{2}}\Gamma\left(\frac{v+1}{2}\right)^{-1}, \quad v \geq 1, \tag{2.10}$$

where $\varepsilon_i(s)$ are standard Gaussian process with correlation function ρ and Γ are the Gamma function. This model is a generalization of Schlather model (Opitz, 2013, Ribatet & Sedki, 2013). Its bivariate cumulative distribution equals

$$Pr\{Z(s_1) \leq z_1, Z(s_2) \leq z_2\} = exp[-\frac{1}{z_1}T_{v+1}\left\{-\frac{\rho(s_1-s_2)}{b} + \frac{1}{b}\left(\frac{z_2}{z_1}\right)^{\frac{1}{v}}\right\}$$

$$-\frac{1}{z_2}T_{\nu+1}\left\{-\frac{\rho(s_1-s_2)}{b}+\frac{1}{b}\left(\frac{z_1}{z_2}\right)^{\frac{1}{\nu}}\right\}], \tag{2.11}$$

where T_ν is the cumulative distribution function of a Student random variable with ν degrees of freedom and $b^2 = \{1-\rho(x_1-x_2)^2\}/(\nu+1)$.

The difference between these three models are the choice of spatial dependence model and the function of correlation coefficients. In practical application like extreme wind speed events, the difference of spatial model means the impact of one extreme wind event to different locations are different, it will affect the distribution of extreme wind speeds to a large extent.

When considering the method to assess the ability of max-stable processes to capture the spatial dependence structure of extreme events, a convenient way is using the extremal coefficient function (Schlather & Tawn, 2003),

$$\theta(s_1-s_2) = -zlogPr\{Z(s_1)\le z, Z(s_2)\le z\} = \mathbb{E}\,[max\{Y(s_1), Y(s_2)\}]. \tag{2.12}$$

The extremal coefficient function takes value in the interval [1,2], the lower bound $\theta = 1$ represent complete dependence and the upper bound $\theta = 2$ is correspond to complete independence (Ribatet & Sedki, 2013). In this thesis, the formula (2.12) has its practical meaning. The extremal coefficients of extreme wind speeds can be obtained from extracting the maximum of stochastic distribution of extreme wind event in location s_1 and s_2. If $\theta = 1$, it means that the extreme wind speed event in location s_1 and s_2 are totally dependent, if there is an extreme wind event happens in location s_1, this wind event definitely happens in s_2. And if the $\theta = 2$, it means one extreme wind event will never affect location s_1 and s_2 at the one time.

2.1.3. Composite likelihood method

In spatial extreme model based on max-stable processes, the format of data are like $\{Z_n(s), s = 1,2,\dots,S; n = 1,2,\dots,N\}$, s represents the stations and n is the block, $Z_n(s)$ is the maximum data in station s and block n. Since one block represents one year in the thesis, so $Z_n(s)$ means annual maximum wind speed data, and the joint inference of data from research area stations is close related to the spatial extreme model analysis, but classical maximum likelihood method is intractable to be used in such problems, because there is no joint probability density functions of annual maximum wind speed for all stations. However, the bivariate cumulative distribution function and marginal distribution of the max-stable processes can be obtained, so we can use composite likelihood method, and focus on the pairwise likelihood, to deal with the model fitting problems. The method is introduced by Padoan et al. (2009).

Assume that $z = (z_1,\dots,z_k)$ are single observations, then the pairwise log-likelihood is

$$l_p(\psi; z) = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\omega_{i,j}logf(z_i, z_j;\psi), \tag{2.13}$$

where $\omega_{i,j}$ are suitable non-negative weights and $f(;\psi)$ are the bivariate density. From the formula above we can see composite likelihoods are linear combination of log-likelihoods (Ribatet, 2013).

Since the composite likelihood estimator has the same regularity condition with the maximum likelihood estimator, and the parameter ψ is recognizable in (2.10), so the maximum pairwise likelihood estimator is shown as follows,

$$\widehat{\psi}_p = \underset{\psi \in \Psi}{\arg\max}\, l_p(\psi; z). \qquad (2.14)$$

2.1.4. Model selection

It is necessary to compare different models and select a best one when facing different spatial extreme models. According to the composite likelihood method, difference between models are based on the marginal distribution parameters and geographic information of stations.

Normally, the model selection method which used on the non-nested model is as follows. Since we pay attention on the method of composite likelihood, we define a composite Kullback-Leibler divergence based on (2.12) and g and f_ψ are the two statistical models that need to compare(Varin and Vidoni, 2005)

$$D_P(f_\psi; g) = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \omega_{i,j}\ \mathbb{E}\left\{log\, \frac{g(Z_i, Z_j)}{f_\psi(Z_i, Z_j; \psi)}\right\}, \qquad (2.15)$$

where $Z = (Z_1, \ldots, Z_k) \sim g$. In (2.14), methods of composite likelihood are adopted to compare the model g and model f_ψ to find which model is more precision and simpler., results of model selection can be judged by the minimize value of composite information criterion because of the composite Kullback-Leibler divergence consists off a linear combination of Kullback-Leibler divergences

$$TIC(f_\psi) = -2l_p(\widehat{\psi_p}; Z) + 2tr\{\hat{J}(\psi_0) * \hat{H}(\psi_0)^{-1}\}, \qquad (2.16)$$

where $\hat{H}(\psi_0)$ and $\hat{J}(\psi_0)$ are consistent estimator of the matrices $H(\psi_0)$ and $J(\psi_0)$. Formula (2.16) is a generalization of the Takeuchi's information criterion (TIC) (Ribatet, 2013). The first part of the formula $\{-2l_p(\widehat{\psi_p}; Z)\}$ describe the precision of model, high score means more precise of the model; the second part $\{2tr\{\hat{J}(\psi_0) * \hat{H}(\psi_0)^{-1}\}\}$is penalty factor, low score means more simple of the model

2.2. Data and research area

2.2.1. Research area

The research area in this thesis is the European parts of Netherlands (Netherland also has territory in three Caribbean islands), which is located in western Europe and lies between latitudes 50° and 54°N, longitudes 3°and 8°E. There is Germany in the east, Belgium and Luxembourg in the south, and the North Sea in the northwest. The Netherlands shares maritime borders with Belgium, the United Kingdom and Germany. It is geographically a very low and flat country, with about 26% of its area and 21% of its population located below sea level, and only about 50% of its land exceeding one meter above sea level (Centraal Bureau voor de Statistiek, 2015). There are 35 weather stations provided by KNMI, the basic information of the weather stations are listed in Table 2.1 Geographic information of weather stations and the location of the weather stations are shown in Figure 2.1.

Station No.	NAME	Lon(east)	Lat(north)	Alt(m)	X(km)	Y(km)	Coast Distance(km)
210	Valkenburg	4.419	52.165	-0.2	88.736	464.426	3.977
225	Ijmuiden	4.575	52.463	4.4	99.779	497.446	0.873
235	De kooy	4.785	52.924	0.5	114.477	548.596	4.811
240	Schiphol	4.774	52.301	-4.4	113.149	479.290	18.955
242	Vlieland	4.942	53.255	0.9	125.261	585.350	0.406
249	Berkhout	4.979	52.644	-2.5	127.346	517.352	23.929
251	Hoorn	5.346	53.393	0.5	152.230	600.616	2.708
260	De Bilt	5.177	52.101	2	140.569	456.885	53.086
265	Soesterberg	5.274	52.13	13.9	147.220	460.069	57.847

267	Stavoren	5.384	52.896	2.6	154.755	545.311	44.719
269	Lelystad	5.526	52.458	-4	164.403	496.599	63.458
270	Leeuwarden	5.755	53.225	1.5	179.532	581.982	26.187
273	Marknesse	5.889	52.703	-3.1	188.884	523.955	82.094
275	Deelen	5.888	52.061	50	189.310	452.532	99.536
277	Lauwersoog	6.196	53.409	3	208.751	602.697	9.744
278	Heino	6.263	52.437	4	214.517	494.603	113.232
279	Hoogeveen	6.575	52.75	15.6	235.158	529.726	85.706
280	Eelde	6.586	53.125	3.5	235.208	571.459	45.654
283	Hupsel	6.65	52.073	29	241.534	454.501	148.188
286	Nieuw Beerta	7.15	53.196	0.2	272.760	580.133	61.311
290	Twenthe	6.897	52.273	34.5	258.001	477.070	142.382
310	Vlissingen	3.596	51.442	8	30.460	385.086	0.082
319	Westdorpe	3.862	51.226	1.4	48.446	360.641	28.993
323	Wilhelminadorp	3.884	51.527	1.4	50.673	394.087	16.312
330	Hoek van Holland	4.124	51.993	12.5	68.224	445.604	0.498
340	Woensdrecht	4.349	51.448	14.9	82.809	384.387	47.491
344	Rotterdam	4.444	51.955	-4.8	90.143	441.043	20.041
348	Cabauw	4.927	51.972	-0.7	123.351	442.613	44.833
350	Gilze-rijen	4.933	51.568	11.1	123.483	397.670	67.407
356	Herwijnen	5.145	51.858	0.9	138.287	429.859	64.388
370	Eindhoven	5.414	51.446	20.3	156.832	384.001	103.330
375	Volkel	5.706	51.657	21.1	177.027	407.520	108.959
377	Ell	5.764	51.197	30	181.304	356.370	138.165
380	Maastricht	5.768	50.91	114	181.749	324.447	157.674
391	Arcen	6.196	51.498	19	211.124	390.096	147.067

Table 2.1 Geographic information of weather stations including the longitude, latitude, altitude, X and Y coordinates.

In Table 2.1, parameter lon and lat represent the longitude (east) and latitude (north) of meteorological stations, and since the distances between stations cannot be calculated directly with the longitude and latitude, so geographic coordinates of stations are needed. Parameter X and Y are the parameter of geographic coordinate which are defined and transformed by the longitude and latitude. The defined geographic coordinate are "GCS_AMERSFOOT", and projection coordinate are "RD_NEW". Amersfoort / RD New is a projected coordinate reference system last revised on 05/27/2005 and is suitable for use in Netherlands.

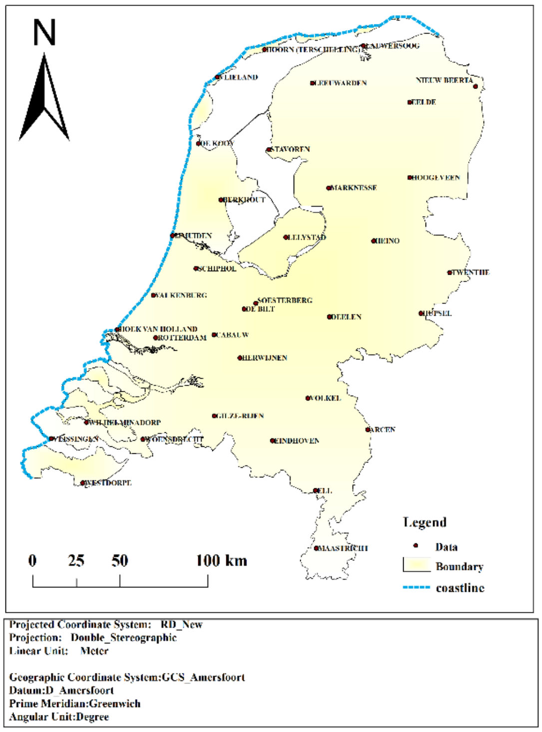


Figure 2.1 Locations of the 35 Netherlands weather stations and the blue line shows the coastline of the Netherlands.

