Newcastle University

School of Mathematics and Statistics

MAS8391 - MMathStat Project

Modelling Environmental Extremes

Student:
Michael French

Supervisor:
Dr. David Walshaw

May 1, 2015

Abstract

In this project, we will investigate and model the environmental extremes of wind speed and temperature in the area of the nuclear power station, Sellafield. In particular, we will look at wind speeds predominantly from the angle of 118 degrees to north, as this is the wind speed direction that Sellafield Ltd are most interested in with regards to the performance of a cooling system. Also, we will look at prolonged conditions of these variables, these types of conditions are of are most concern as extended periods have a more severe negative impact on the cooling system. We will assess the univariate and bivariate extremes of these two variables, from this we will find parameter estimates for Generalised Pareto Distributions and the logistic model. From this, we will obtain return levels and provide scenarios in which we predict may occur in various timeframes assuming that the variables follow the same meteorological patterns of the time of the dataset.

Contents

1 Introduction

1.1 Background Information

At Sellafield Ltd, they are interested in the performance of one of their building's cooling system, the system is based on natural convection due to the heat generator within, the primary characteristic to the cooling system is that possesses no fans. This buildings system performance(P) depends on both temperature(T) of the inlet air and its speed and direction. The impact of the parameter, external temperature, to the internal temperature is directly proportional. So, an increase in external temperature of 5 degrees celsius will increase the internal temperature by 5 degrees celsius.

$$P \propto -T$$

The impact the external wind parameter(W) has on the buildings system isn't as clear cut as temperature, certain wind directions namely, 118^o to north inhibit the air flow the most, this is understood to be due to the pressure differentials between the inlet and the outlet which results from wind flowing over the outlet stacks. Thus, the internal environment heats up as the wind prevents the escape of the air. It has been observed that wind speeds under 3m/s have very little effect on the system at all, however, from this point there is an increasing effect on the system, it is believed that there is a speed squared dependence.

$$P \propto -W^2$$

1.2 The Aim

There have been studies that suggest that the heat capacities of the internal components take up to 4 hours for the store internals to fully respond to a change in external conditions, however, around 80-90% of the effect manifests after only one hour. So, what the Sellafield Ltd company are most concerned with is weather conditions that persist for 1 hour or more. What I have been asked, is to present a 1 in 100 and 1 in 10000 years probability of the extremes of worst case temperature, wind direction and wind speed. Also, what Sellafield Ltd Company are most interested in is the combination of these factors and their net impact on the system. So what we will do in order to account for these conditions, we will perform our analysis on the one hour and two hour moving averages of the data in order to find prolonged weather conditions that may be of concern to the performance of the cooling system. We will model these environmental extremes and supply the researchers at Sellafield Ltd with extreme scenarios with their apprioriate probability of occurence.

1.3 Modelling Environmental Extremes

So, what we need to model this behaviour and evaluate these probabilities, we need a model that is tailored to measuring extreme values, there are two types of model that do this, Classical models and Threshold Models.

1.3.1 Classical Models[9][5]

Suppose you have $X_1, X_2, ...,$ as an independent and identically distributed sequence of random variables, then define

$$M_n = max\{X_1, ..., X_n\},$$

From this you need the limiting distribution of M_n as $n \to \infty$, however, this is *degenerate* and we need to work with a normalised version.

1.3.2 The Extremal Types Theorem [10]

If there exist sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$Pr\{(M_n - b_n)/a_n \leq z\} \to G(z) \;\; as \;\; n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\begin{aligned} I : G(z) &= e^{e^{-(\frac{z-\beta}{\gamma})}}, && -\infty < z < \infty; \\ II : G(z) &= e^{-(\frac{z-\beta}{\gamma})^{-\alpha}}, && z > \beta; \quad [G(z) = 0, z \leq \beta] \\ III : G(z) &= e^{-(-\frac{z-\beta}{\gamma})^{\alpha}}, && z < \beta; \quad [G(z) = 1, z \geq \beta] \end{aligned}$$

for parameters $\gamma > 0$, β and $\alpha > 0$. These families are referred to as Gumbel[1], Fréchet and Weibull distributions respectively.

1.3.3 The Generalised Extreme Value Distribution (GEV)[9][5]

These families can be generalised into a single distribution known as the Generalised Extreme Value Distribution (GEV). The GEV's cumulative distribution function is as follows:

$$G(z) = e^{-[1+\xi(\frac{z-\mu}{\sigma})]^{-\frac{1}{\xi}}}, \tag{1}$$

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, and where μ, $\sigma > 0$ and ξ are *location*, *scale* and *shape* parameters respectively.

1.3.4 The Block Maxima Approach

Say you have a sequence $X_1, X_2, ...$ as defined in section 1.3.1, you specify a block size of n where n is reasonably large. Break this sequence into blocks of size n and obtain the maximum observation from each block. Now you have a sequence of maxima $M_{(1)}, M_{(2)}, ..., M_{(N)}$, you then fit the GEV to the sequence of maxima and then this gives you a basis for statistical inference using maximum likelihood estimation to find the optimum parameter values for the GEV models cumulative distribution function shown by equation (1).

1.3.5 Threshold Models

The second type of model that can be used to model environmental extremes is known as a threshold model, this model fits to all the data points above a certain specified threshold, this approach uses a lot more of the data and is less wasteful of the data than the block maxima approach. In order to do this, we use a distribution known as the Generalised Pareto Distribution.

1.3.6 The Generalised Pareto Distribution (GPD)[9][5]

Under certain conditions, if it exists, the limiting distribution as $u \to \infty$ of $(X - u|X > u)$ has a Generalised Pareto Distribution of the form,

$$H(y) = 1 - max\left(0, 1 + \left(\frac{\xi y}{\sigma}\right)^{-1/\xi}\right), \tag{2}$$

where $Y = X - u$, $\sigma(\sigma > 0)$ and $\xi(-\infty < \xi < \infty)$. Σ and ξ are the scale and shape parameters respectively. From the cumulative distribution, it would suggest that the GPD doesn't exist as $\xi = 0$. However, by taking the limit as $\xi \to 0$, we obtain the cumulative distribution of the exponential distribution with the mean equal to the scale parameter σ, so the GPD at $\xi = 0$ is like so,

$$H(y) = 1 - exp\left(\frac{-y}{\sigma}\right), \tag{3}$$

defined for the values $y > 0$.

1.3.7 The Return Levels of the Threshold Model[7]

So, for the threshold models, we need to find return levels in order to calculate the apprioriate probabilities of say an 1 in 10000 year occurence, so we need to find the $Pr(X > x)$ where X is the random variable in question. Say we have set a threshold u, then for we have the following conditional distribution,

$$Pr(X > x|X > u) = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi},$$

So, by multiplying both sides by the $Pr(X > u)$, we get the left hand side to become $Pr(X > x)$ by the using conditional probability, thus it follows that we get,

$$Pr(X > x) = \lambda_u\left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}, \tag{4}$$

where $\lambda_u = Pr(X > u)$. Now that we have an expression for $Pr(X > x)$, we set Equation (4) to equal $1/m$ where m will be the number of observations, giving us the probability of $X > x$ once every m observations. By doing this, we get,

$$\lambda_u\left[1 + \xi\left(\frac{x_m-u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m}, \tag{5}$$

where x_m is the level that is exceeded once every m observations.

We want to make x_m the subject of the equation, rearranging Equation (5) gives,

$$x_m = u + \frac{\sigma}{\xi}[(m\lambda_u)^{\xi} - 1] \tag{6}$$

This gives us a return level for the values you expect the random variable X to exceed once every m observations. So, to produce an N-year return level, we set $m = Nxn_y$, thus by substituting this into Equation (6) the N-year return level is as follows,

$$z_N = u + \frac{\sigma}{\xi}[(Nn_y\lambda_u)^{\xi} - 1] \tag{7}$$

where N is the number of years and n_y is the number of observations per year.

1.3.8 Exceedances over the Threshold Method

In order to model data using threshold models, you will choose an apprioriate threshold value such that a GPD explained in Section 1.3.6 is a good model for the exceedances, then fit the GPD to the exceedances using maximum likelihood estimation to find the inferred values for the parameters of the GPD. Once you have this model, you can use the model to produce return levels from equation (7) in Section 1.3.7.

1.3.9 The bivariate logistic model[5][1]

For a bivariate analysis for environmental extremes, there is a bivariate model namely, the logistic model, this model allows for dependence between the two variables y_1 and y_2 with a dependence parameter α. The CDF is as follows:

$$F(y_1, y_2) = e^{-(y_1^{-1/\alpha} + y_2^{-1/\alpha})^{\alpha}} \tag{8}$$

where y_1 and y_2 are the appropriately transformed Fréchet margins. The α parameter ranges from 0 to 1, going from dependent to independent.

Theorem: limiting distributions for bivariate extremes[5][9]

It can be shown that:

$$Pr(x_i, y_i) \to G(x, y)$$

where (X_i, Y_i) are independent and identically distributed with standard Fréchet marginal distributions, and G is non-degenerate. If this is the case, G has the form:

$$G(x, y) = e^{-V(x,y)}; \qquad x > 0, \quad y > 0 \tag{9}$$

where:

$$V(x, y) = 2\int_0^1 max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

where H is a distribution function defined on the range [0,1] which satisfy the mean constraint:

$$\int_0^1 \omega \, dH(\omega)$$

By using Equation (4) in section 1.3.7, we get our $H(x)$ to be:

$$H(x) = 1 - \lambda_u \left[1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} \tag{10}$$

In order to obtain a standard Fréchet margin for our logistic model, modelling threshold exceedances, we need to perform the following transformation:

$$\tilde{x} = -\left(log\left(1 - \lambda_{u_x}\left(1 + \frac{\xi_x(x - u_x)}{\sigma_x}\right)\right)\right)^{-1}$$

So we apply this transformation to both margins, we achieve the form described by Equation (9), where $V(x, y) = (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}$ for the logistic model.

1.3.10 Joint return period estimation for the logistic model[6]

The return period of the occurrence $X_i > x_i, i = 1, 2..$ is expressed as follows:

$$T_{x_i} = \frac{1}{1 - F(x_i)}$$

where $F(x_i) = Pr(X_i < x_i)$. So, by introducing a new variable x_2, the joint return period $T(x_1, x_2)$ of X_1 and X_2 can be condensed down to three possible events of interest. Given any pair of points, we have the following possibilities:

1. $X_1 > x_1$ and $X_2 < x_2$
2. $X_2 > x_2$ and $X_1 < x_1$
3. $X_1 > x_1$ and $X_2 > x_2$

These possibilities can be represented by: [4]

$$T_{(x_1,x_2)} = \frac{1}{1 - F(x_1, x_2)} \tag{11}$$

This form does not account for there being more than one extreme occurrence in one year. To account for this, we can rewrite Equation (11) in the form:

$$R = \frac{1}{kF^*(H_s, T)}$$

in which k is the mean number of extreme events per year and $F^*(H_s, T)$ is the bivariate exceedance probability function of H_s and T.

1.4 The return level standard error[8]

The Delta method can be used to derive the following formula to calculate the standard errors for our return levels defined in section 1.3.7:

$$Var(\hat{z}_r) \approx \nabla z_N^T V \nabla z_N \tag{12}$$

Furthermore, we can create expressions for our return levels as follows using the following operator:

$$\nabla = \left[\frac{\partial}{\partial\sigma}, \frac{\partial}{\partial\xi}\right] \tag{13}$$

So, by partially differentiating Equation (7) with respect to σ and ξ using the operator described by Equation (13), we get:

$$\nabla z_N^T = [\xi^{-1}((Nn_y\lambda_u)^{\xi} - 1), -\sigma\xi^{-2}((Nn_y\lambda_u) - 1) + \sigma\xi^{-1}(Nn_y\lambda_u)^{\xi}log(Nn_y\lambda_u)] \tag{14}$$

These will be evaluated at the point estimates for the parameters of a given model $(\hat{\sigma}, \hat{\xi})$.

2 The Data

The company Sellafield Ltd has supplied us with 7 years and 5 months of data points that were measurements taken of the wind speed/direction and temperature at several mast heights, these measurements were taken at the frequency of one in every ten minutes. These data points range from the start of 2007 all the way to 2014 up to and including May. The plot of all the variables in Appendix 1 shows the missing blocks points to be discussed in Section 2.1, it shows the seasonal variation of temperature throughout the years that you would expect, the wind speeds seem to be fairly random, with plenty of spikes corresponding to large storms which will be the data of most interest to us.

2.1 Issues with the data

We have a large portion of missing data points in 2008 at the beginning of the year whilst the site was under refurbishment. After this, throughout the year, they had a regular check up resulting in a few missing points from thereafter. Further points were missing from wind speeds at the 16m mast height. However, in terms of our analysis, this will not affect it substantially as the missing values were not caused by say extreme conditions, if this was the case and each missing value corresponded to extreme conditions, then the analysis would suffer dramatically as the data would be biased. These data points can be treated as random and completely independent of the variables in question, wind speed and temperature. Also, in the data recording process, due to the nature of the recordings, there were inconsistencies in the mast heights frequency of data, for temperature and wind speed/direction, some mast heights had too many missing values, or weren't recorded for time periods as long as a year. The differing levels of data over the mast heights resulted in us having to only use three mast heights, 2m, 10m and 16m as we plan to use a bivariate model, we needed the points to be at the same height to be compared fairly.