2.2.2. Data

The data used in the thesis are wind speed series (km/h) observed at 35 stations located all around the Netherlands, and they are obtained from the Royal Netherlands Meteorological Institute (http://www.knmi.nl/home). In the thesis, we set one year as a block and extract annual maximum wind speeds from the raw data, daily maximum wind speed, for the time period 1971--2014. As a result, we can obtain 44 data for one station and set data from all stations into one sequence. We write the sequence as $\{z_s^n, s = 1,2, \ldots ,35; n = 1,2, \ldots ,44\}$, where *s* represent stations and *n* represent years.

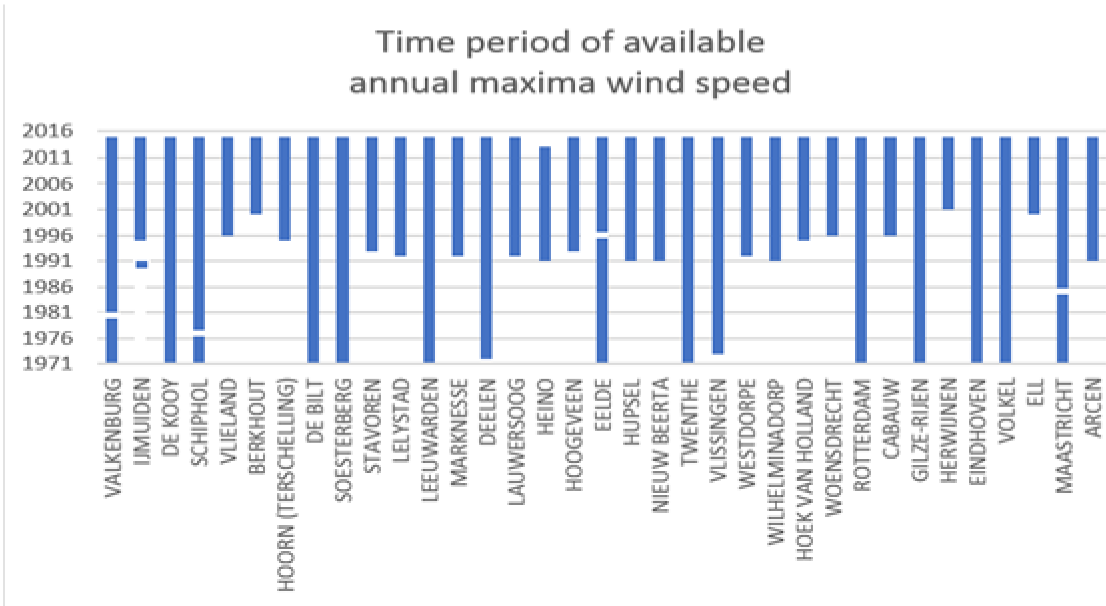
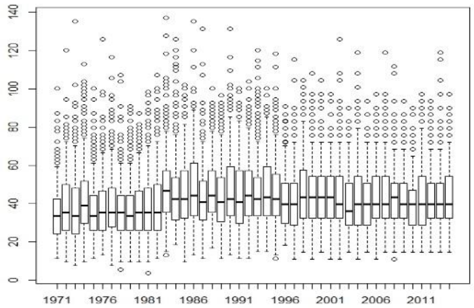
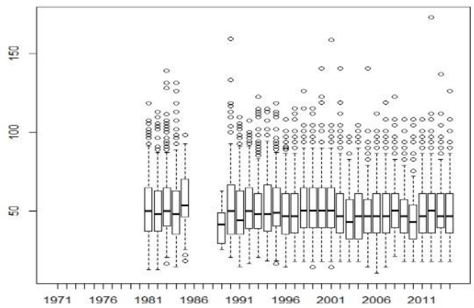


Figure 2.2 Time period of available annual maximum wind speed.

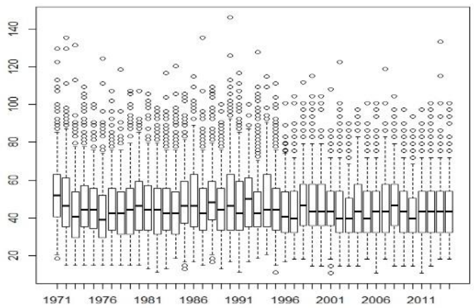
Figure 2.2 shows that there are some data missing in part of the stations. Wind speed data are the basic of further research. But in actual measurements, data missing and exceptions might be happened because of the instrument failure, this will affect the reasonableness and completeness of the research. In order to fully and effectively use the measured data and get the more precise research results, it is essential and necessary to assess the quality and validity of extreme wind speed data, and detailed examinations of extreme wind data are needed before the research.



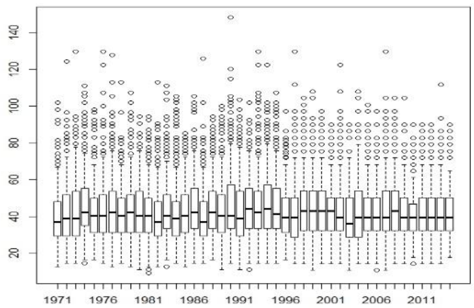
210 VALKENBURG



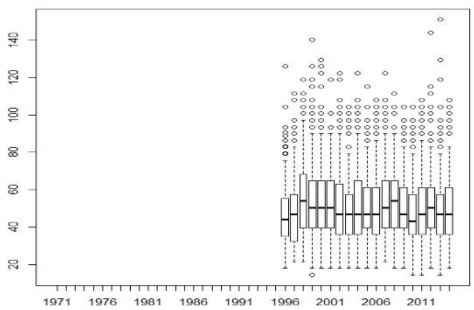
225 IJMUIDEN



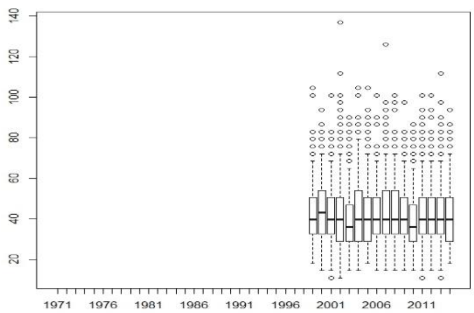
235 DE KOOY



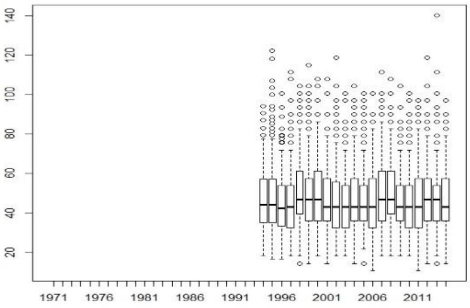
240 SCHIPHOL



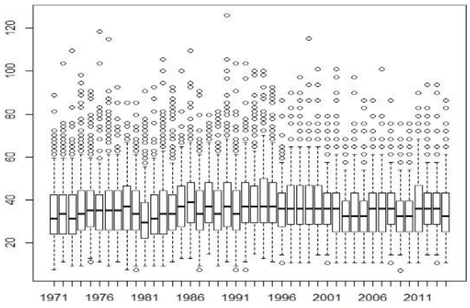
242 VLIELAND



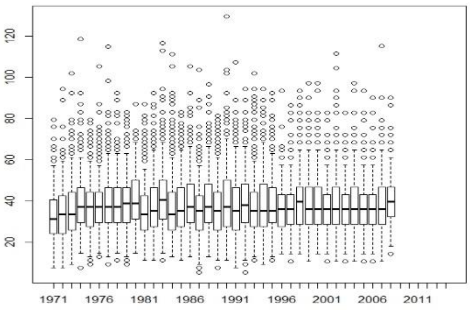
249 BERKHOUT



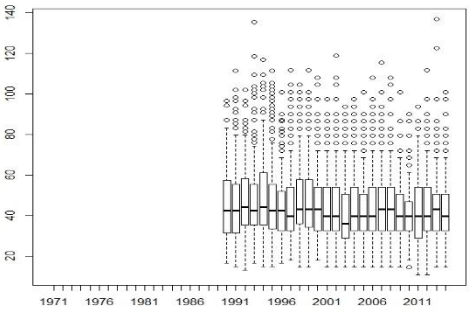
251 HOORN



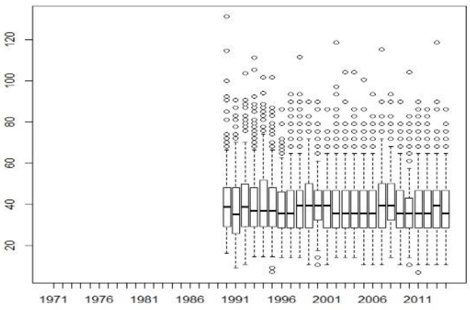
260 DE BILT



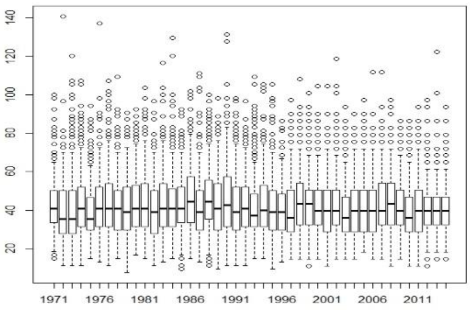
265 SOESTERBERG



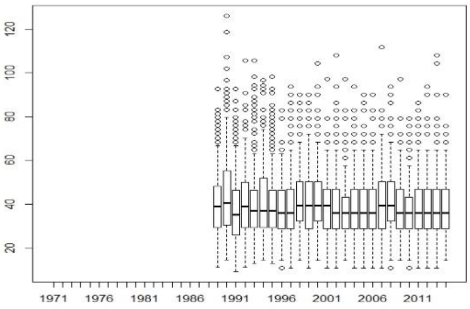
267 STAVOREN



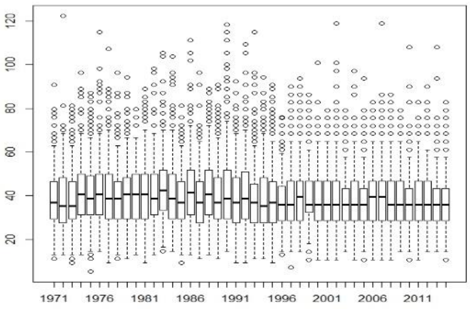
269 LELYSTAD



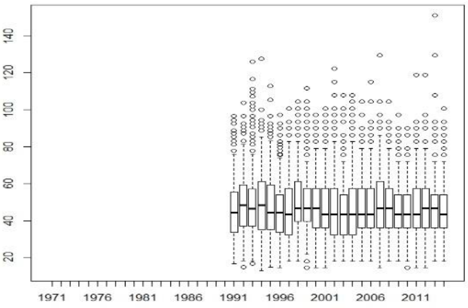
270 LEEUWARDEN



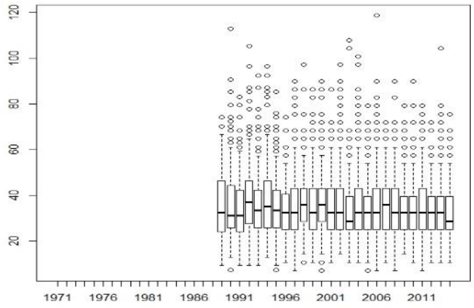
273 MARKNESSE



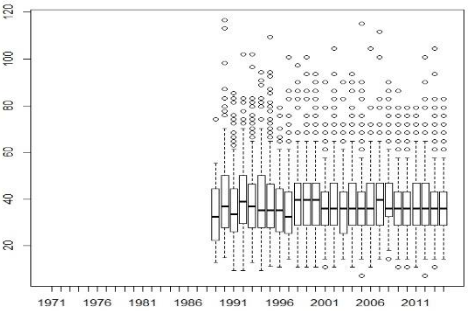
275 DEELEN



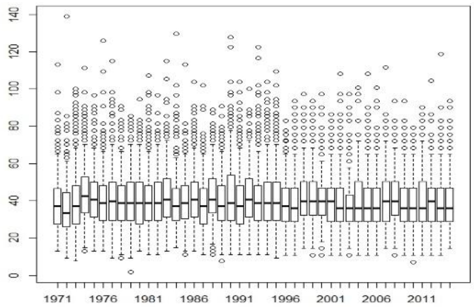
277 LAUWERSOOG



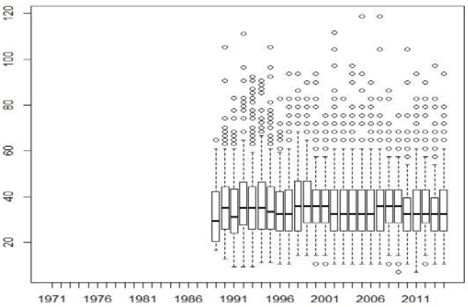
278 HEINO



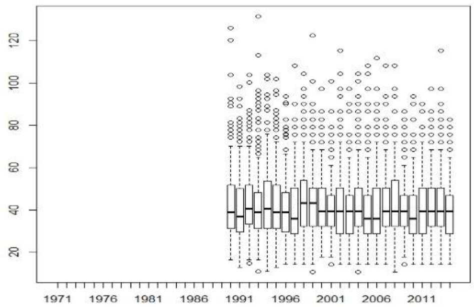
279 HOOGEVEEN



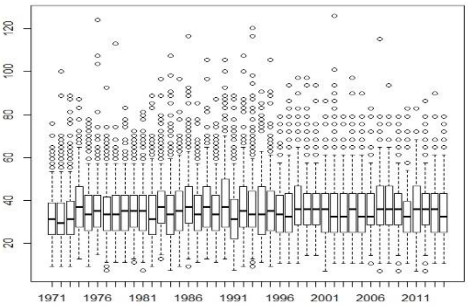
280 EELDE



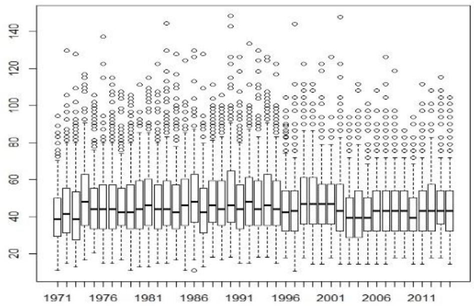
283 HUPSEL



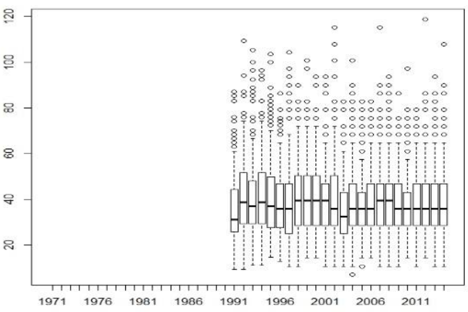
286 NIEUW BEERTA



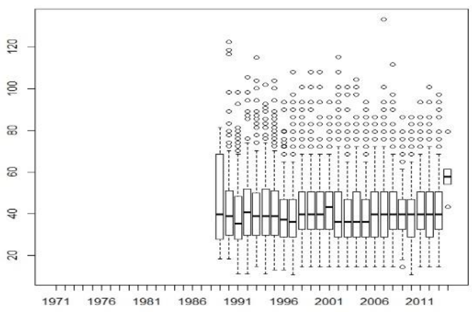
290 TWENTHE



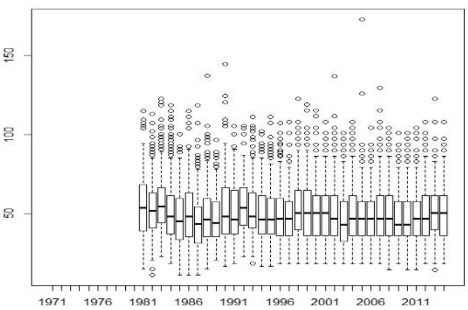
310 VLISSINGEN



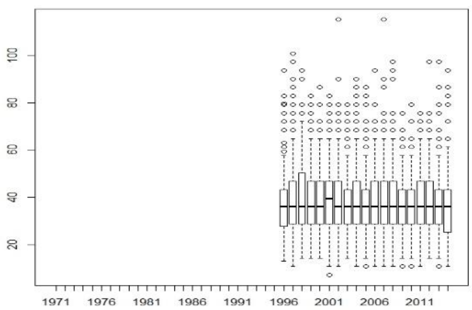
319 WESTDORPE



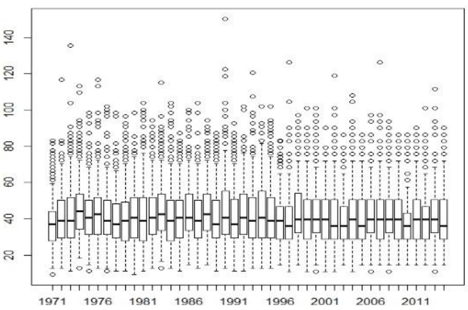
323 WILHELMINADORP



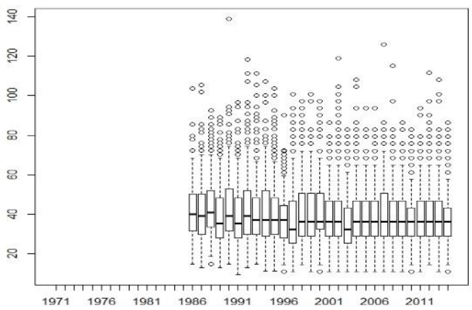
330 HOEK VAN HOLLAND



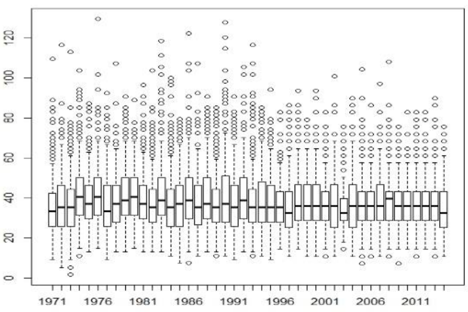
340 WOENSDRECHT



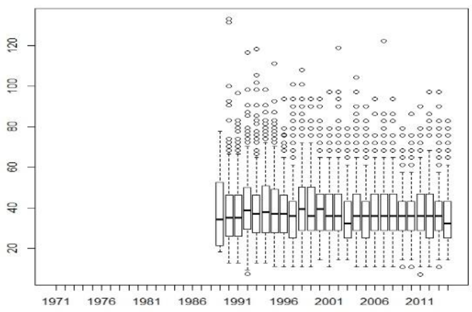
344 ROTTERDAM



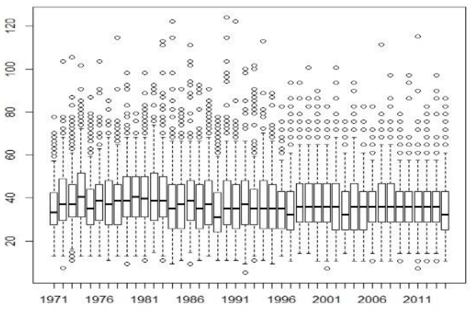
348 CABAUW



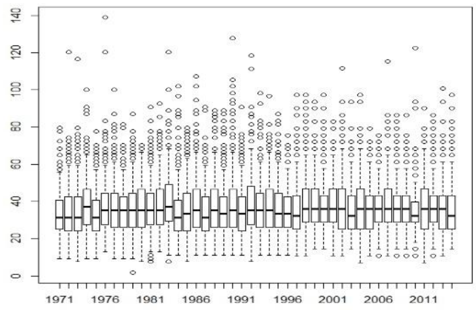
350 GILZE-RIJEN



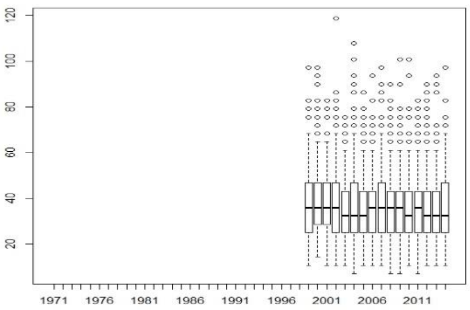
356 HERWIJNEN



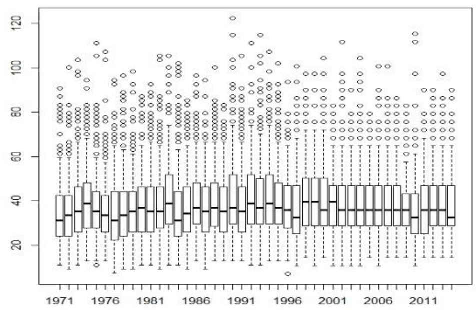
370 EINDHOVEN



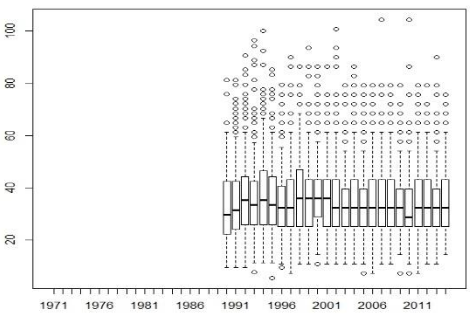
375 VOLKEL



377 ELL



380 MAASTRICHT



391 ARCEN

Figure 2.3 Boxplot of daily maximum wind speed (km/h) from 1971 to 2014 in 35 stations.