2.2 The Wind Speed Component

For the aims of this project, we have an interest in the wind speeds relative to the angle 118^o to north. So we need to work out the wind speed component relative to this angle. In Appendix 2, we can see that from 190 to 320 degrees has the highest wind speeds predominantly, this is very reassuring for Sellafield Ltd as the most extreme behaviour seems to lie in the angle range that does not have any effect on the cooling system. However, we appear to have a spike in high wind speeds within the range 100-140 degrees circled in the box plots, this range contains the most influential angle of concern. In order to calculate this wind speed contribution we will multiply all the wind speeds by the cosine of the difference in angle from 118^o.
We use,

$$y_{adj,i} = y_i cos(118 - d_i)$$

where $y_{adj,i}$ is the ith wind speed component, d_i is the ith wind direction and y_i is the ith wind speed.

This applies the following cosine curve to each point in the dataset.

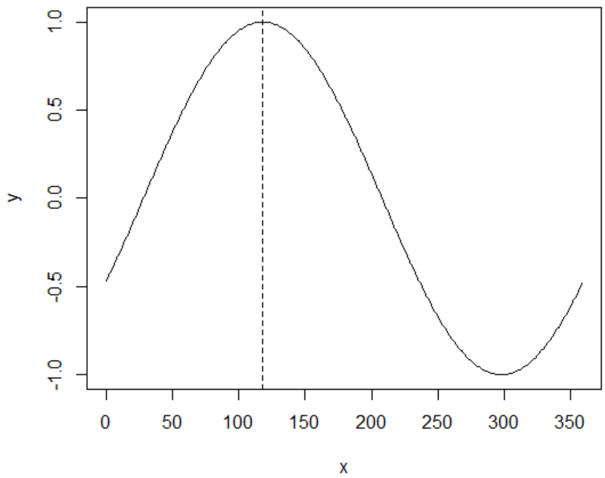

Figure 1: Cosine curve centered at 118^o

From Figure 1, it can be seen that the angle 118^o the cosine is equal to one as this angle has the most influence over the cooling system, with a decay in contribution as you veer away from this angle.

By doing this we have evaluated all the wind speed contributions, by setting on the negative values to zero, in Appendix 3, we produced the same boxplots as before to get a clearer picture of the data points we have to start our analysis on. From these boxplots we have a better picture of the behaviour of the wind speeds of concern.

3 Univariate Analysis

For our univariate analysis we will use the exceedances over threshold method using a generalised pareto distribution(GPD) in Section 1.3.6. In order to do this we must first choose an appriorate threshold value such that the GPD is a suitable model for the exceedances $(X - u_0|X > u_0)$. We shall then fit the GPD to the observed excesses using maximum likelihood estimation to find our model for the variable, from this we can supply return levels.

3.1 Wind Speed

In this section we will find some univarate models for the wind speed, we will assume independence between the exceedances that we model. This may be an issue, as extreme wind behavour usually comes as storms which are sequences of dependent values. So, what may be the case is that one storm could mimic a region of points to be giving us the same amount of information as that number of independent points. The problem with this, is that the dependent points will be only giving us the same amount information as few independent points.

3.1.1 Threshold Choice

We must first decide upon a threshold value for wind speed, we will do this with the aid of parameter stability plots and mean residual life plots being produced by R packages 'evd'[12] and 'ismev'[11] respectively. These plots will give us a good indication of our choice of threshold for the variable wind speed.

3.1.2 Mean residual life plots

Firstly, we will look at the mean residual life plots, by using [3] we know that the if all the exceedances x_i above a certain threshold say u_0 then the mean value of $(x_i - u)$ plotted against $u > u_o$ should give a linear plot if the GPD is the correct model for the exceedances. So, by looking at the mean residual life plots found in Appendix 4, we can chose suitable thresholds for each mast heights. From observation, 13.5 for 2m, 18.3 for 10m and 19.9 for 16m.

However, an issue with these choices is that we have very little data above the threshold chosen, we have a dataset of length 389952 and are using only 24,21 and 36 data points for mast heights 2m,10m and 16m respectvely, we have a lot of data wastage this suggests that the high end values are extremely dependent on each other as these collection of points of a large storm, so the original assumption that the exceedances are independent may be wrong. What we are observing is most likely a small collection of storms. We shall look at parameter stability plots to confirm our choice of threshold.

3.1.3 Parameter Stability Plots

In these plots, if we chose a particular threshold, we need to ensure that all further threshold choices greater than the chosen threshold has 95% confidence intervals that contain the chosen threshold. So, this assesses the stability of our estmates for ξ and σ. So, from looking at appendix 5, we confirm these are our threshold choices, as the parameters only seem stable after our choice of thresholds for the wind speed, this is a problem because we cannot model the extremal behaviour of the wind speed over approximately 7 years using such a small subset of the data. We need to be at least using approximately using 1-5% of the data to model the extremal behaviour reliably. We need to account for this dependence.

3.1.4 Declustering[9]

In order to account for the dependence we encountered with our wind speed data in section 3.1 will use a method called declustering. As our assumption of independence failed, we must condense these dependent regions, '*clusters*'. Say one storm goes on for a long period of time, these points are telling us and mimicking the same information of the same amount of independent observatons. So to get around this, we need to condense these regions down to one point which will be approximately independent to the other cluster peaks. We will take the most extreme value in each cluster throughout the entire series of points. This enables us to still be able to model the most extreme behaviour and fit a GPD to these '*cluster peaks*'.

To decluster a data set we need to decide the following:

1. The declustering threshold u
2. The declustering cooldown interval κ

The declustering threshold u is the value that triggers the start of the cluster, once this has been exceeded you initiate the start of a cluster. The declusterng cooldown interval is the number of consecutive values that fall below the threshold in order for us to classify that the cluster has ended and all the dependence has passed from the current cluster. This approach is referred to as the *peaks over threshold approach* [3]

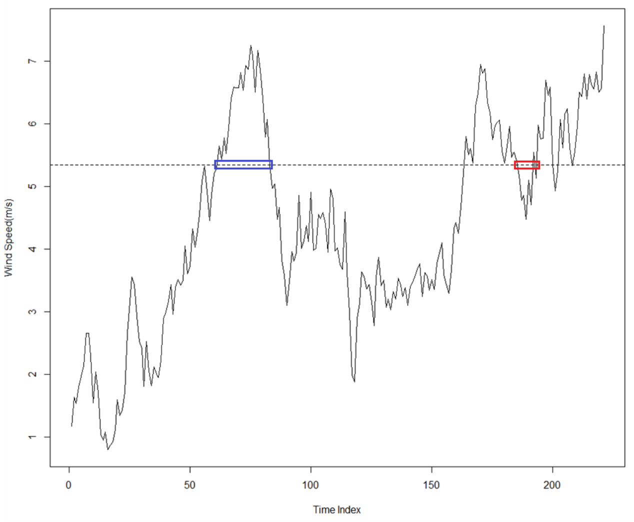

Figure 2: Declustering Example, threshold represented by the dotted line

From Figure 2, you can clearly see the first clusters beginning and end illustrated by the blue box, with the cluster peak being the most extreme value in that region. In this example,we assume that the cooldown interval has been satisfied thus the cluster has ended. However, looking at the second cluster, you can see that the cooldown interval may have not been satisfied, illustrated by the red box, this is an occurence where the cluster is not deemed to have finished, so the cluster continues and the cooldown period is resets and restarts when the current data string falls below the declustering threshold again.

The declustering threshold choice

The choice of the declustering threshold for the wind speeds were based on the percentage of data above the declustering threshold, we used trial and error to find a compromise between this and the consequential very large clusters caused by very low choices of the declustering threshold. With the declustering threshold being too low we get very little data and long clusters, this was caused by long periods of time being considered a single cluster and very rarely was the cooldown period satisfied. With a very high declustering threshold choice we got a similar problem as very little values exceeded the threshold so we ended up with very little cluster peaks. We went for roughly 15% of the data to be above the declustering threshold, this gave us a reasonable amount of data to fit the GPD to, and no substantially long clusters, clusters that extended for longer than a month were a benchmark.

The declustering cooldown interval choice

The choice of the declustering cooldown interval needed to be sufficiently long such that the long storms that may dip below the threshold for a while during certain times of the day were still considered the same storm and not considered two separate storms as this would go against why we are declustering. But needed to be sufficiently short to ensure that we don't merge two storms together and lose a storm in the process of declustering. We used a cooldown period of 5 days, this seemed to be a reasonable cooldown period for a storm to be said as finished, we assumed extreme storms wouldn't prolong for this amount of time.

3.1.5 The declustered wind speed data

For the declustered data set, we decided upon the declustering thresholds of 4, 5 and 7 for mast heights 2m, 10m and 16m respectively. Now we need to return to deciding upon a threshold choice for our GPD model for the exceedances of the declustered peaks. So, we need to produce mean life residual plots explained in Section 3.1.2 for the new data set.

From looking at the mean residual life plots from Appendix 6, we can see that there is a clearly a linear relationship that can be achieved for all mast heights. This suggests that we have succeeded in removing most of the dependence caused by the clusters and we can now model the extreme behaviour of the wind speed cluster peaks. From the mean residual life plots we decided to chose the following thresholds for each wind speed mast height.

Mast Height(metres)	2	10	16
Threshold(m/s)	7.8	11.1	12.6
Num. of Exceedances	182	168	161

Table 1: Threshold Choices for declustered wind speeds

3.1.6 The fitted wind speed models

2m wind speed model

For our first model at mast height 2m, the GPD fits really well for our choice of threshold $7.8(\kappa = 4)$, this gives us 182 cluster peak exceedances out of 1469 exceedances, $x_i; i = 1, ..., 182$, by using maximum likelihood estimation we get:

$$\hat{\sigma} = 2.006(0.190) \qquad \hat{\xi} = -0.118(0.0592).$$

Diagnostics for the 2m wind speed model

By looking at the diagnostics plots in Appendix 7, we can see from the probability and quantile plots that the model is fitting very well. We have a few stray values on the end of the quantile plot but this is not anything to be concerned about. We can see the shape of the density and the return level plot giving our predictions about the probabilty of an exceedance in years graphically.

10m wind speed model

For our second model at mast height 10m, the GPD fits really well for our choice of threshold $11.1(\kappa = 6)$, this gives us 168 cluster peak exceedances out of 1259 exceedances, $x_i; i = 1, ..., 168$, by using maximum likelihood estimation we get

$$\hat{\sigma} = 2.321(0.251) \qquad \hat{\xi} = -0.0679(0.0763).$$

Diagnostics for the 10m wind speed model

By looking at the diagnostics plots in Appendix 8, we can see from the probability and quantile plots that the model is fitting very well. We have one outlier in the quantile plot but this is nothing to be too concerned about, this is may due to the variation in the data. The return level plot giving our predictions about the probabilty of an exceedance in years. The density looks very similar to an exponential distribution, but this is because the parameter ξ is very close to zero, as discussed in section 1.3.6 and by Equation (3).

16m wind speed model

For our third model at mast height 16m, the GPD fits really well for our choice of threshold 12.6($\kappa = 7$), this gives us 161 cluster peak exceedances out of 1131 exceedances, $x_i; i = 1, ..., 161$, by using maximum likelihood estimation we get

$$\hat{\sigma} = 2.624(0.288) \qquad \hat{\xi} = -0.0897(0.0767).$$

Diagnostics for the 16m wind speed model

By looking at the diagnostics plots in Appendix 9, we can see from the probability and quantile plots that the model is fitting very well. We have a few stray values on the end of the quantile plot, but this is of no concern. We can see the shape of the density and the return level plot giving our predictions about the probabilty of an exceedance in years graphically.

Stability of the parameters

We must check the stabilty of the parameters for our choice of threshold for all models, we need to do this to confirm and investigate that our choice of threshold makes the GPD the correct model for the excesses.

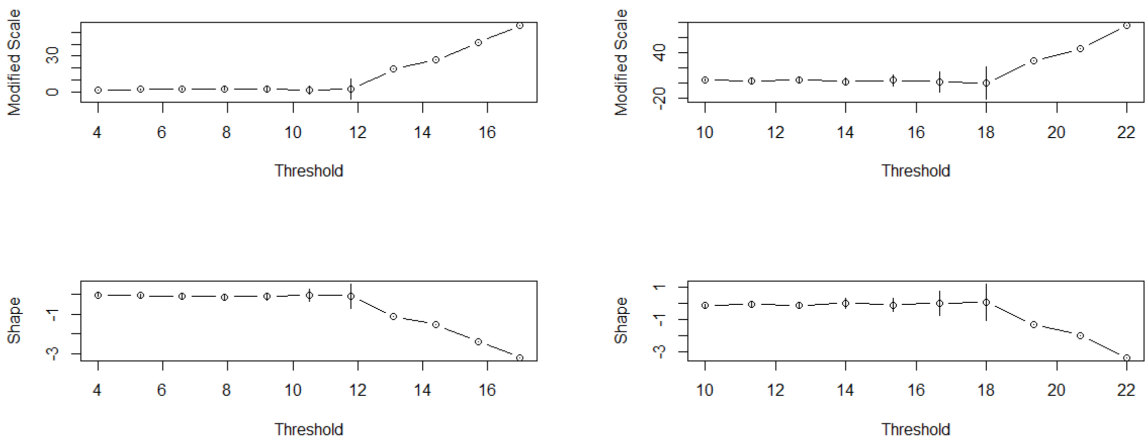

Figure 3: Parameter Stability plots of 2m and 10m wind speed respectively

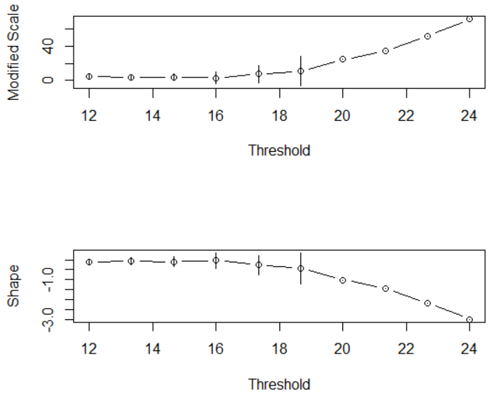

Figure 4: Parameter Stabilty Plots of 16m wind speed

From the Figures 3,4 above, the points with no error bar are representations of the an error bar that extends beyond the range of the y-axis. By observing these plots, we can be reassured of our choices of threshold for all mast heights as the parameters seem to be stable throughout. However, the plots seem to suggest parameter stability for lower thresholds than we have chosen. This needs further investigation, if we can push the threshold lower and still maintain a good fit to the GPD model, then we will have more data to use for our analysis and we will be able to better predict the extreme behaviour of the wind speed.

3.1.7 The final wind speed models

So, we will fit the models to a GPD, but slowly decrease the threshold towards the declustering threshold looking for any deterioration in fit as we lower the threshold. By performing this task, we see very little deterioraton of fit, for all mast heights we manage to get back to the declustering threshold and the data still fits really well. You can see this by the diagnostic plots in Appendices 10-12 of all the models with there new thresholds of 4,6,7 for the mast heights 2m,10m and 16m respectively. So, by doing this, we are using all our cluster peaks for our analysis of the extreme behaviour and have independent observations. We can see that by using all the observations that ξ parameters have all become closer to zero as the densities seemed to approach the Exponential distribution from their previous density.

Finally, our new models for the univariate wind speed are GPD with the following scale and shape parameters:

Mast Height(metres)	2	10	16
Threshold(m/s)	4	6	7
κ	4	6	7
$\hat{\sigma}$	1.860(0.0696)	2.409(0.100)	2.746(0.122)
$\hat{\xi}$	-0.0215(0.0268)	-0.00606(0.0307)	-0.0134(0.0332)
Num. of Exceedances	1469	1259	1131

Table 2: Threshold Choices for declustered wind speeds(lower threshold), standard errors in parentheses to 3 s.f.

3.2 Temperature

In this section, we will fit univarate models to the temperatures extreme behaviour, we will assume independence between the exceedances that we model. This may be an issue, as extreme temperatures usually come for prolonged periods(e.g. heatwaves) which are sequences of dependent values. So, what may be the case is that one heatwave could be interpretated as several independent extremes. Also, what we will have to also consider, is that temperature has a daily cycle, which is good for Sellafield Ltd, as prolonged heated conditions are naturally rare due to the cooling at night.

3.2.1 Mean residual life plots

By looking at the mean residual life plots of the temperatures at all all mast heights in Appendix 13, as we need to guarantee a linear relationship between the mean excess and all thresholds above a certain chosen threshold explained in section 3.1.2, we decide to go for the thresholds 23.5, 23.8 and 23.2 for the mast heights 2m, 10m and 16m respectively as this gives an approximate linear relationship from these threshold onwards. These thresholds have a respective 483, 324 and 411 exceedances above there corresponding thresholds, this is quite small. We must check parameter stability plots to check our choice of thresholds.

Mast Height(metres)	2	10	16
Threshold(Celsius)	23.5	23.8	23.2

Table 3: Threshold Choices for Temperatures

3.2.2 Parameter stability plots

By looking at the parameter stability plots of the temperatures at all mast heights given in Appendix 14, our choice of threshold seems to be reassured, however, the last few thresholds seem disconcerting, suggesting that our choice of threshold may be incorrect. We shall fit the model and assess its fit to investigate our choice of threshold further.

3.2.3 Fitting the GPD to the exceedances

We fitted a GPD model to the exceedances of temperatures for all mast heights with our chosen thresholds shown in Table 3, by fitting these exceedances we get the diagnostic plots shown in Appendices 15-17. It is very clear in all cases that the probability plot shows that the models aren't fitting well. The quantile plots also suggest a lack of fit with the data. The return level plots are very bad also, as the alot of the data is falling out of the range of the 95% confidence intervals and is not very helpful to predicting the extreme behaviour. We need to consider the daily cycle of temperature into our analysis of the data and we need to account for some of the dependence that temperature will have when measured as close as 10 minute periods.

3.2.4 Block maxima

In this section, we will take block maxima of the temperature, in particular we will look at the daily maxima of temperatures in order to remove some of the dependence in the data and this method also accounts for the daily cycle that temperature possesses. As explained in section 1.3.4, we will simply take the maximum temperature value for each day in the whole data set, we will then perform our analysis on this new dataset using a threshold model, this new sequence of observations will hopefully fit the GPD model better than the raw data.

3.2.5 Fitting to the Daily Maxima

In this section, we will fit the GPD models to the exceedances of the daily maxima, again, using maximum likelihood estimation.

Choosing a Threshold value

By using Appendix 18, we chose the illustrated thresholds which are 16.2, 16.1 and 16.1 for mast heights 2m,10m and 16m respectively. From these thresholds seem to give a linear relationship afterwards.

Mast Height(metres)	2	10	16
Threshold(Celsius)	16.2	16.1	16.1
Num. of Exceedances	604	487	506

Table 4: Threshold Choices for Daily Temperatures

Parameter Stability
We must check for parameter stability to reinforce or to find any problem our choice of threshold for the daily temperatures.

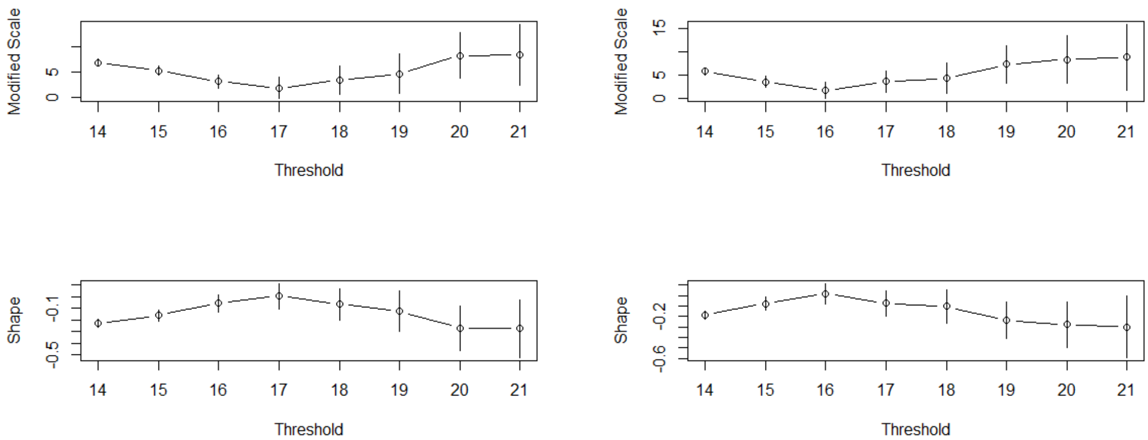

Figure 5: Parameter Stability plots of 2m and 10m daily temperatures respectively

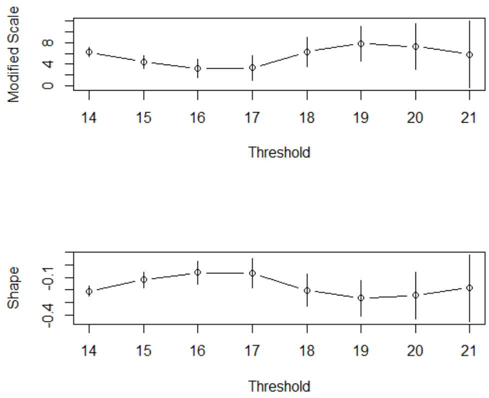

Figure 6: Parameter Stabilty Plots of 16m daily temperatures

From figures 5 and 6, we can see that our parameters are stable for our choice of threshold, the threshold choice of 15 may be a possibility in order to be a good fit to the GPD. We will investigate this possibility further as we assess the deterioration of fit as we decrease the threshold.

Deterioration of fit

By steadily decreasing the threshold from the chosen thresholds, we get slight changes, our new thresholds are 16, 15 and 15 for the mast heights 2m , 10m and 16m respectively. These are our final choices for the threshold for our univariate model of the daily temeratures.

3.2.6 The final temperature models

So, our final models for the univariate daily temperatures are GPD with the following scale and shape parameters:

Mast Height(metres)	2	10	16
Threshold(Celsius)	16	15	15
$\hat{\sigma}$	2.225(0.120)	2.374(0.116)	2.612(0.125)
$\hat{\xi}$	-0.0568(0.0372)	-0.0766(0.0330)	-0.120(0.0321)
Num. of Exceedances	761	776	776

Table 5: Models for the daily temperatures, standard errors in parentheses to 3 s.f.

We can see by the diagnostic plots of the models in Appendices 19-21, that the daily temperature exceedances are fitting very well, by looking at the probability and quantile plots, the fit exceedances seem to fit well, each mast height has a few stray endpoints on the quantile plot but that's of no concern. We have graphical return levels for our prediction of how likely such high temperatures are to occur. All the densities look very similiar to the Exponential distribution as the ξ value is quite close to zero.

3.3 Moving average wind speed model

The company Sellafield Ltd are also interested in prolonged weather conditions, stated in Section 1.2, a way of transforming the data to make it more descriptive of this behaviour we will use moving averages. We will use the periods of most concern which are one and two hour intervals, this will give us a dataset of prolonged weather conditions so we can then do our ordinary analysis on this new dataset to model the prolonged weather conditions for both one and two hour periods. This will give us a very good idea on the type of conditions that Sellafield Ltd should be aware of the possibility of.

3.3.1 Simple moving average

A moving average is a chain of averages calculated by successive overlapping segments of equal size of a series of values. So, given we have a dataset $(x_1, x_2, ..., x_n)$ and we have segment size k. Then the moving average will be as follows:

$$
\begin{aligned}
MA_1 &= \frac{1}{k}\sum_{i=1}^{k}(x_i) \\
MA_2 &= \frac{1}{k}\sum_{i=2}^{k+1}(x_i) \\
\vdots \quad & \vdots \quad \vdots \quad \vdots \\
MA_m &= \frac{1}{k}\sum_{i=n-k}^{n}(x_i)
\end{aligned}
$$

We will use the moving averages of the wind speed component relative to 118^o so that we get a model for the prolonged conditions of the wind speed direction. This is of particular interest researchers at Sellafield Ltd company. Firstly, we will fit the data assuming that the data points have no dependence structure.

3.3.2 Mean life residual plots

By looking at Appendix 22, we can see that we have a similar situation to the previous analysis in Section 3.1 of the wind speed, this was expected as we have moving averages of the original dataset which also possessed this dependence structure, by taking moving averages, we would have introduced more dependence between observations. By attempting to fit these exceedances using the threshold choices depicted in Appendix 22 to a GPD, we get very poor fits as expected, so let us go straight to declustering our data of moving averages.

3.3.3 The declustered moving averages

Whilst declustering a dataset of moving averages, we must also consider that we have introduced some dependence in the data, so we will have to consider this when deciding upon a declustering cooldown interval. We went for the following declustering thresholds 4, 6 and 7 for the mast heights 2m,10m and 16m respectively. Our choices for a cooldown interval were our original cooldown interval from section 3.1.7 plus the moving average period. By looking at appendix 23, we can see that the declustering has succeeded in removing the majority of the dependence of the moving averages and will be good fits to the GPD as all mean life residual plots show a linear relationship.

3.3.4 The declustered moving average models

We fit these exceedances to the GPD, as in Section 3.1, the parameter stability plots seem to suggest we can push the threshold choice for the GPD back to the declustering thresholds. By doing this and checking the deterioration of fit, we can push the threshold choice back. By looking at the diagnostics in Appendix 24, we can see that the probability plots and the quantile plots don't detect anything alarmingly wrong with the fit suggesting the GPD is a good fit to the data. A few outliers near the tail of the quantiles appear, but are of no concern.