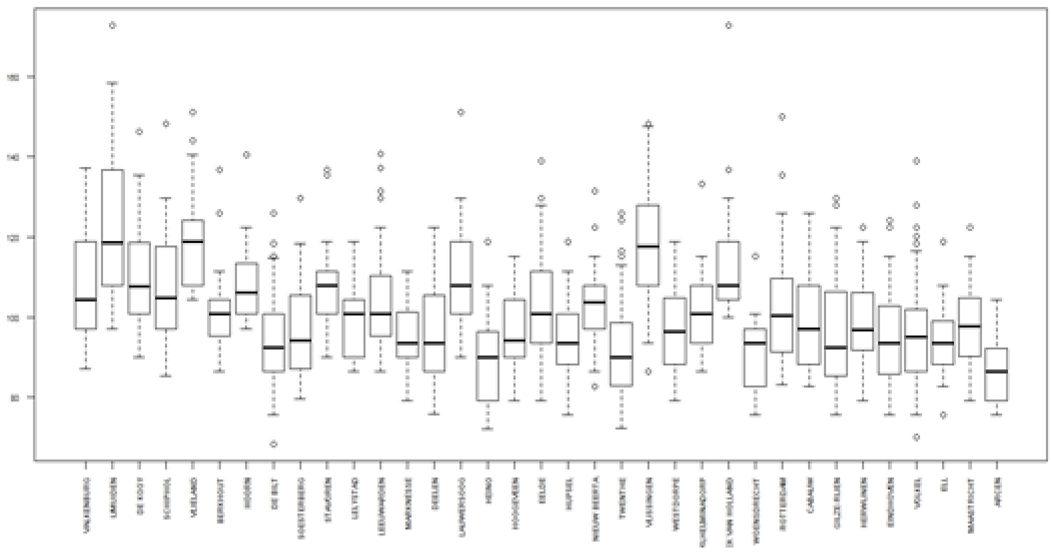


Figure 2.4 Boxplot of annual maximum wind speed (km/h) from 1971 to 2014 in 35 stations.

In statistical analysis, boxplots can show the distribution function of data clearly. Figure 2.3 and Figure 2.4 are the boxplot of daily and yearly maximum wind speed in 35 meteorological stations. Boxplot figures has 5 parts, the bottom and top of the box are always the first and third quartiles, and the band inside the box is always the second quartile (the median), if the band is not in the middle of the box, it shows that the data contains skewness.

From the Figure 2.3 we see some regularities of distribution of daily maximum wind speeds in time and space series. For single stations, the difference between the top and bottom and the length of boxes are not too much, that means the distribution of daily maximum wind speeds in one locations are similar in time series (Massart et al., 2005). From the aspect of space, since the second quartile of boxplot can roughly represent the average daily maximum wind speed in each location, from Figure 2.3 we can see that the larger wind speed the station has, the closer coast distance the station has. It is also a reflection on the results got from Ribatet (2013), the coast distance from stations may has some connection with the distribution of maximum wind speeds. From Figure 2.4, the boxplot of annual maximum wind speed in each station, we see clearly that the difference because of the length of the boxes are not the same.

3. RESULTS

3.1. Data selection

The occurrence of missing data is a serious research problem, and the discontinuity of meteorological records becomes a major limitation of research and adds complexity to the analysis. Data in this research are daily maximum wind speeds provided by the Royal Netherlands Meteorological Institute (KNMI). Our aim was to extract the yearly maximum wind speeds. As Figure 2.2 shows, due to the instrument failure, human error and environmental influence, part of data provided by KNMI are missing. The occurrence of missing data should be taken into consideration. Typically it is important to know how many data are missing in one year and how they will actually affect the quality of yearly data and the results of the research.

To solve this problem, the occurrences of annual maximum wind speeds in each month are counted clearly, wind speeds have seasonal characteristics and most high wind speeds concentrate in a few months. That means that wind speeds in specific months have a higher probability to become yearly maximum wind speed, whereas wind speed in other months have a much lower probability. Therefore, data in different months have different influence, also call those as different weights, on yearly maximum wind speed. The resulting histogram is shown in Figure 3.1.

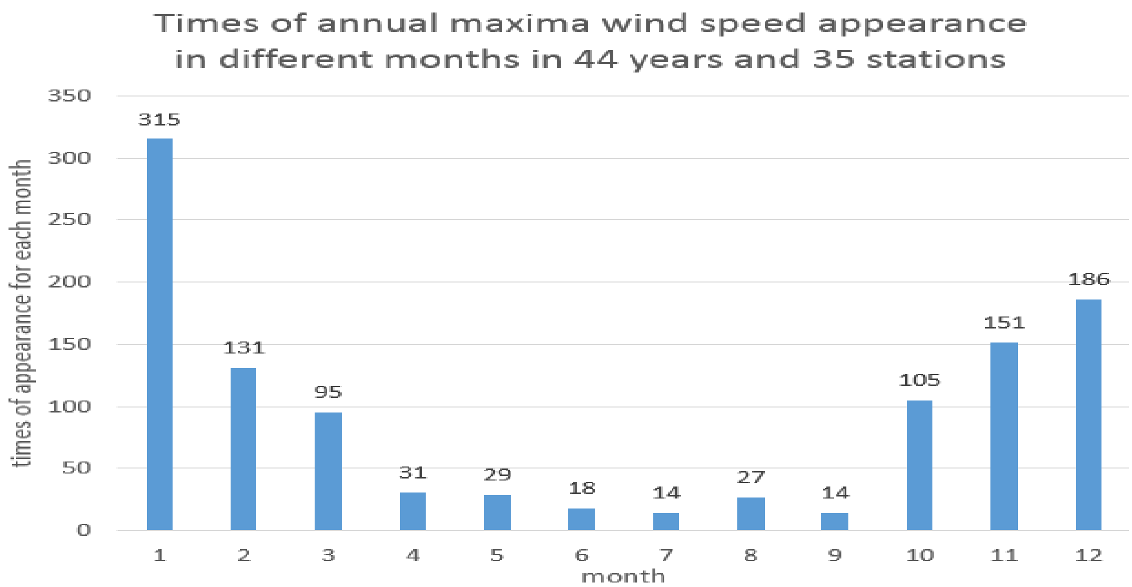


Figure 3.1 Times of annual maximum wind speed (km/h) appears in different months in 35 stations during 44 years.

From the results we observe that compared with other months, annual maximum wind speed have much higher probability to occur from January to March and from October to December, and lower probability to occur in the other six months. So we can simplify the problem of data selection and restrict the range of missing data from one year to these six months.

After simplifying the data quality questions, the next step is to find the number of missing data in one year which will not affect the quality of data and the results of the research. We call the number of missing data the threshold. If the number of missing data in one station in a specific year is larger than the threshold, the quality of wind speed data in these years and stations cannot be regarded as reliable. Therefore data in this year should not be used in the research. If the number of missing data is smaller than the threshold, the quality of data is reliable and the data will be used. We calculate the number of the missing data from January to March and October to December from 35 stations and 44 years. The results are shows in Figure 3.2.

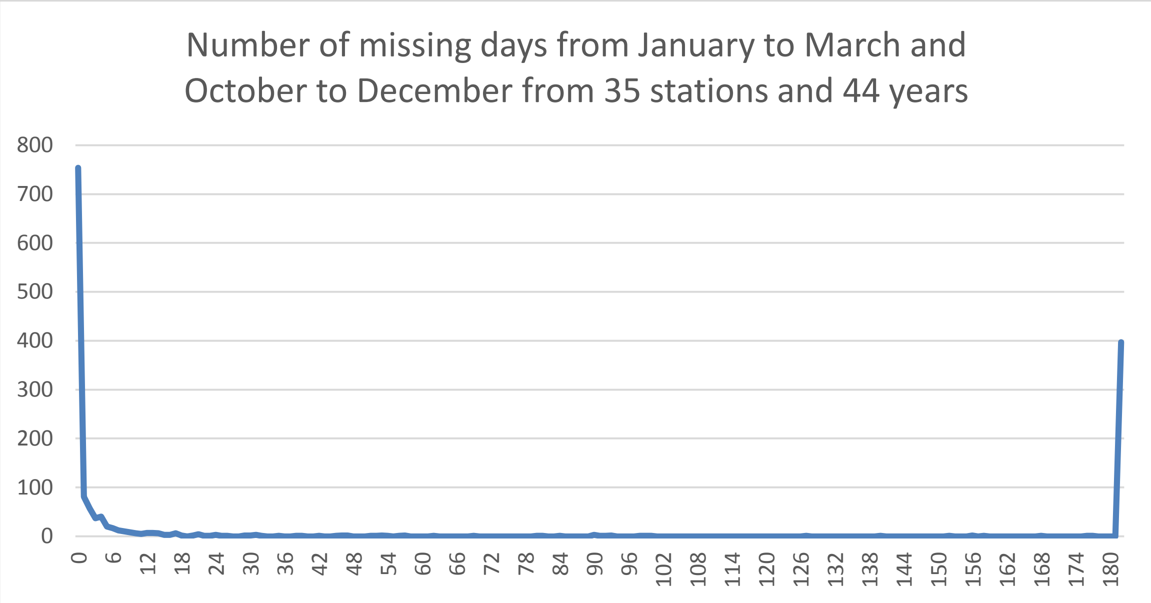


Figure 3.2 Number of missing days from January to March and October to December from 35 stations and 44 years. The starting points 0 and ending point 182 has the largest number of 702 and 397.

Missing days	0	1	2	3	4	5	6	7	8	9
Number	702	44	57	42	35	25	26	30	18	9

Missing days	10	11	12	13	14	15	16	17	18	19
Number	10	13	8	9	6	2	1	5	1	0

Missing days	20	21	22	23	24	25	26	27	28	29
Number	6	1	0	2	2	2	0	0	0	1

Table 3.1 Number of missing days from January to March and October to December from 35 stations and 44 years. This table shows numbers from 0 to 29 to give direct information on potential threshold.

Figure 3.2 shows that nearly 2/3 raw data in 44 years and 35 stations have no missing data, but also 1/3 data have missing data, whereas the number of missing data mostly fluctuate between 1 to 7 and 182. These numbers are not the appropriate choice for a potential threshold, because if a threshold between 1 and 7 is chosen, many data cannot be used. Also 182 missing data mean half-year data is missing. These will affect the results of the research. According to several studies, the missing data should not exceed 10% of the whole data (Köse, 2004). Therefore the range of potential thresholds is chosen between 8 and 18 (as there is no year with 19 missing data). It will affect only a few data since they account for only 4% and 10% of the whole days in 6 months. For those reasons, I consider 8—18 is an acceptable and reasonable range of threshold.

3.2. Model selection

Two steps are taken to decide the threshold of data selection. The first is to find the best trend surface model and the second is using this model to find the best fitting threshold. Focusing on the first step, theoretically, the marginal distribution of a max-stable processes is the generalized extreme value (GEV) distribution. It has three parameters, the location distribution μ, the scale parameter σ and the shape parameter ξ. For concrete applications, different places have different distributions. Therefore, pointwise marginal distribution should be allowed to occur locally. Defining a trend surface for the generalized extreme value parameters can solve this problems. The trend surface is the model which represents the marginal distribution of a max-stable process. Since wind speeds have connections with longitude, latitude and altitude, but there is no great change in the altitude of the Netherlands. Therefore longitude and latitude of meteorological stations are set as the parameters in the trend surface models. From the results of Ribatet (2013), it follows that there is a correlation between the distribution of maximum wind speed and the distance to the coast. Therefore, coast distance is set as a model variable. In this thesis, we assume several marginal distribution models with different forms (Table 3.2). Parameters $\mu(s)$, $\sigma(s)$, $\xi(s)$, $X(s)$, $Y(s)$ and *coast(s)* represent the location, scale and shape parameters of the generalized extreme value distribution, the geographic coordinate and the coast distance from the location $s \in S$, respectively.

M1 $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M2* $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0$	*M3* $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$

M4 $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0$	*M5* $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M6* $\mu = \beta\mu0 + \beta\mu1*X(s) + \beta\mu2*Y(s)$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0$

M7 $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M8* $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0$	*M9* $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$

M10 $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0$	*M11* $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M12* $\mu = \beta\mu0 + \beta\mu1*X(s)$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0$

M13 $\mu = \beta\mu0$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M14* $\mu = \beta\mu0$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s) + \beta\sigma2*Y(s)$ $\xi = \beta\xi0$	*M15* $\mu = \beta\mu0$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$

M16 $\mu = \beta\mu0$ $\sigma = \beta\sigma0 + \beta\sigma1*X(s)$ $\xi = \beta\xi0$	*M17* $\mu = \beta\mu0$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0 + \beta\xi1*X(s) + \beta\xi2*Y(s)$	*M18* $\mu = \beta\mu0$ $\sigma = \beta\sigma0$ $\xi = \beta\xi0$

M19 μ= $\beta\mu0$ +$\beta\mu1$*coast(s) σ= $\beta\sigma0$ + $\beta\sigma1$*coast(s) ξ=$\beta\xi0$ + $\beta\xi1$*coast(s)	*M20* μ= $\beta\mu0$ +$\beta\mu1$*coast(s) σ= $\beta\sigma0$ + $\beta\sigma1$*coast(s) ξ=$\beta\xi0$	*M21* μ= $\beta\mu0$ +$\beta\mu1$*coast(s) σ= $\beta\sigma0$ ξ=$\beta\xi0$ + $\beta\xi1$*coast(s)
M22 μ= $\beta\mu0$ +$\beta\mu1$*coast(s) σ= $\beta\sigma0$ ξ=$\beta\xi0$	*M23* μ= $\beta\mu0$ σ= $\beta\sigma0$ + $\beta\sigma1$*coast(s) ξ=$\beta\xi0$ + $\beta\xi1$*coast(s)	*M24* μ= $\beta\mu0$ σ= $\beta\sigma0$ + $\beta\sigma1$*coast(s) ξ=$\beta\xi0$
M25 μ= $\beta\mu0$ σ= $\beta\sigma0$ ξ= $\beta\xi0$ + $\beta\xi1$*coast(s)		

Table 3.2 Potential marginal distribution models with different variables and forms.

In the equation in Table 3.2, $\beta_{\mu0}$, $\beta_{\sigma0}$ and $\beta_{\xi0}$ are the constant parameters of the location distribution μ, the scale parameter σ and the shape parameter ξ , $\beta_{\mu1}$, $\beta_{\sigma1}$ and $\beta_{\xi1}$ are the coefficients of parameter X or coast(s) in coast distance model from M19 to M25, and $\beta_{\mu2}$, $\beta_{\sigma2}$ and $\beta_{\xi2}$ are coefficients of the parameters Y of the trend surface model.

As explained in chapter 2.4, Takeuchi's information criterion (TIC) is used to decide which model is fitting best. The results are shown in the Table 3.3.

Threshold 8 model selection results								
M1	M3	M4	M5	M2	M21	M19	M20	M6
8231.754	8234.629	8235.506	8236.844	8237.034	8237.525	8239.563	8242.032	8242.937
M22	M9	M11	M7	M10	M12	M8	M25	M23
8244.887	8346.354	8354.747	8354.826	8358.144	8367.256	8368.611	8393.521	8396.129
M17	M15	M13	M16	M24	M18	M14		
8398.761	8400.463	8401.767	8417.876	8420.142	8424.952	8428.736		

Threshold 9 model selection results								
M20	M2	M19	M1	M6	M7	M22	M4	M23
8481.106	8481.268	8483.277	8483.653	8485.038	8486.248	8486.634	8489.011	8490.679
M11	M24	M21	M17	M12	M25	M10	M18	M16
8490.764	8490.985	8491.157	8491.491	8491.797	8491.802	8492.369	8492.781	8493.453
M14	M9	M8	M15	M5	M13	M3		
8493.902	8493.93	8494.501	8500.123	8504.551	8514.971	8540.718		

Threshold 10 model selection results								
M1	M3	M21	M5	M4	M2	M19	M20	M6
8343.319	8345.73	8346.836	8347.433	8348.462	8348.561	8349.658	8350.714	8353.165
M22	M9	M7	M11	M10	M12	M8	M25	M23

8354.13	8464.986	8466.554	8467.352	8471.508	8479.007	8483.471	8506.373	8508.929
M15	M17	M13	M16	M24	M14	M18		
8511.296	8514.693	8518.028	8531.04	8532.19	8532.369	8537.365		

Threshold 11 model selection results								
M21	M19	M20	M3	M1	M4	M5	M22	M2
8298.889	8301.353	8301.846	8303.082	8303.457	8305.095	8306.606	8307.27	8309.439
M6	M9	M7	M11	M10	M12	M8	M25	M23
8313.46	8402.875	8403.224	8405.094	8415.651	8422.364	8424.375	8445.018	8447.65
M17	M13	M16	M24	M18	M14	M17		
8451.27	8456.234	8470.402	8471.291	8476.802	8480.823	8451.27		

Threshold 12 model selection results								
M1	M3	M4	M5	M19	M2	M21	M20	M6
8442.545	8444.268	8445.248	8446.773	8446.804	8447.79	8448.177	8448.717	8451.951
M22	M9	M7	M11	M10	M12	M8	M25	M23
8452.631	8560.874	8563.106	8563.602	8571.465	8580.02	8584.544	8608.346	8610.73
M17	M15	M13	M16	M24	M14	M18		
8616.881	8617.676	8619.414	8633.087	8633.574	8633.995	8639.384		

Threshold 13 model selection results								
M1	M3	M4	M5	M2	M19	M21	M20	M6
8495.541	8497.622	8497.824	8499.468	8500.285	8501.173	8501.383	8501.867	8504.3
M22	M9	M7	M11	M10	M12	M8	M25	M23
8505.83	8614.505	8622.533	8622.976	8626.002	8634.458	8638.885	8662.815	8665.208
M15	M13	M17	M16	M24	M14	M18		
8668.115	8671.561	8671.675	8687.732	8687.938	8688.418	8693.934		

Threshold 14 model selection results								
M1	M3	M4	M5	M21	M2	M19	M20	M6
8540.593	8541.984	8542.674	8544.191	8544.78	8545.203	8545.242	8546.727	8549.155
M22	M9	M7	M11	M10	M12	M8	M25	M23
8550.376	8662.981	8669.259	8669.605	8672.873	8681.219	8685.695	8711.144	8715.341
M15	M17	M13	M16	M24	M18	M14		
8718.368	8718.843	8721.667	8735.821	8736.29	8742.101	8748.546		

Threshold 15 model selection results								
M1	M3	M4	M5	M2	M21	M19	M20	M6
8565.275	8567.627	8567.683	8568.506	8570.163	8570.35	8571.042	8571.483	8574.154
M22	M9	M7	M11	M10	M12	M8	M25	M23
8575.02	8687.061	8692.59	8693.129	8696.967	8705.194	8710.134	8735.091	8741.522
M17	M13	M15	M16	M24	M14	M18		
8743.192	8743.781	8743.984	8760.156	8760.628	8760.971	8766.377		

Threshold 16 model selection results								
M1	M4	M3	M5	M19	M21	M2	M20	M6
8587.937	8589.988	8590.081	8591.48	8592.372	8592.603	8592.715	8593.83	8596.508
M22	M9	M7	M11	M10	M12	M8	M25	M23
8597.21	8707.511	8710.396	8712.68	8719.89	8727.939	8732.899	8758.901	8763.059
M17	M15	M13	M16	M24	M18	M14		
8766.844	8767.631	8770.632	8784.033	8784.543	8790.161	8796.578		

Threshold 17 model selection results								
M1	M4	M3	M5	M2	M21	M19	M20	M6
8629.824	8632.064	8632.074	8633.007	8634.469	8634.845	8636.081	8636.212	8638.512
M22	M9	M7	M11	M10	M12	M8	M25	M23
8639.705	8753.808	8759.23	8759.644	8763.541	8771.742	8773.28	8802.443	8805.033
M17	M15	M13	M16	M24	M18	M14		
8810.7	8811.384	8812.199	8827.847	8828.34	8834.049	8840.206		

Threshold 18 model selection results								
M1	M4	M3	M21	M5	M19	M2	M20	M6
8643.919	8646.182	8646.425	8646.853	8647.013	8648.224	8648.513	8650.264	8652.443
M22	M9	M11	M7	M10	M12	M8	M25	M23
8653.622	8767.534	8769.965	8773.56	8779.137	8786.009	8787.184	8817.088	8819.597
M15	M17	M13	M16	M24	M18	M14		
8820.853	8825.307	8826.716	8842.636	8843.164	8848.702	8854.531		

Table 3.3 Results of Takeuchi's information criterion (TIC) in model selection. Models are sorted by score of Takeuchi's information criterion (TIC).

The results show that according to the results of Takeuchi's information criterion (TIC), in 11 tests, 9 tests among threshold 10 and 12--18 have the same results that M1 has the lowest TIC score and is the best fitting model. For the other two tests, thresholds 9 and 11 have different outcomes. So it can be concluded that M1 is the best fitted model in most cases. The format of M1 is presented in Table 3.2.

3.3. Threshold selection

After identifying a model that for most cases fitting well, the next step is to find the optimal threshold. In this step, a model sensitivity analysis, the test on the influence of model results on different inputs, is used to decide upon the optimal threshold. The range of thresholds has been discussed earlier in this thesis. We selected the same range from 8 to 18 as before. The results have presented in **错误!未找到引用源。**.

threshold	loc1	loc2	loc3	scale1	scale2	scale3	shape1	shape2	shape3
8	76.42618	-0.08108	0.06714	14.02343	-0.002566	-0.004724	-0.281954	-0.000971	0.000783
9	85.06437	-0.01907	0.02656	11.02118	-0.010086	0.007287	-2.741e-1	-3.476e-5	3.892e-04
10	76.57774	-0.08136	0.06712	13.77651	-0.002763	-0.004165	-0.310351	-0.000891	-0.000891
11	77.31800	-0.07627	0.06370	13.66727	-0.007641	-0.002041	-0.304113	-0.000749	0.000757
12	76.49291	-0.08201	0.06760	13.70476	-0.003357	-0.003766	-0.313992	-0.000856	0.000820
13	76.22114	-0.08214	0.06820	13.50392	-0.003488	-0.003334	-0.297530	-0.000849	0.000781

14	76.28477	-0.08240	0.06822	13.55185	-0.003217	-0.003535	-0.300906	-0.000846	0.000784
15	76.41206	-0.08195	0.06788	13.61339	-0.002627	-0.003845	-0.300801	-0.000872	0.000789
16	76.47423	-0.08309	-0.08309	13.62017	-0.08309	-0.003715	-0.003715	-0.000819	0.000820
17	76.40538	-0.08256	0.06812	13.58352	-0.003216	-0.003650	-0.269505	-0.000842	0.000712
18	76.42618	-0.08333	0.06835	13.57343	-0.003725	-0.003475	-0.305263	-0.000800	0.000774

Table 3.4 Results of model sensitivity test on parameters of marginal distribution model M1 with different threshold.