So, our final models for the moving average wind component model are GPDs with the following parameters:

Mast Height(metres)	2	10	16	2	10
Threshold(m/s)	4	6	7	4	6
$\hat{\sigma}$	1.388(0.0906)	1.570(0.129)	1.885(0.147)	1.357(0.0942)	1.442(0.133)
$\hat{\xi}$	-0.131(0.0405)	-0.0971(0.0500)	-0.113(0.0434)	-0.134(0.0404)	-0.0727(0.0574
Num. of Exceedances	755	732	689	644	627

Table 6: Moving average wind speed component models, standard errors in parentheses to 3 s.f.

3.4 Univariate Return Levels

In this section, we will calculate various return levels for our pre-existing models from Sections 3.1.7, 3.2.6 and 3.3.4 to work out the types of temperatures and wind speeds you will expect over certain periods of time and how likely these extremes are to occur. We will be using the return level formula derived in Section 1.3.7. All standard errors will be calculated using the delta method in Section 1.4 using Equations (12) - (14).

3.4.1 Wind speed models

In this section, we will look at the types of extremes of wind speed we expect assuming there is no severe changes in behaviour throughout the future years caused by global warming as an example.

Return Level(years)	10	100	1000	10000	100000
Wind speed extreme(m/s)	17.08595	20.63455	24.01213	27.22693	30.28678
Standard error	1.002018	1.735527	2.643892	3.707875	4.910085

Table 7: Return levels for the declustered wind speeds(mast height 2m)

Return Level(years)	10	100	1000	10000	100000
Wind speed extreme(m/s)	23.60025	28.86458	34.05599	39.17549	44.22407
Standard error	1.54139	2.766577	4.344059	6.262696	8.512264

Table 8: Return levels for the declustered wind speeds(mast height 10m)

Return Level(years)	10	100	1000	10000	100000
Wind speed extreme(m/s)	21.51129	27.2962	32.90483	38.34257	43.61462
Standard error	0.9206395	1.984886	3.450757	5.289414	7.475657

Table 9: Return levels for the declustered wind speeds(mast height 16m)

From Tables 7 to 9 above, we can see the types of wind speed we would expect if the wind speeds behave in the same meteorological manner as the past ≈ 7.4 years. We have the once every ten years to the once in every hundred thousand years probability, the standard errors were calculated using the delta method explained in Section 1.4 . The tables clearly show that the higher the mast height the more extreme wind speeds you should expect, this is as expected as you increase height, wind conditions get more severe.

3.4.2 Daily Temperatures

In this section, we will look at the types of extremes in temperature we predict to see assuming there is no dramatic changes in behaviour throughout the future years.

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Temperature extreme(Celsius)	28.52947	31.79626	34.66251	37.17734	39.38382
Standard error	1.175846	2.061709	3.104662	4.251589	5.460847

Table 10: Return levels for the daily temperatures(mast height 2m)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Temperature extreme(Celsius)	27.76595	30.71077	33.17914	35.24817	36.98244
Standard error	1.663737	2.787769	4.033241	5.330795	6.63075

Table 11: Return levels for the daily temperatures(mast height 10m)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Temperature extreme(Celsius)	27.32278	29.60709	31.34051	32.6559	33.65407
Standard error	1.275873	2.028923	2.789123	3.508604	4.162581

Table 12: Return levels for the daily temperatures(mast height 16m)

From Tables 10 to 12, we can see that the daily temperatures extreme values decrease as you increase the height, this would be as expected as it tends to be colder on average at higher altitudes. Also, it can be noticed that as the temperatures rise, then the wind decreases, there seems to be a negative correlation between the two variables.

3.4.3 Moving average models

We will find the return level values for the prolonged conditions of one and two hour periods of the wind speed relative to the direction 118^o which are of most interest for Sellafield Ltd. We will firstly look at the one hour period moving averages.

3.4.4 One hour period

These will be the expected extreme values of wind speed that will prolong for a one hour period.

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	9.905473	11.12056	12.01835	12.68171	13.17184
Standard error	0.4458034	0.7343665	1.033134	1.316255	1.57124

Table 13: Return levels for the moving average wind speeds(mast height 2m, one hour)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	12.90784	14.76444	16.249099	17.4363	18.38568
Standard error	0.6458221	1.152773	1.733454	2.336138	2.928027

Table 14: Return levels for the moving average wind speeds(mast height 10m, one hour)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	14.98515	16.97438	18.50734	19.68869	20.59908
Standard error	0.651473	1.110859	1.622311	2.136116	2.623587

Table 15: Return levels for the moving average wind speeds(mast height 16m,one hour)

From Tables 13 to 15, we have our prolonged conditions of concern for one hour periods, these are all above the wind speeds that will effect the cooling system. These extremes are going to have a high negative impact on the cooling system.

3.4.5 Two hour period

These will be the expected extreme values of wind speed that will prolong for a two hour period.

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	9.62822	10.82966	11.71304	12.362691	12.84041
Standard error	0.4170993	0.6870849	0.9699182	1.23896	1.481357

Table 16: Return levels for the moving average wind speeds(mast height 2m, two hour)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	12.58022	14.60791	16.31917	17.76339	18.98224
Standard error	0.7104049	1.336769	2.096703	2.93089	3.796772

Table 17: Return levels for the moving average wind speeds(mast height 10m, two hour)

Return Level(years)	1 in 10	1 in 100	1 in 1000	1 in 10000	1 in 100000
Wind speed extreme(m/s)	14.60015	16.76729	18.52686	19.955519	21.11548
Standard error	0.7157696	1.283911	1.951398	2.658649	3.366217

Table 18: Return levels for the moving average wind speeds(mast height 16m,two hour)

From Tables 16 to 18, we have our prolonged conditions of concern for two hours, a similar picture to the one hour period return levels, we have very high prolonged wind speeds that will have a negative impact on the cooling system.

4 Bivariate Analysis

In this section, we will fit a bivariate model to both temperature and wind speed to model the combinations of both factors to assess the types of conditions the researchers at Sellafield Ltd company can expect. We will supply some extreme scenarios of combinations of wind speed and temperature and how likely these are to occur. We will use a software package in R called 'evd'[12] to fit these models to the exceedances. So, by using the logistic model in Section 1.3.9, we will use maximum likelihood estimates to infer our parameters for our bivariate peaks over the thresholds, the method is described in [2], we need to assume independence between the bivariate pairs.

4.1 Bivariate Datasets

For our moving average datasets, we will still have dependence within the points in their raw state, we have a problem that the wind speeds in the univariate case needed to be declustered and the temperatures we fit to the daily maxima. This makes it hard to model the bivariate case as we don't have independent bivariate pairs to fit the logistic model to. As a solution, we will look at all the declustered wind speed peaks, then take the associated temperature value for the given interval of time this extreme occured. However, this will only give us the wind speeds extreme behaviour, we then do the same for the temperature. We decluster the temperatures and take the associated wind speed from these peaks. This is where we need to be careful as our bivariate pairs need to be approximately independent to fit the logistic model, we need to check if we have any cases that we have extracted the same peak from both datasets. We need to also find any potentially dependent observations (e.g two points lying in the same cluster lying too close to one another to be considered approximately independent). By doing this we find a few identical points in both datasets, we then removed one of the repeated points, we find a few overlaps in the dataset cluster intervals, however, none were substantially close to question that we have enough dependence to severely effect our model. Our assumptions seem to hold. All wind speeds considered in the bivariate analysis will be the wind speed component discussed in Section 2.2.

4.2 Bivariate one hour moving average models

In this section, we will look at the bivariate pairs for the prolonged conditions of one hour, we will find some scenarios in which we expect to see as a combination of the two variables. Throughout the analysis in this section, the wind speed component will be the symbolised as y_1 and temperature y_2.

Mast Height(metres)	2	10	16
$\lambda_{u_{y_1}}$	0.444	0.360	0.359
$\hat{\sigma}_{y_1}$	1.198(0.0677)	1.435(0.0847)	1.128(0.0729)
$\hat{\xi}_{y_1}$	-0.134(0.0212)	-0.147(0.0133)	0.000606(0.0384)
$\lambda_{u_{y_2}}$	0.429	0.704	0.703
$\hat{\sigma}_{y_2}$	1.604(0.0955)	1.342(0.0796)	1.511(0.827)
$\hat{\xi}_{y_2}$	-0.0200(0.0334)	0.1044(0.0419)	0.0200(0.327)
$\hat{\alpha}$	0.999(0.000)	1.000(0.000)	0.999(0.000)

Table 19: Parameter estimates of the bivariate models for the one hour moving average dataset, standard errors in parentheses to 3 s.f.

From Table 19, we can see that in all models, there is a striking inference that both the variables are independent at the extremes. A lot of the ξ parameter values are close to zero.

4.3 Bivariate two hour moving average models

In this section, we will look at the two hour prolonged conditions bivariate pairs to find some scenarios in which we expect to see as a combination of the two variables. Throughout the analysis in this section, the wind speed component will be the referred to as x_1 and temperature x_2.

Mast Height(metres)	2	10	16
$\lambda_{u_{x_1}}$	0.399	0.324	0.3244
$\hat{\sigma}_{x_1}$	1.109(0.0654)	1.301(0.0908)	1.126(0.0925)
$\hat{\xi}_{x_1}$	-0.0999(0.0319)	-0.113(0.0239)	-0.0282(0.0435)
$\lambda_{u_{x_2}}$	0.736	0.706	0.736
$\hat{\sigma}_{x_2}$	1.632(0.0106)	1.344(0.0792)	1.647(0.0947)
$\hat{\xi}_{x_2}$	0.00398(0.0419)	0.0914(0.0401)	0.000518(0.0345)
$\hat{\alpha}$	0.999(0.000)	0.999(0.000)	0.999(0.000)

Table 20: Parameter estimates of the bivariate models for the two hour moving average dataset, standard errors in parentheses to 3 s.f.

From Table 20, We have a similar picture to the one hour moving average in Section 4.2, we can see that in all models, there is a huge inference that both the variables are independent at the extremes. many ξ values are close to zero.

4.4 Bivariate joint return levels

In the bivariate form, we have the return levels as curves instead of exact points like in Section 3.4. So, we will use a graphical representation of the return levels for each model to describe the extreme combinations of the two variables. These lines will describe the types of extreme combinations we expect assuming that wind and temperatures behaviour doesn't change dramatically in the given periods of time.

4.4.1 One hour moving average bivariate model

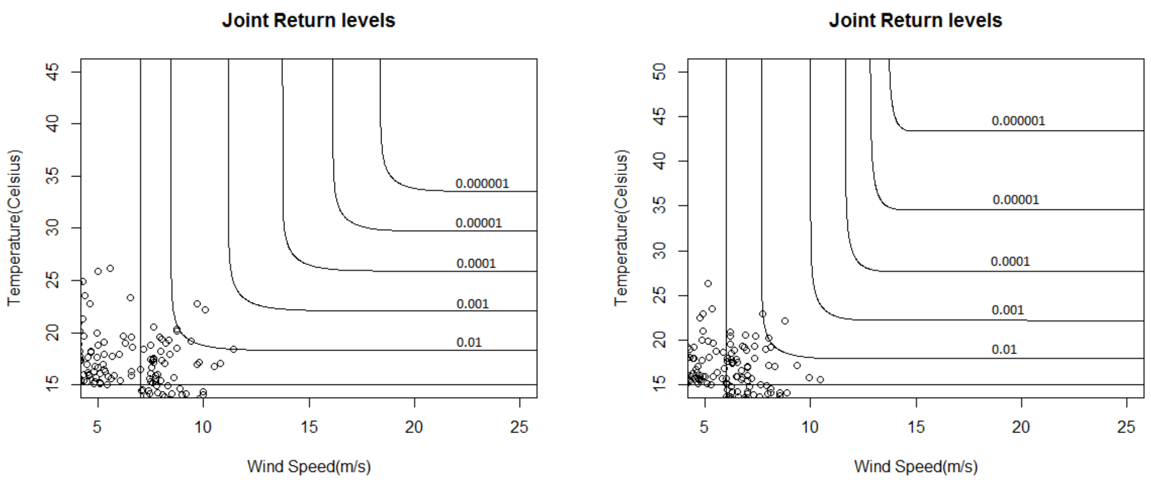

Figure 7: Return level plots for one hour moving average model(mast heights 2m and 10m respectively)

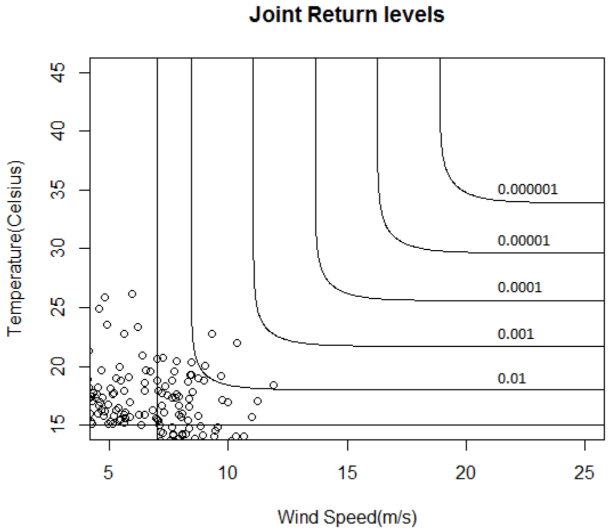

Figure 8: Return level plot for one hour moving average model(mast height 16m)

From observation of the extreme scenarios that can occur for the bivariate model depicted in Figures 7 and 8, comparing it to the univariate models we can see that the extreme pairs are much less extreme, so combinations of the two variables are much less extreme when you account for the interaction between both wind speed and temperature, this will be more comforting for Sellafield Ltd company.