错误!未找到引用源。 shows that for different thresholds, small changes happen to the model parameters. In general, this means that the model selection and parameter estimation were not very sensitive to threshold selection. So, basically, any threshold between 8 and 18, except 9, can be chosen, because according to the results of sensitivity analysis, no matter which threshold is used, the results of the research turned out to be the same. In this research, the threshold was somewhat arbitratily set to 14, in the middle of the explored threshold value and corresponding to the time span of two weeks.

3.4. Results of marginal distribution and spatial dependence model

Considering the trend surface model, no spatial dependence was taken into consideration. Thus, the next step is to find the most appropriate spatial dependence model. In the section 2.1.2, several spatial dependence models were introduced. Since these models are not nested, Takeuchi's information criterion (TIC) can be used to decide which is the best fitted model. The results are presented in **错误!未找到引用源。**,

Model	Extremal-t	Schlather	Brown-Resnick
Results	213169.7	213358.9	213427.8

Table 3.5 Results of spatial dependence model selection according to the Takeuchi's information criterion (TIC) method. Models are sorted by scores of Takeuchi's information criterion (TIC).

错误!未找到引用源。 shows that TIC values are not quite similar, but the extremal-t model has a slightly smaller TIC score, therefore it is the best fitting spatial dependence model. Also, the best fitting trend surface of the extremal-t model has been obtained, and is presented in equation (2.17).

$$\begin{aligned} \mu &= 73.86_{(1.9)} - 0.083_{(0.004)} * X(s) + 0.072_{(0.003)} * Y(s), \\ \sigma &= 17.17_{(1.2)} - 0.00058_{(0.003)} * X(s) - 0.0129_{(0.002)} * Y(s), \\ \xi &= -0.34_{(0.09)} - 0.0011_{(0.0002)} * X(s) + 0.001_{(0.0002)} * Y(s). \end{aligned} \tag{2.17}$$

In this model, the subscripts are the associated standard errors. From the results of model fitting we can see that the marginal distribution of extreme wind speed in the Netherlands has a connection with the X and Y coordinate of the location, because location parameter μ has a negative and positive correlation with X and Y, respectively. For a concrete application it means that the value of parameter μ will increase from east to west and from south to north; The scale parameter σ, has a negative correlation with the parameter X and has no connection with parameter Y, so the value of parameter σ increaseS from east to west. The shape parameter ξ , also, has a negative correlation with parameter X and positive correlation with parameter Y and it increases from east to west and from south to north. From the model we further see that the parameter ξ has a relative small coefficient, therefore the longitude and latitude have little influence on the shape parameter of marginal distributions. From the above analysis we can see that the distribution of extreme wind speed have a correlation with the coast distance, the largest wind gusts are observed along the coastline.

3.5. Comparison and prediction

After achieving the best max-stable model for extreme wind speed in the Netherlands, the summary of characteristics and the prediction of regional extreme events was studied with the best fitting model. For regional extreme events, the characteristics are the spatial dependence and marginal distribution of certain locations. In particular, we consider three aspects.

The first aspect is the comparison of the return level for return period of 25-year, 50- year, 75-year and 100-year in 35 stations for the generalized extreme value (GEV) distribution in a single location and the marginal distribution in spatial models. The fitting methods to the generalized extreme value (GEV) parameters were the calculation of the maximum likelihood function by means of the log-likelihood function (2.18).

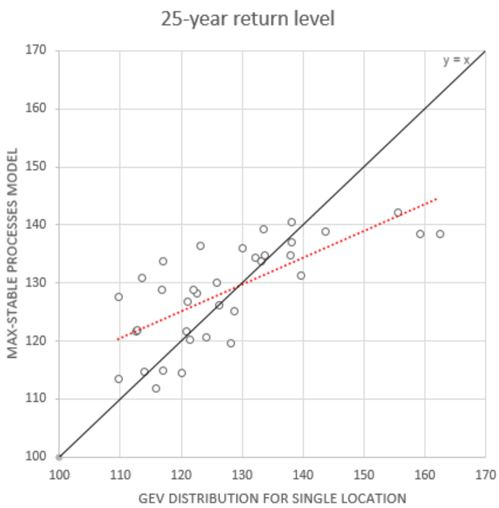
$$L(\theta) = L(\mu,\sigma,\gamma) = -nln\sigma - \sum_{i=1}^{n}\left(1+\gamma\left(\frac{x_i-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}} - (1+\frac{1}{\gamma})\sum_{i=1}^{n} ln[1+\gamma(\frac{x_i-\mu}{\sigma})] \quad (2.18)$$

in where formula (2.18), $\theta = (\mu,\sigma,\gamma)$. The $\hat{\mu},\hat{\sigma},\hat{\gamma}$ are the maximum likelihood estimators of the parameter vector (2.18) with which the function obtains the maximum likelihood value. The results are shown in Table 3.6 and Figure 3.3.

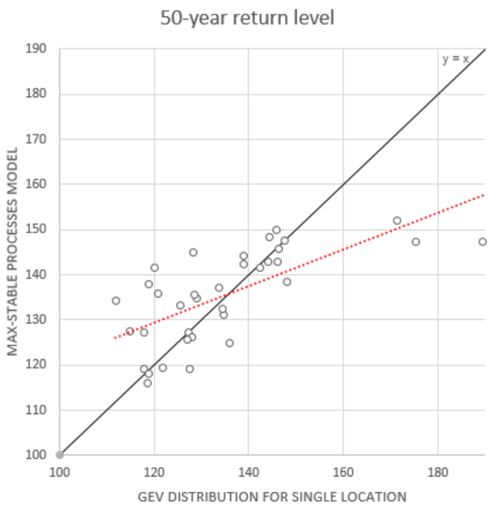
No.	Station	Return level for 25 years		Return level for 50 years		Return level for 75 years		Return level for 100 years	
		Single GEV	Model	Single GEV	Model	Single GEV	Model	Single GEV	Model
210	Valkenburg	138.20	137.10	146.29	145.71	151.01	150.78	154.35	154.41
225	Ijmuiden	162.43	138.35	175.26	147.26	183.17	152.58	188.96	156.41
235	De kooy	138.16	140.57	145.77	150.03	150.25	155.78	153.45	159.98
240	Schiphol	137.94	134.75	146.13	142.85	150.91	147.60	154.29	150.97
242	Vlieland	155.64	142.05	171.25	151.89	181.77	157.96	189.93	162.43
249	Berkhout	130.03	135.94	139.01	144.31	144.57	149.28	148.67	152.84
251	Hoorn	133.60	139.25	144.36	148.39	151.59	153.98	157.19	158.07
260	De Bilt	116.77	128.82	120.79	135.71	122.90	139.63	124.29	142.36
265	Soesterberg	122.66	128.14	129.01	134.90	132.63	138.74	135.16	141.42
267	Stavoren	132.17	134.28	138.81	142.29	142.68	147.03	145.42	150.43
269	Lelystad	122.07	128.80	128.52	135.69	132.32	139.64	135.03	142.41
270	Leeuwarden	133.09	133.72	142.37	141.59	148.07	146.28	152.23	149.65
273	Marknesse	109.63	127.63	111.81	134.29	112.90	138.11	113.61	140.79
275	Deelen	120.86	121.65	127.32	127.27	131.00	130.37	133.58	132.49
277	Lauwersoog	139.60	131.22	148.06	138.55	153.11	142.90	156.75	146.01
278	Heino	112.60	121.70	117.98	127.32	120.98	130.44	123.04	132.59
279	Hoogeveen	112.81	121.76	114.81	127.38	115.75	130.52	116.33	132.69
280	Eelde	128.79	125.08	134.72	131.25	138.02	134.78	140.30	137.25
283	Hupsel	117.08	114.86	121.87	119.45	124.49	121.91	126.26	123.55
286	Nieuw Beerta	124.10	120.72	128.08	126.14	130.19	129.19	131.61	131.30
290	Twenthe	120.10	114.53	127.59	119.08	131.99	121.51	135.13	123.14
310	Vlissingen	143.66	138.76	147.55	147.69	149.51	152.92	150.76	156.63
319	Westdorpe	116.96	133.74	120.18	141.56	121.82	146.03	122.89	149.15
323	Wilhelminadorp	123.21	136.49	128.17	144.92	130.95	149.82	132.87	153.28
330	Hoek van Holland	159.17	138.53	189.53	147.45	213.33	152.73	233.67	156.50

340	Woensdrecht	113.49	130.80	118.76	138.05	121.73	142.16	123.78	145.01
344	Rotterdam	133.63	134.80	144.08	142.89	150.54	147.60	155.29	150.93
348	Cabauw	125.88	130.06	133.83	137.18	138.57	141.25	141.98	144.09
350	Gilze-rijen	126.24	126.11	134.42	132.49	139.19	136.04	142.58	138.49
356	Herwijnen	121.00	126.81	125.62	133.32	128.11	136.98	129.79	139.51
370	Eindhoven	121.42	120.29	127.04	125.70	130.13	128.62	132.24	130.60
375	Volkel	128.05	119.54	135.95	124.83	140.45	127.70	143.60	129.64
377	Ell	114.03	114.65	117.84	119.24	119.87	121.64	121.23	123.24
380	Maastricht	115.78	111.90	118.70	116.12	120.20	118.30	121.18	119.72
391	Arcen	109.68	113.55	118.79	117.98	124.77	120.31	135.16	121.85

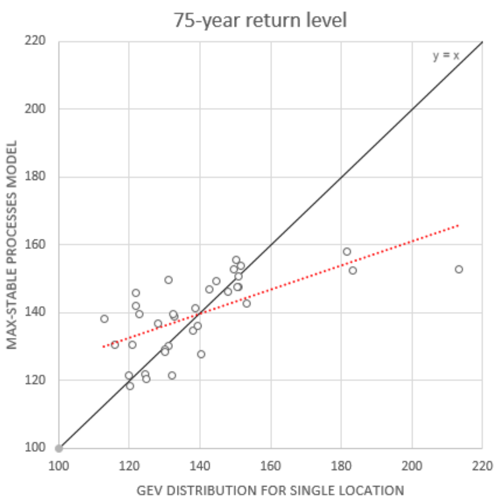
Table 3.6 Comparison of predicted return level in 25-year, 50-year, 75-year and 100-year return period in each stations by generalized extreme value (GEV) distribution and max-stable processes model.