4.4.2 Two hour moving average bivariate model

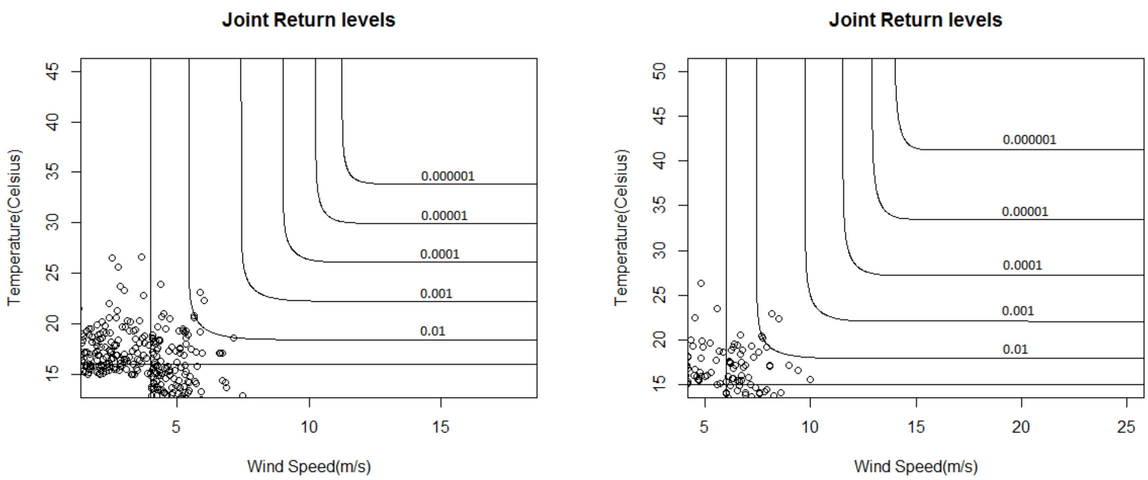

Figure 9: Return level plots for two hour moving average model(mast heights 2m and 10m respectively)

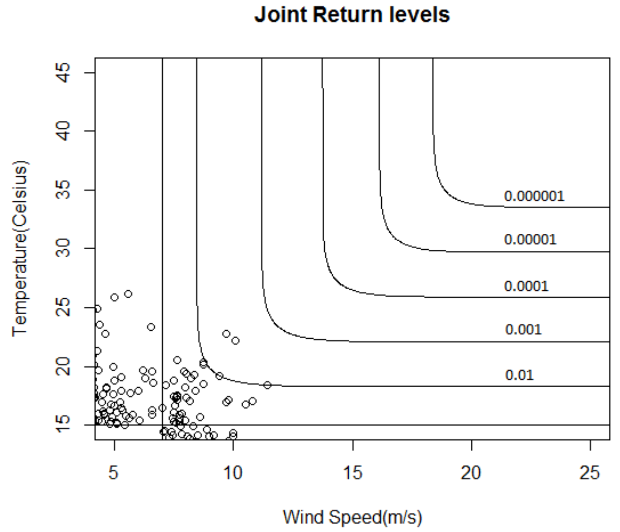

Figure 10: Return Level Plot for two hour moving average model(Mast height 16m)

From Figures 9 and 10, we have a very similar picture to that of in Section 4.4.1, the bivariate extremes seem much less severe to there univariate models, again, very comforting for Sellafield Ltd company.

5 Conclusion

5.1 Overview, discussions and limitations

Overall, we have modelled the environmental extremes of the wind speeds and temperature in the area of Sellafield, the nuclear power station. We have done this both by looking at the univariate and bivariate cases, we have looked at prolonged conditions of these factors and also the most extreme ten minute spike at three mast heights 2m, 10m and 16m. From this, we have found estimates of the most extreme cases we expect to see in the form of return levels. These are described in tabular form for the univariate case in Section 3.4, and in graphical form for the bivariate case in Section 4.4, from these, just by inspection it is very clear that scenarios are all of interest for Sellafield Ltd company as these conditions predicted are well above the temperatures and wind speeds of concern. Sellafield Ltd company may decide to use these univariate values and bivariate lines in order to simulate possible scenarios that may occur at the provided probabilities, thus they can continue there research in the performance of there cooling system. As the building itself will only be in operational for less than hundred years, the first two return levels of the moving average will be the most useful, however, the research being performed by Sellafield Ltd is to ensure safety, thus, the return levels that will outlive the building are necessary to be absolutely sure that there's no chance of anything drastically will go wrong with the cooling system even at the most unlikely prolonged conditions. However, the results we have found are all within the range of concern, importantly, including the one in ten year return level. By first observation of the univariate cases, the extremes seem very high, however, when we look at the bivariate model, the conclusions change, the combination of the two variables lead to less extreme pairs of values. This is as expected, intuitively, as the wind speed increases, conditions generally are cooler due to the nature of diffusion of heat energy in the atmosphere. Clearly, accounting for the interaction and possible combinations of temperatures and wind speeds at there extremes gives us a much more comforting and less extreme picture of the behaviour of both variables. A few limitations to the analysis are that we have created return levels based on the assumption that the following years follow the same meteorological patterns as the past 7.4 years, this ignores any potential enironmental change over the furture years which may occur due to global warming. Furthermore, the method of declustering is very subjective and the personal opinion of what the ideal cooldown and declustering threshold can influence your statistical influence.

5.2 Areas for improvement and future work

With the univariate analysis, we could have chose not to decluster the data, we could have fit the GPD to all exceedances, thus ignoring the dependence. From this you could adjust the inference by inflation of standard errors to take into account for the reduced information. This also models the temporal dependence. For both univariate and bivariate analysis, we could have performed a bayesian analysis to further improve our model, given an apprioriate prior supplied by an expert. The dataset we had was large, however, 7.4 years is a very small dataset when calculating the highly unlikely return levels(e.g thousand years) that we were calculated in this project. I we had more measurements dating back further, we will have improved our model and give ourselves a better idea of the behaviour of the weather as we would have had a better idea of the behaviour of the two variables in the Sellafield area. Also, we could have tested several types of models in the bivariate case, such as bilogistic model which allows for asymmetry. From this point, having seen that the dependence parameter infers independence, we could have further investigated this matter, and perhaps try a model that doesn't allow for dependence thus reducing the parameters in the model.

6 Bibliography

References

[1] GUMBEL, E. J. *Distributions des Valeurs Extremes en Plusieurs Dimensions*, 1960, Publ. Inst. Statist, Univ. Paris 9.

[2] Coles, S. *Introduction to Statistical Modeling of Extremes Values*, 2001, Springer Basel, New York.

[3] Davison and Smith. *Models for Exceedances over High Thresholds*, 1990,Wiley for the Royal Statistical Society.

[4] YUE, S. & WANG, *A comparison of two bivariate extreme value distributions*, 2004 ,Stochastic Environmental Research 18.

[5] Jones, Mari,*Characterising and modelling time-varying rainfall extremes and their climatic drivers* ,2012 ,Newcastle University.

[6] Panagiota Galiatsatou & Panayotis Prinos, *Bivariate models for extremes of significant wave height and period*, 2007, Dept. of Civil Eng.

[7] Cai Juan Juan, Fougres Anne-Laure and Mercadier Ccile,*Environmental data : multivariate Extreme Value Theory in practice*, 2013, Journal of the French Society of Statistics.

[8] *How can I estimate the standard error of transformed regression parameters in R using the delta method?* , UCLA Statistical Consulting Group, http://www.ats.ucla.edu/stat/sas/notes2/, accessed April 2015.

[9] Lee Fawcett and Dave Walshaw,June 2008, Newcastle University, *Modelling Environmental Extremes*, http://www.mas.ncl.ac.uk/ nlf8/shortcourse/fullset.pdf.

[10] Fisher and Tippett, *International Encyclopedia of Statistical Science*,1928, Springer Berlin Heidelberg.

[11] Original S functions written by Janet E. Heffernan with R port and R documentation provided by Alec G. Stephenson. (2012). ismev: An Introduction to Statistical Modeling of Extreme Values. R package version 1.39. http://CRAN.R-project.org/package=ismev

[12] A. G. Stephenson. evd: Extreme Value Distributions. R News, 2(2):31-32, June 2002. URL: http://CRAN.R-project.org/doc/Rnews/

7 Appendices

7.1 Appendix 1 - Plots of Raw Data

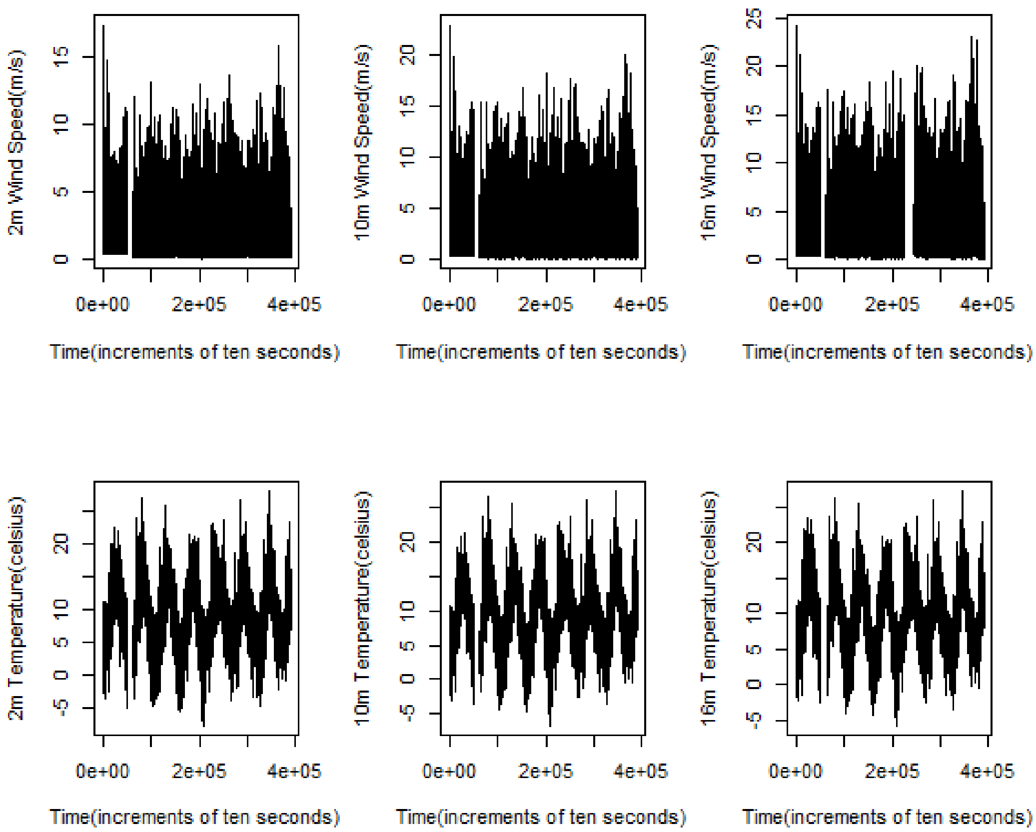

Figure 11: Raw data versus time

7.2 Appendix 2 -Wind speeds relative to direction

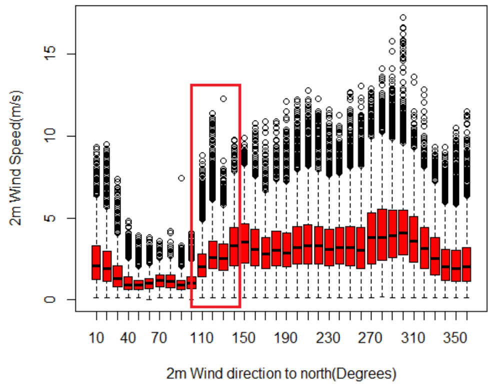

Figure 12: Boxplot of wind speed and direction at mast height 2m

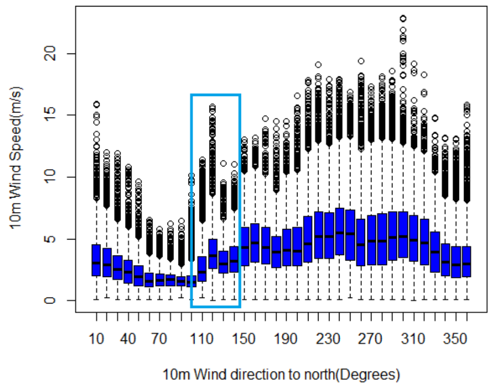

Figure 13: Boxplot of wind speed and direction at mast height 10m

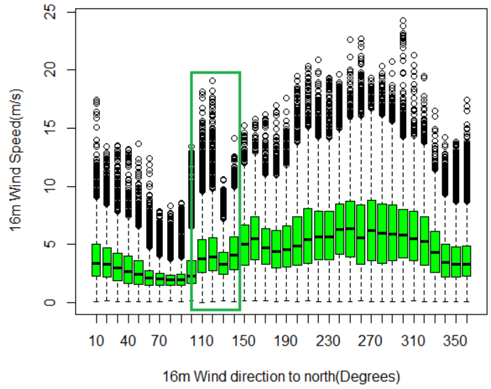

Figure 14: Boxplot of wind speed and direction at mast height 16m

7.3 Appendix 3 -Wind speed component relative to 118 degrees

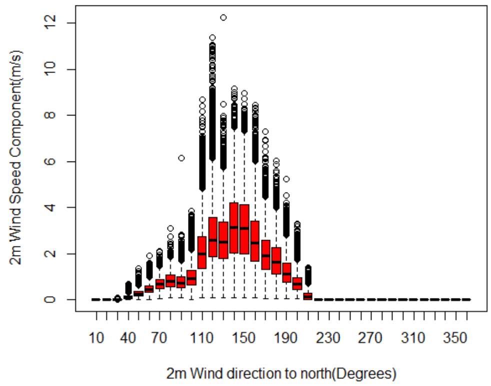

Figure 15: Boxplot of wind speed component and direction at mast height 2m

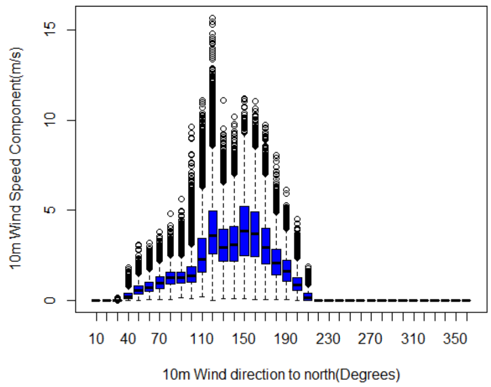

Figure 16: Boxplot of wind speed component and direction at mast height 10m

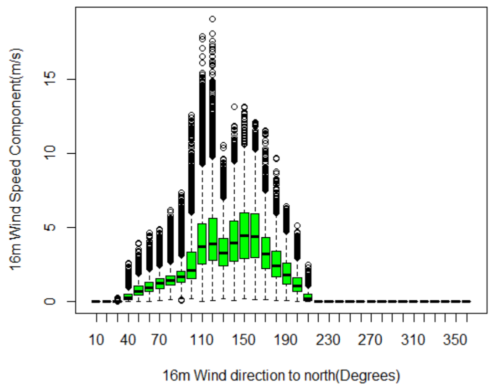

Figure 17: Boxplot of wind speed component and direction at mast height 16m

7.4 Appendix 4 - Mean residual life plots for wind speeds

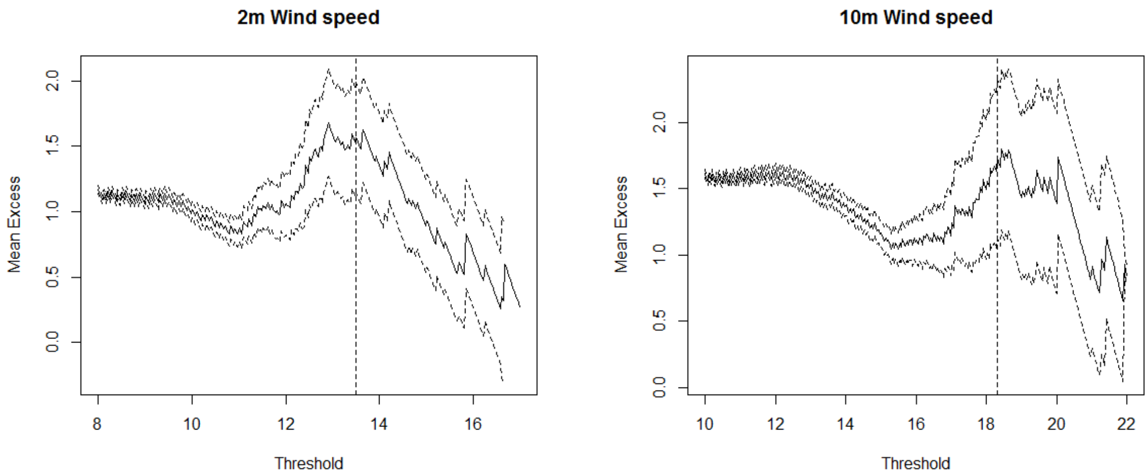

Figure 18: MRL plots of 2m and 10m wind speed respectively

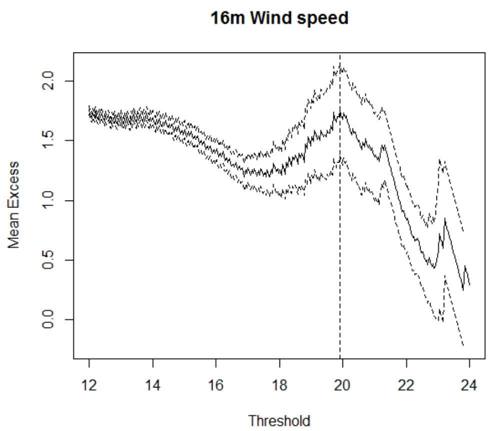

Figure 19: MRL Plot of 16m wind speed

7.5 Appendix 5 - Parameter Stability Plots for wind speeds

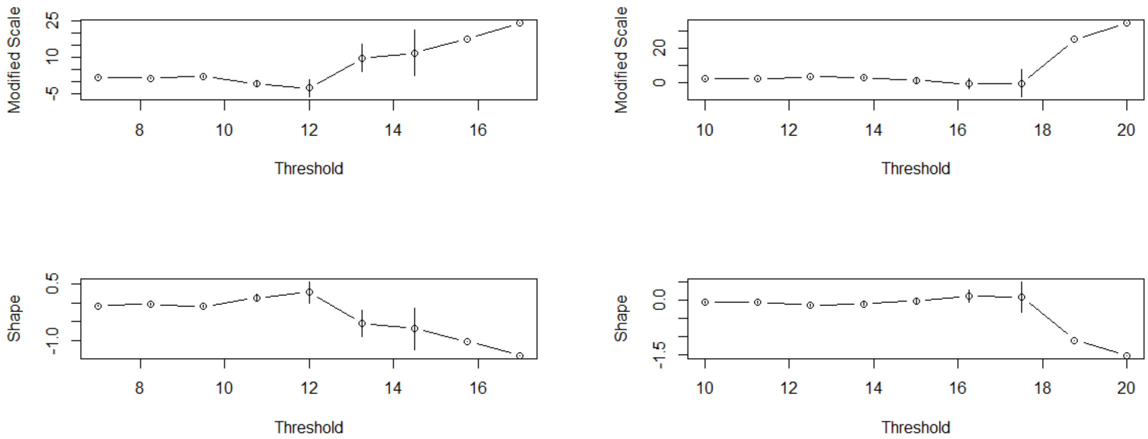

Figure 20: Parameter Stability plots of 2m and 10m wind speed respectively

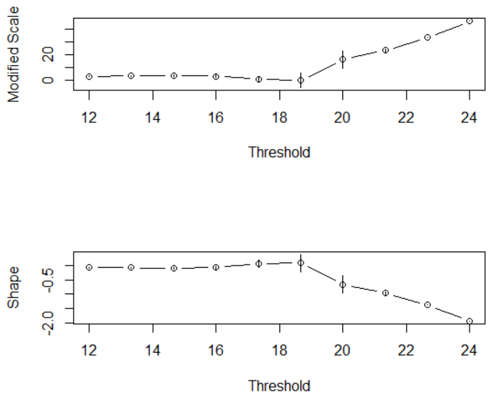

Figure 21: Parameter Stability Plots of 16m wind speed