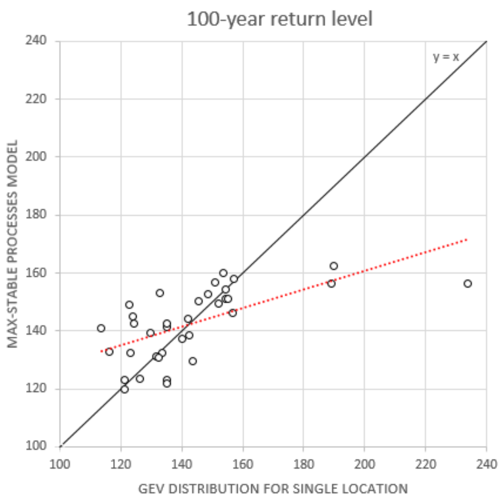
a. 25-year return period



b. 50-year return period



c. 75-year return period

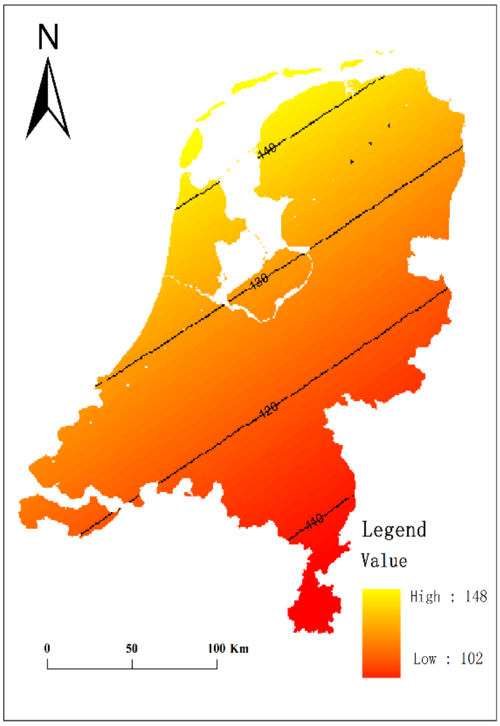


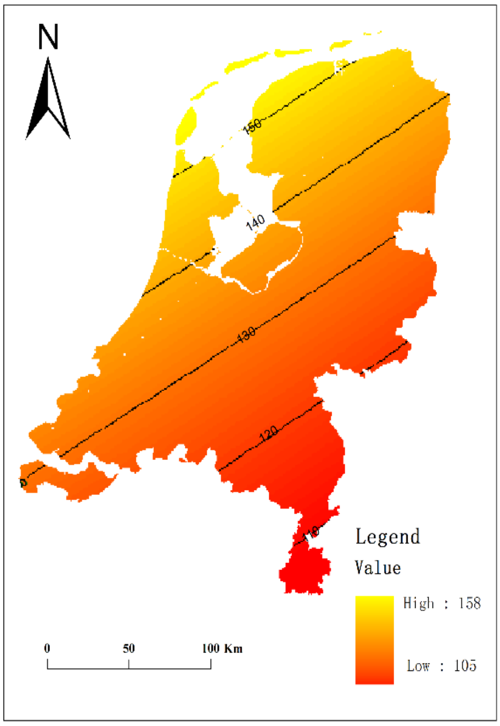
d. 100-year return period

Figure 3.3 Comparison of predicted return level in each stations by generalized extreme value (GEV) distribution and max-stable processes model. The black line in figure is y=x and the red line is the fitted line by least square method.

From Table 3.6 we see that, for the generalized extreme value (GEV) distribution model in a single location, the predicted values of the return level have the highest degree of dispersion. Small values are much smaller and large values are much larger than the max-stable processes model return level. This can also be seen from Figure 3.3. In Figure 3.3, the x-axis is the value of return level obtained from the generalized extreme value (GEV) distribution in single locations, the y-axis is the value of return level obtained from the max-stable processes model, thus providing the scatter diagram. There are two lines in the figure: the red dotted line is the least square line of predicted return level, and the black line is the line $y=x$. From the two lines we see that the slope of red line is smaller than that of the black line. This means that compared with max-stable processes model, the return level obtained from the generalized extreme value (GEV) distribution model in single locations has higher values in stations which have a relative large annual maximum wind speed, but are underestimated in stations which have relatively small annual maximum wind speeds. This difference can be regarded as the result of modifications from max-stable process models to traditional generalized extreme value (GEV) distribution model. This can also explained by max-stable processes model considers the characteristics of spatial dependence. For further analysis, parameters in the generalized extreme value (GEV) distribution model are flexible in the sense that they are all fitted by data in single locations. So there are no intrinsic links between locations. Max-stable process models can be separated into two components, spatial dependence and marginal distribution. For the spatial dependence, the model represents spatial correlation between different locations. So all data from different stations should be used in the model at the same time, and the results will be influenced by all stations. For the marginal distribution, since method of trend surface model has been used, the variable of trend surface in this thesis is projection coordinate X and Y. Since research areas are not far away and they have similar geographic features, the results of marginal distribution models will be similar, and the results of max-stable process models will be smoother and have less degree of dispersion than the generalized extreme value (GEV) distribution.

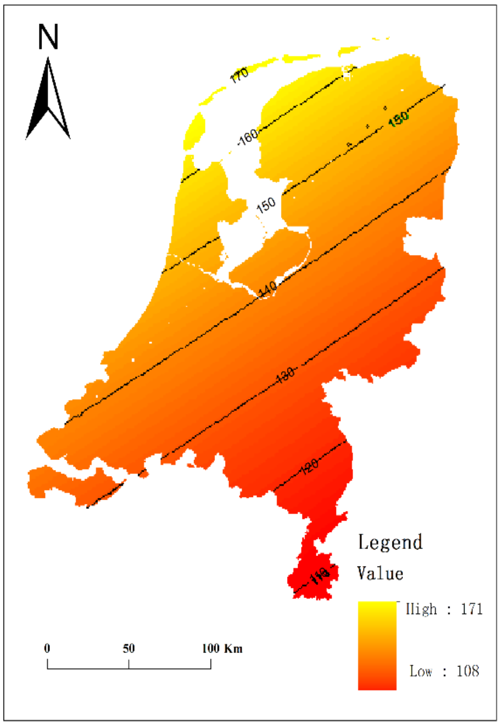
The second aspect is the prediction of maximum wind speed in individual locations. Regional extreme wind speed events analysis can predict wind speed in locations where there are no meteorological stations, and this is why for such analyse, spatial extreme models are superior to multivariate extreme models, as the objects of max-stable processes modelling is an area, but not individual meteorological stations. The method of prediction is specific for the whole Netherlands to grid nodes and calculates the value of return levels for each node using the model that we obtained. Figure 3.4 is the map of return level in 25-year, 50-year, 100-year and 200-year return period in the Netherlands based on the fitted extremal-t model in the Netherlands. From the figure we see that the red parts in the south-east have relatively low return levels, and the yellow parts in the north-west have much higher return levels, and this distribution is familiar with the coast distance. As expected, the largest wind speeds have a connection with the coastline, but the maps indicate that the correlations are not strong enough. This conclusion is the same for model selection.

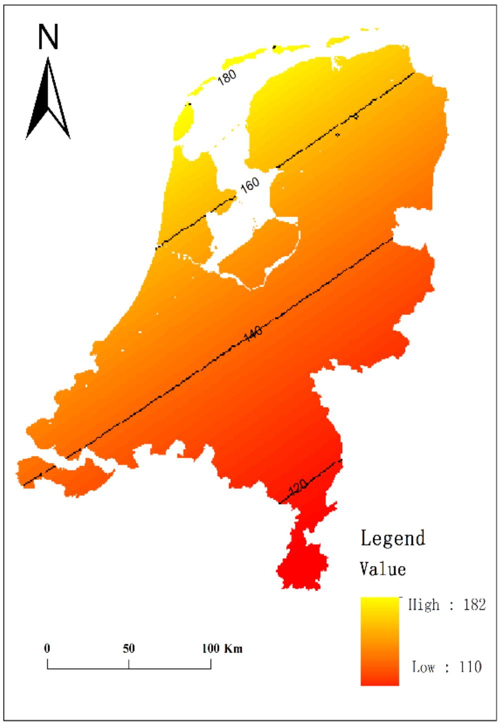




a. Return level in 25-year return period

b. Return level in 50-year return period



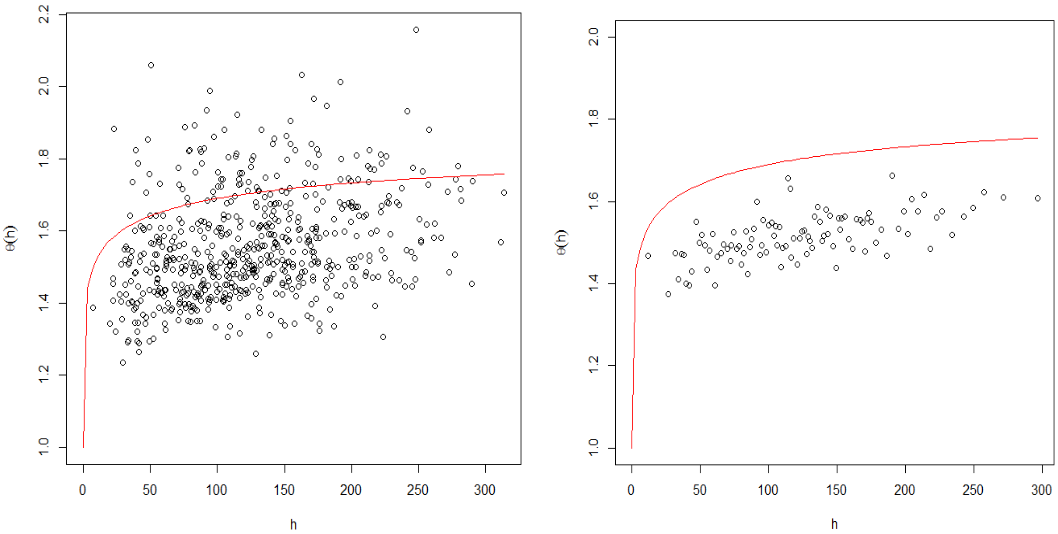


c. Return level in 100-year return period

d. Return level in 200-year return period

Figure 3.4 Map of return level in 25-year, 50-year, 100-year and 200-year return period in the Netherlands.

The third aspect is the spatial dependence of models. The extremal coefficient functions are a convenient way to assess whether the model can reflect the spatial correlation structure in regional areas well or not. From Figure 3.5 we see that with the longer distance, the spatial dependence weakens. The fitted curve indicates that if the distance of two stations equals 50km, then the extremal coefficient is slightly more than 1.6, and if the distance becomes larger than 250km, the spatial dependence of two stations has a weak correlation with the extremal coefficient is around 1.8.



a. Pairwise estimates; b. Binned estimates (with 100 bins)

Figure 3.5 Comparison between fitted extreme coefficient function and empirical F-madogram cloud.