7.6 Appendix 6 - Mean residual life plots for declustered wind speeds

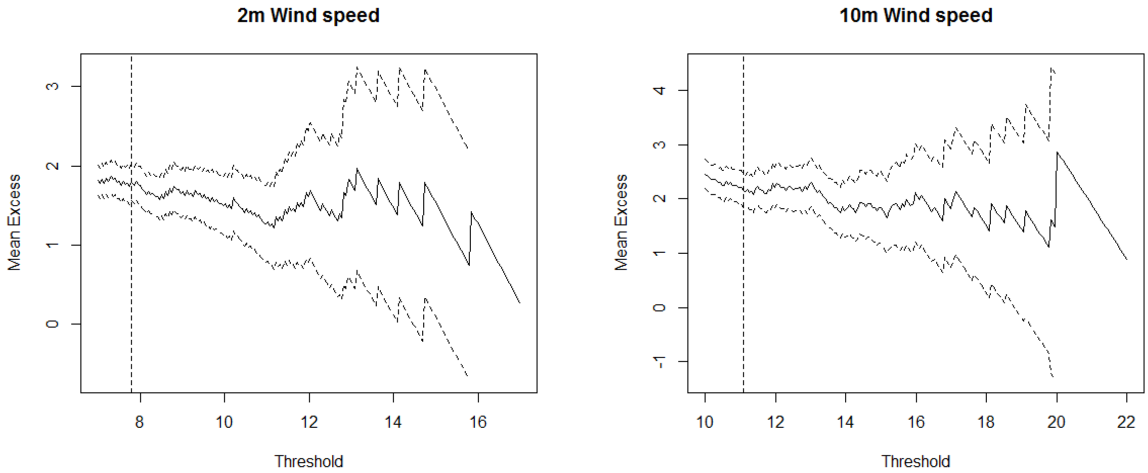

Figure 22: MRL plots of 2m and 10m declustered wind speeds respectively

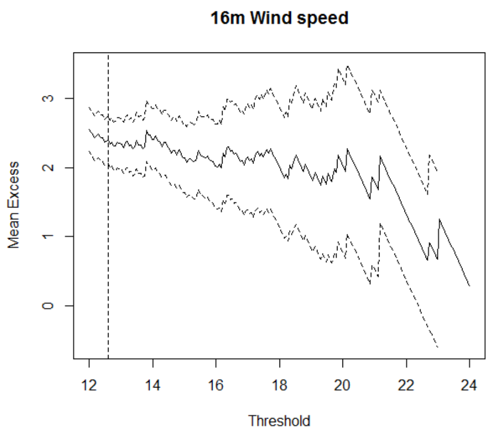

Figure 23: MRL plot of 16m declustered wind speed

7.7 Appendix 7 - Diagnostic plots for declustered 2m wind speed model

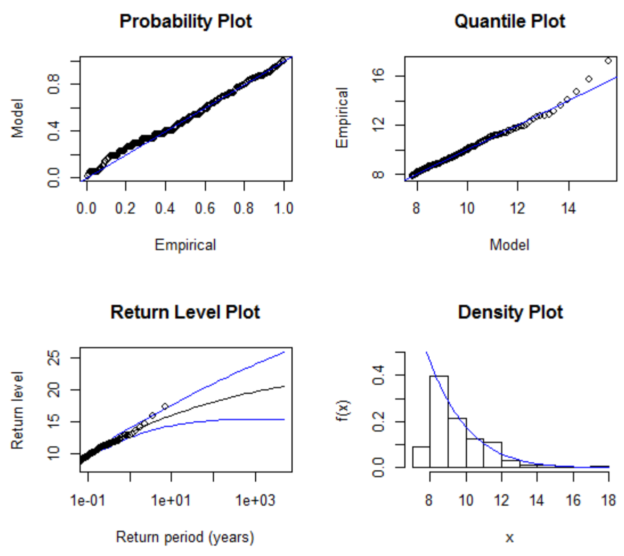

Figure 24: Diagnostic plots for declustered 2m wind speed model

7.8 Appendix 8 - Diagnostic plots for declustered 10m wind speed model

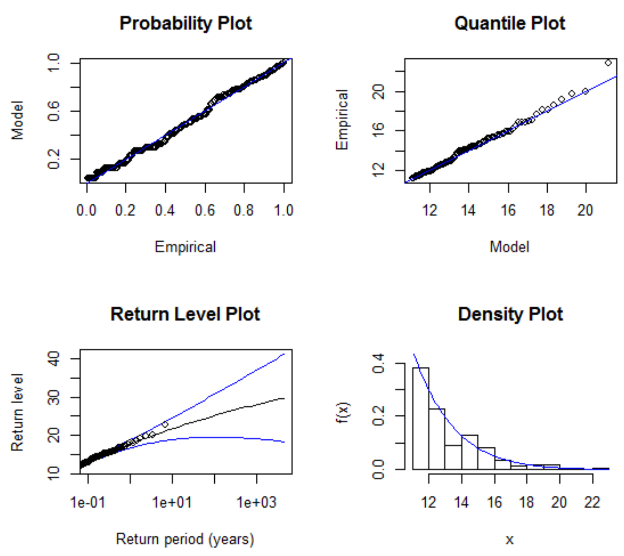

Figure 25: Diagnostic plots for declustered 10m wind speed model

7.9 Appendix 9 - Diagnostic plots for declustered 16m wind speed model

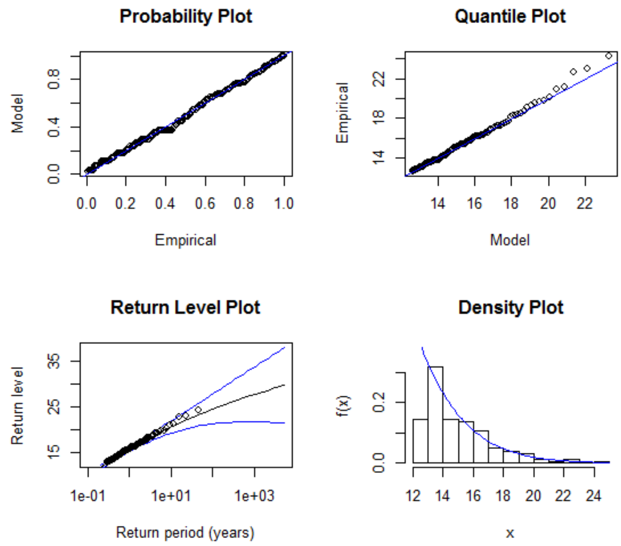

Figure 26: Diagnostic plots for declustered 16m wind speed model

7.10 Appendix 10 - Diagnostic plots for declustered 2m wind speed model(lower threshold)

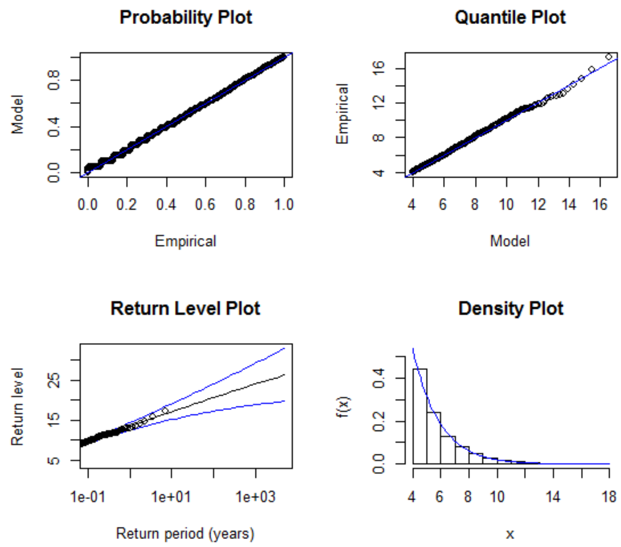

Figure 27: Diagnostic plots for declustered 2m wind speed model(lower threshold)

7.11 Appendix 11 - Diagnostic plots for declustered 2m wind speed model(lower threshold)

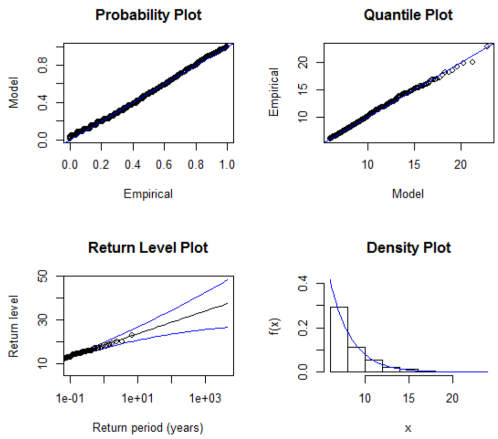

Figure 28: Diagnostic plots for declustered 10m wind speed model(lower threshold)

7.12 Appendix 12 - Diagnostic plots for declustered 16m wind speed model(lower threshold)

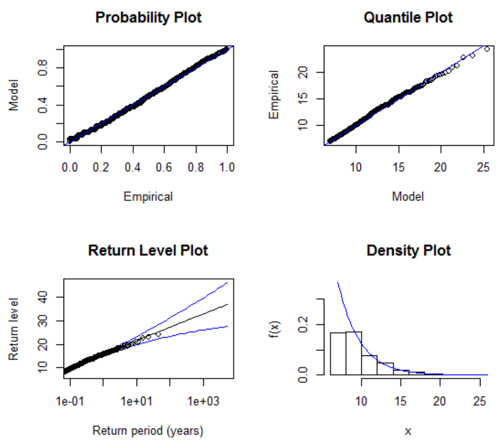

Figure 29: Diagnostic plots for declustered 16m wind speed model(lower threshold)