4. DISCUSSION

The aim of building spatial extreme value models is to analyse statistical regularities of regional extreme events. Due to the characteristics of extreme events, classical statistical methods in this thesis fall short when aiming to solve such problems in practical applications. Therefore, spatial extreme models based on max-stable processes are proposed and developed recently. Max-stable processes, which focus on the property of regionality and on randomness of spatial extreme events, can separate the statistical properties of extreme events into two components, spatial dependence, i.e., the spatial relationship of different locations, and marginal distribution, i.e., the statistical regularities of extreme events at single locations.

In this thesis, annual maximum wind speed data from 35 meteorological stations measured between 1971 and 2014 were used to model the distributions and spatial dependences of extreme wind speed events in the Netherlands. The specific methods for modelling was to parameterize the distribution of max-stable processes, this was done by using the composite likelihood method. After fitting the model, predictions can be made via the two dimensional joint distribution function. Even though the composite likelihood method has disadvantages in terms of the efficiency, it can fit the spatial extreme model with all stations' data at the same time and obtain an asymptotic and unbiased estimators of the model parameters and asymptotic normality of the estimations (Ribatet, 2013). According to the method of model fitting and selection methods described in chapter 3.2 and 3.3, models of marginal distribution and spatial dependence can be obtained. The model of the marginal distribution is presented in formula (2.17), whereas the model of spatial dependence is the extremal-t model. Then, several spatial analyses and comparisons between spatial models and generalized extreme value (GEV) distribution were carried out to find regularities in the distributions and spatial dependences of extreme wind speed events in the Netherlands. Comparing the results with other similar studies (Ribatet, 2013), we see that both researches obtain similar results, even though we have different geographic and projection coordinates, data selection methods and model variables. From the results of the analysis, we notice that the maximum wind speed increases from south to north and from east to west, whereas the maximum wind speed have a negative correlation with the distance to the coastline. To some degree we can say that a higher distance from the coastline to the station corresponds with a lower annual wind speed.

Since De Haan discovered the spectral representation of max-stable processes (de Haan, 1984), statistical modelling of spatial extreme events has been studied widely and much progress has been achieved. There are, however, some problems and difficulties in the both theoretical and application, which need further discussion and improvement.

1 Interpretation of the missing data
One aim of the thesis was to model extreme wind speed events in the Netherlands, understand their characteristics and correlation and predict the occurrence regularities and probabilities. The characteristics of annual maximum wind speed must be determined by using at least several years' maximum wind speed data. Therefore completeness should be taken into consideration. In the thesis, nearly half of the meteorological stations has missing data. So it is really important to minimize the effect of missing data on modelling as a spatial data quality issue.

From past studies we obtained that missing data should not exceed 10% according to standards (Köse, 2004). So in this thesis, 10% is set as a threshold, meaning that the number of missing data in a single year should be less than 19 (10.4%). There are two reasons that 1 to 7 days cannot be chosen as threshold. Firstly, for a whole year data, the number of missing days from 1 to 7 days is relatively small, so the probability that the annual maximum wind speed appears in these days is low; secondly, from table 3.1 we see that the number of missing days from 1 to 7 is really large. If one of which value is chosen, large amounts of data meeting this conditions will be deleted, this will affect the results of the research.

There are different factors that affect the precision during data selection. Data selection is really important in this thesis. If the data are selected too loosely, the results of the research will not be the regularity of extreme wind speed but extreme as well as normal wind speed. However, if the data are selected too tightly, the results of the research will not show characteristics of extreme wind speed completely. So in the future, efforts should be made to propose better and more precise methods of data selection.

2 Usage of other variables in the statistical models
The model for extreme wind speed can be complicated because of several variations from different sources, but can as well as simple, just constant with no variation. Some variations depend on the geographical feature of the research area, like hills, valleys, river bluffs, buildings and trees. In this thesis, we induced three variables in the model, the projection coordinates X and Y and the distance from coast to location s, *coast(s)*, but other geographical features are not considered in the model. There are no remarkable hills, valleys and river bluffs in the Netherlands. Buildings and trees, however, are shown everywhere around the country, which will lead to irregular surfaces and produce more friction and turbulence. The higher the friction, the lower the wind speed is. But for now, inclusion of data on building and trees which can affect wind speed and how to affect it are hard to define. Therefore, efforts need to be done to improve the precision of spatial models in future studies.

Since the weather model can be applied widely, the method of modelling and analysis used in this thesis can be extended towards other places. One of the difficulties in the research is to fit the max-stable processes. The maximum composite likelihood estimator implies a loss in efficiency and shows typically numerical instabilities (Ribatet, 2013). Therefore, better inferential procedures should be studied for future research and analysis. Also we should do efforts to improve the quality of the raw data, the determination of model variables and the criteria of model selection, like we can choose those years which have no or few missing data to do the research. This will improve the quality of the data. Also, since the wind speed has obvious relationships with geographical features of the research area, so it is preferable to do study on the topographic and geomorphic conditions of the research area. This will helpful to improve the accuracy of the model.

5. CONCLUSION

The conclusions of the thesis are reported in four parts, data selection, model selection split into marginal distribution and spatial dependences, and the spatial analysis of annual maximum wind speeds events in the Netherlands.

1 Data selection:
Since the occurrence of missing data is a common phenomenon in meteorological studies, this topic was addressed in this thesis. Methods should be applied to maintain the quality of data. In the thesis, data from January to March and October to December are chosen during the data selection since these six months have much higher probabilities to be the month which annual maximum wind speed occurs. Meanwhile, 10% of the whole data was set as the threshold of missing data. If the number of missing data is smaller than the threshold, it will be regarded as fine data and if the number of missing data is larger than the threshold, the quality of data does not meet the standard and data from this year will be given up. Model sensitivity tests are used to get the best fitted model after data selection. The results show that there are almost no differences between 10 and 18 (10% data), so we rather arbitrarily choose the value 14 as the threshold of missing data.

2 Marginal distribution model:
Composite likelihood methods was used to fit the marginal distribution of spatial extreme models. The method use a parameters of the marginal distribution and geographic coordinates to build regression models, the maximizing composite likelihood estimation has asymptotic unbiasedness and normality. Based upon the composite likelihood methods, the best marginal distribution model for extreme wind speed in the Netherlands is obtained and shown in formula (2.17). The model shows that the extreme wind speed in the Netherlands has a connection with the longitude and latitude of the stations. Also, to some degree, it has a connection with the coast distance of the meteorological stations. The results show that a short distance from the coastline corresponds with a large extreme wind speed for the locations have.

3 Spatial dependence model:
The spatial dependence is well explained by the extreme coefficient. In this thesis, the pairwise extreme coefficient in two different stations was estimated as it explains partly the interconnections and influences of extreme wind speed events in the Netherlands. With the help of Takeuchi's information criterion (TIC), the extremal-t model is chosen as the best model among the three models. Figure 3.5 shows that stations have a correlation with each other, and that the correlation reduced with the increasing distance.

4 Spatial analysis:
Finally, the thesis predicted the return level for different return periods with max-stable processes models. It also made a comparison between the prediction results from generalized extreme value (GEV) distribution and from marginal extreme distribution fitted by max-stable processes. The results show that the max-stable processes model was smoother and had a lower dispersion than the generalized extreme value (GEV) distribution. This is partly because the parameters of generalized extreme value (GEV) distribution are flexible, whereas parameters of the max-stable process model are defined by trend surface with coordinate X and Y. The return level map also shows that the largest wind speeds were connected with the distance to the coastline, but the maps show that this correlation is not very strong.

LIST OF REFERENCES

Bortot, P., & Gaetan, C. (2013). Multivariate Extremes, 109–140.

Brown, B. M., & Resnick, S. I. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability.14*(4), 732-739.

Butler, A., Heffernan, J. E., Tawn, J. A., & Flather, R. A. (2007). Trend estimation in extremes of synthetic North Sea surges. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *56*(4), 395–414.

Casson, E., & Coles, S. (1999). Spatial Regression Models for Extremes. *Extremes*, *1*(4), 449–468.

Centraal Bureau voor de Statistiek. (2015). Windenergie op land; productie en capaciteit per provincie. Retrieved January 14, 2016, from http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=70960NED&D1=0-1,6,9&D2=a&D3=22-23&VW=T

Chen, J., Lei, X., Zhang, L., & Peng, B. (2015). Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China. *PLoS ONE*, *10*(2), 1–10.

Coles, S., & Casson, E. (1998). Extreme value modelling of hurricane wind speeds. *Structural Safety*, *20*(3), 283–296.

D.L. Massart, Smeyers-Verbeke, J., Caprona, X., & Schlesier, K. (2005). Visual Presentation of Data by Means of Box Plots. *LC GC Europe*, *18*(4), 215–218.

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, *27*(2), 161–186.

De Haan, L. (1984). A spectral representation for max- stable processes. *The Annals of Probability*, *Volume 12*(Number 4), 1194–1204.

Faranda, D., Lucarini, V., Turchetti, G., & Vaienti, S. (2011). Numerical Convergence of the Block-Maxima Approach to the Generalized Extreme Value Distribution. *Journal of Statistical Physics*, *145*(5), 1156–1180.

García-cueto, O. R., & Santillán-soto, N. (2007). Modeling Extreme Climate Events: Two Case Studies in Mexico. *Statistical Methodology*, (i).

De Haan, L., & Ferreira, A. (2006). Extreme Value Theory : An Introduction.

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546.

Kabluchko, Z., Schlather, M., & De Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, *37*(5), 2042–2065.

Köse, R. (2004). An evaluation of wind energy potential as a power generation source in Kütahya, Turkey. *Energy Conversion and Management*, *45*(11-12), 1631–1641.

Lee, Y., Yoon, S., Murshed, M. S., Kim, M.-K., Cho, C., Baek, H.-J., & Park, J.-S. (2013). Spatial modeling of the highest daily maximum temperature in Korea via max-stable processes. *Advances in Atmospheric Sciences*, *30*(6), 1608–1620.

Opitz, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, *122*, 409–413.

Padoan, S. A, Ribatet, M., & Sisson, S. A. (2009). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, *105*(489), 263–277.

Palutikof, J., Brabson, B., Lister, D., & Adcock, S. (1999). A review of methods to calculate extreme wind speeds. *Meteorological*, *6*, 119–132.

Ribatet, M. (2013). Spatial extremes: Max-stable processes at work. *Journal de La Société Française de Statistique*, *154*(2), 156–177.

Ribatet, M., & Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de La Société Française de* , *154*(1), 138–150.

Sang, H., & Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, *16*, 407–426.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, *5*(1), 33–44.

Schlather, M., & Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, *90*(1), 139–156.

Shaby, M. J. R. B. A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *Annual Applied Statistics*, *29*(6), 997–1003.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Unpublished Manuscript, University of Surrey*, 1–32.

Varin, C., Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Advances in Applied Probability*, *443*, 461–483.