7.13 Appendix 13 - Mean residual life plots for temperature

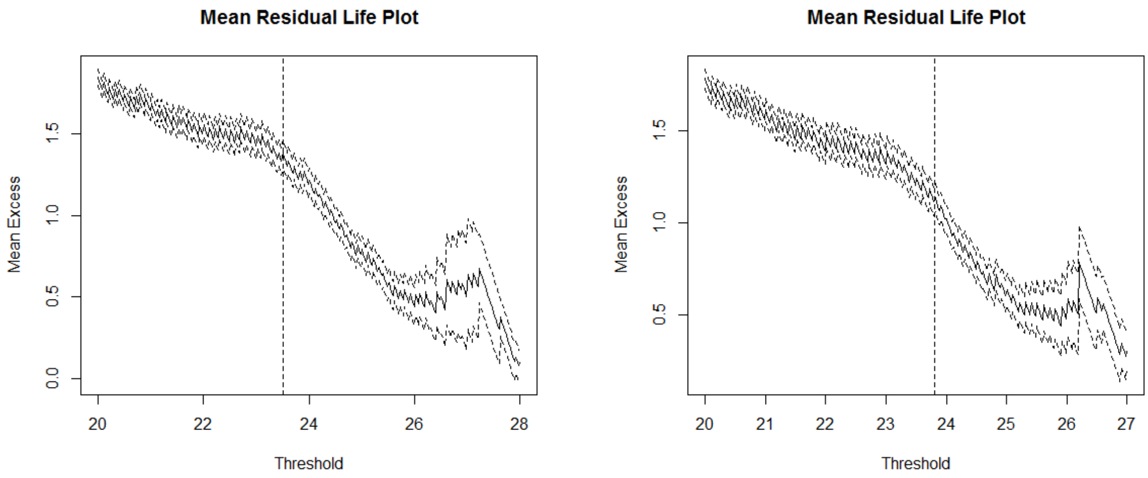

Figure 30: MRL plots of 2m and 10m temperature respectively

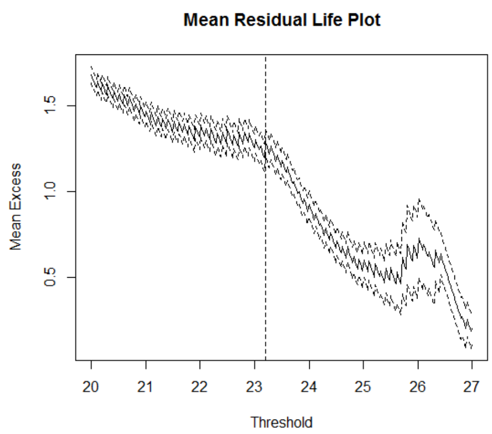

Figure 31: MRL Plot of 16m temperature

7.14 Appendix 14 -Parameter stability plots for temperature

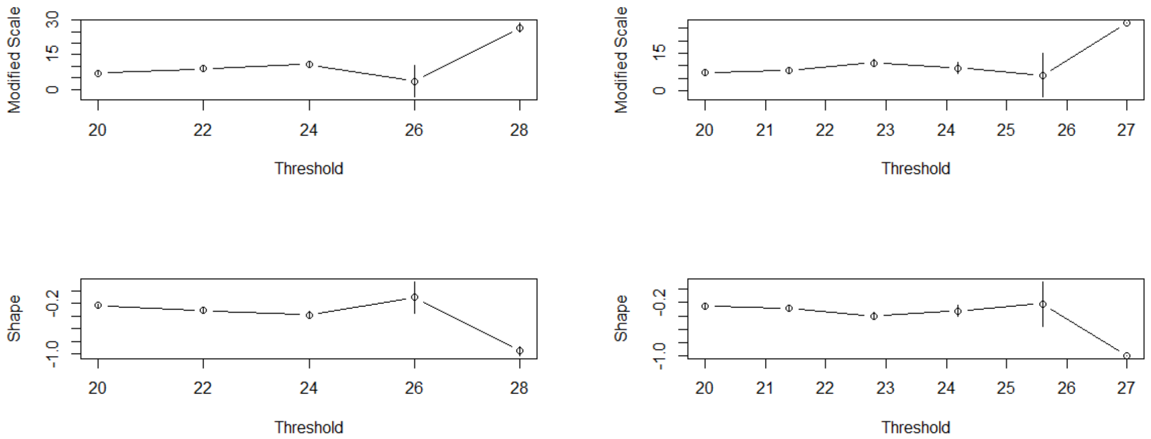

Figure 32: Parameter stability plots of 2m and 10m temperature respectively

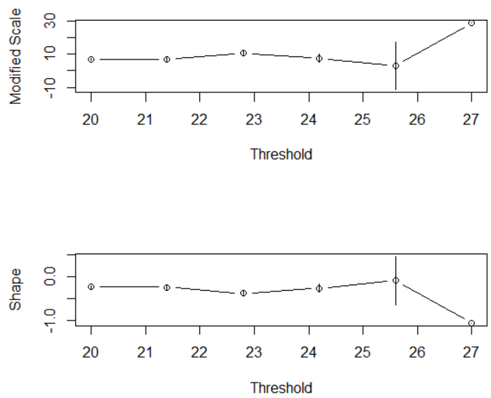

Figure 33: Parameter stability plot of 16m temperature

7.15 Appendix 15 - Fitted model of 2m temperature

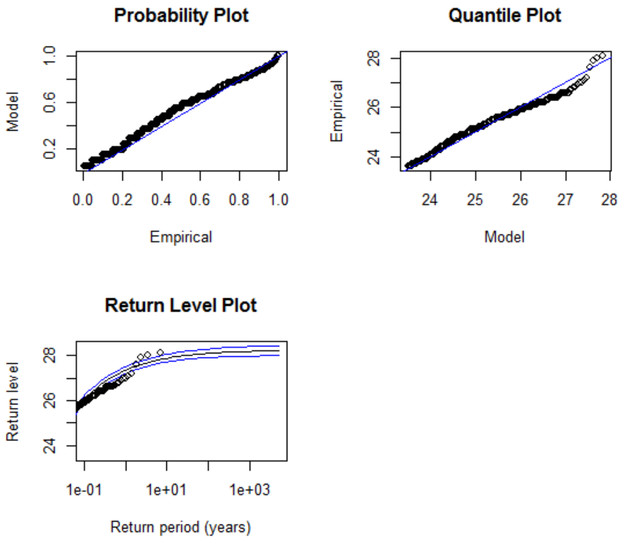

Figure 34: Diagnostics of 2m temperature model

7.16 Appendix 16 - Fitted model of 10m temperature

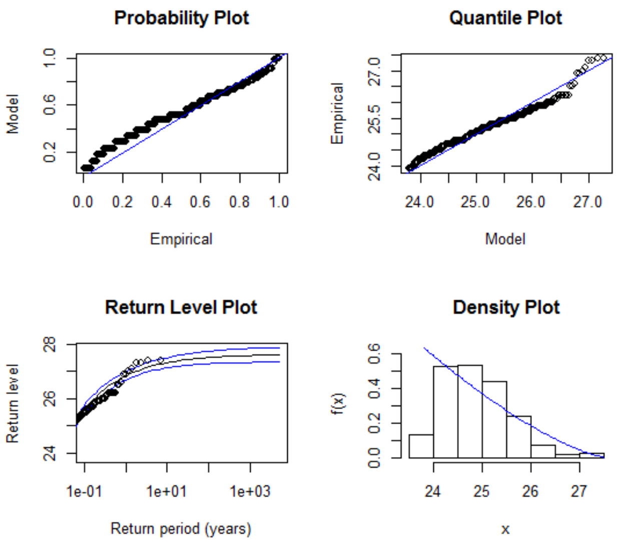

Figure 35: Diagnostics of 10m temperature model

7.17 Appendix 17 - Fitted model of 16m temperature

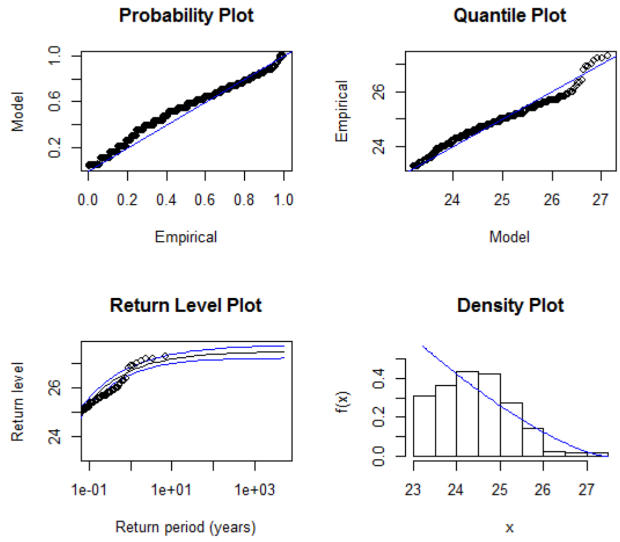

Figure 36: Diagnostics of 16m temperature model

7.18 Appendix 18 - Mean residual life plots of daily temperatures

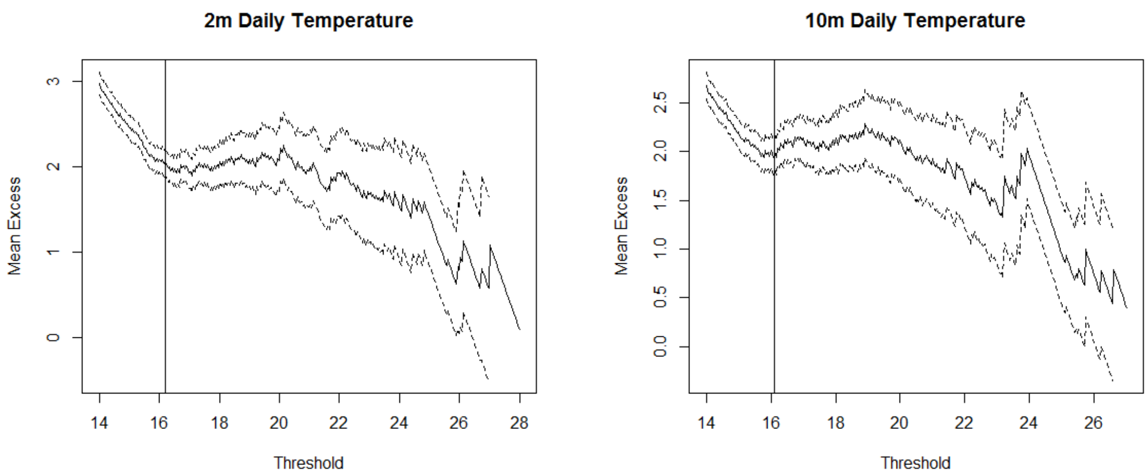

Figure 37: Parameter stability plots of 2m and 10m daily temperatures respectively

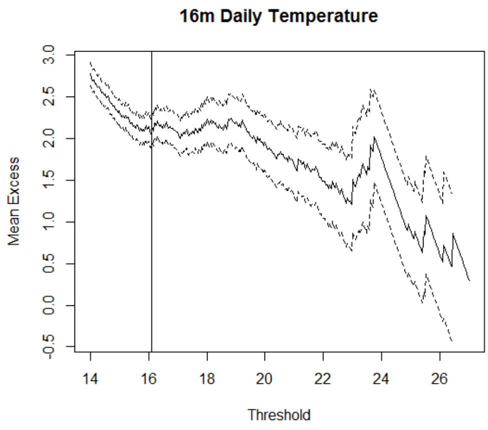

Figure 38: Mean residual life plot of 16m daily temperature

7.19 Appendix 19 - Diagnostic Plots of 2m daily temperatures

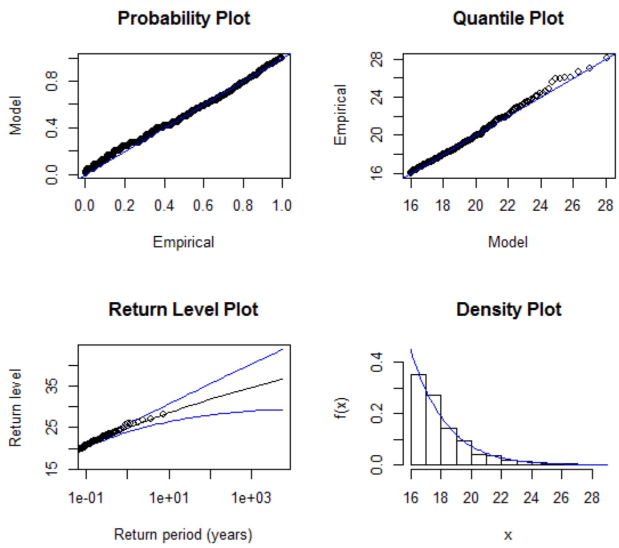

Figure 39: Diagnostic Plots of 2m daily temperatures

7.20 Appendix 20 - Diagnostic Plots of 10m daily temperatures

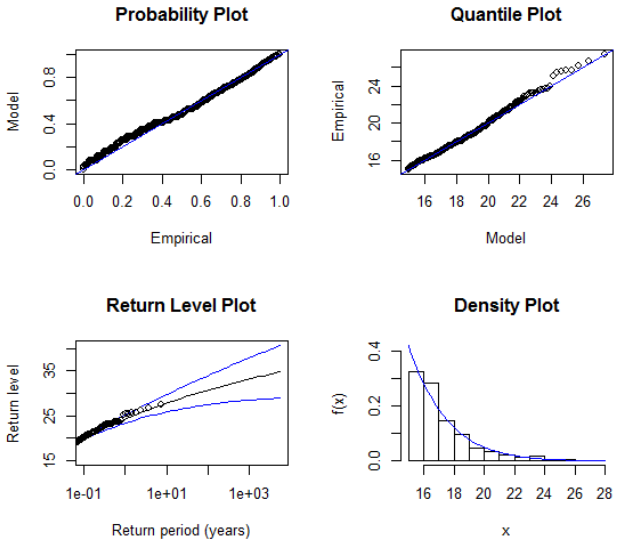

Figure 40: Diagnostic Plots of 10m daily temperatures

7.21 Appendix 21 - Diagnostic Plots of 16m daily temperatures

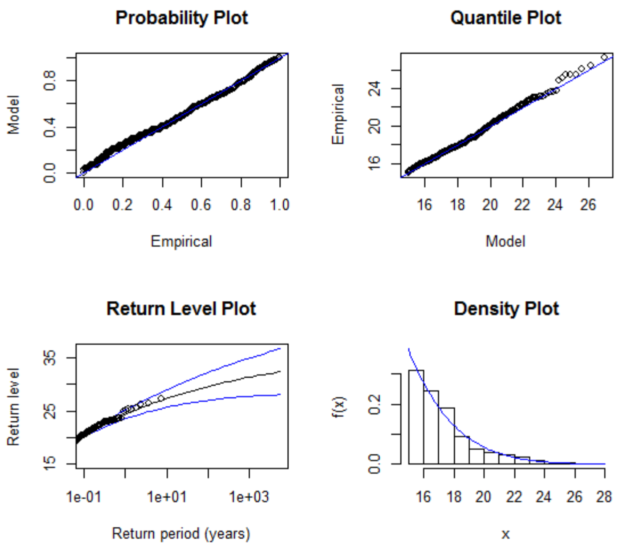

Figure 41: Diagnostic Plots of 16m daily temperatures

7.22 Appendix 22 - Mean residual life plots of wind speed component moving averages

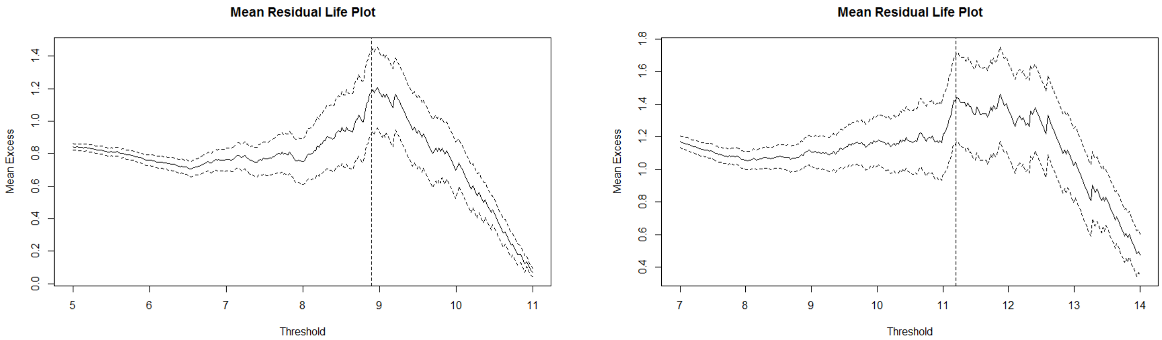

Figure 42: MRL plots of 2m and 10m moving average(one hour) wind speeds respectively

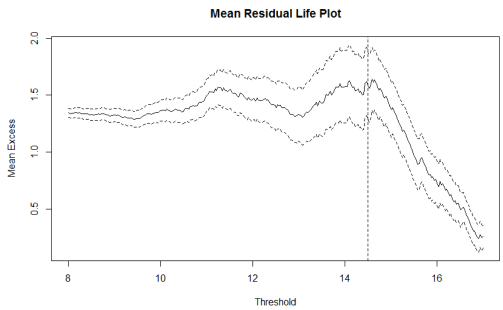

Figure 43: 16m Moving Average(one hour)

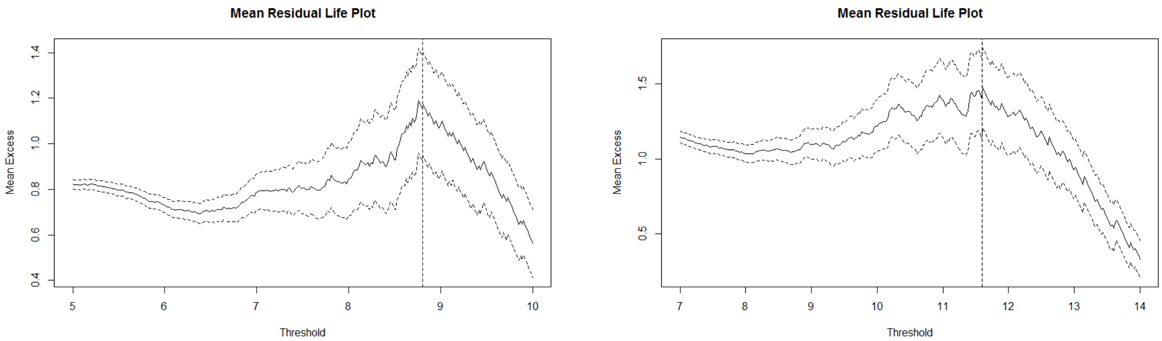

Figure 44: MRL plots of 2m and 10m moving average(two hour) wind speeds respectively

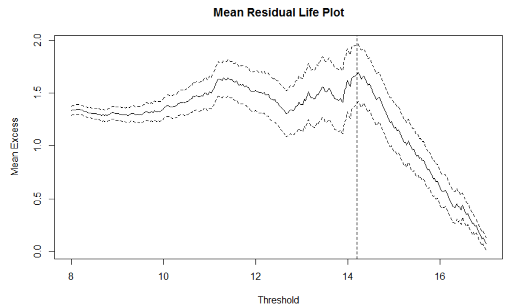

Figure 45: 16m Moving Average(two hour)

7.23 Appendix 23 - Mean residual life plots of wind speed component moving averages after declustering

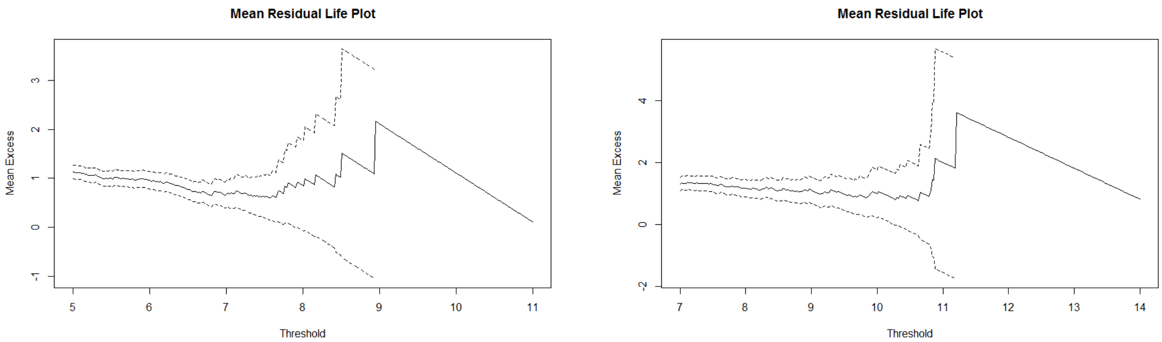

Figure 46: MRL plots of 2m and 10m moving average(one hour) wind speeds respectively after declustering

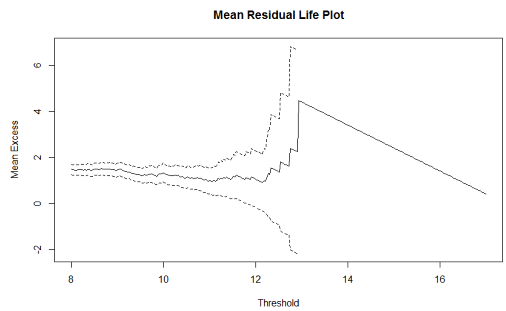

Figure 47: 16m Moving Average(one hour)

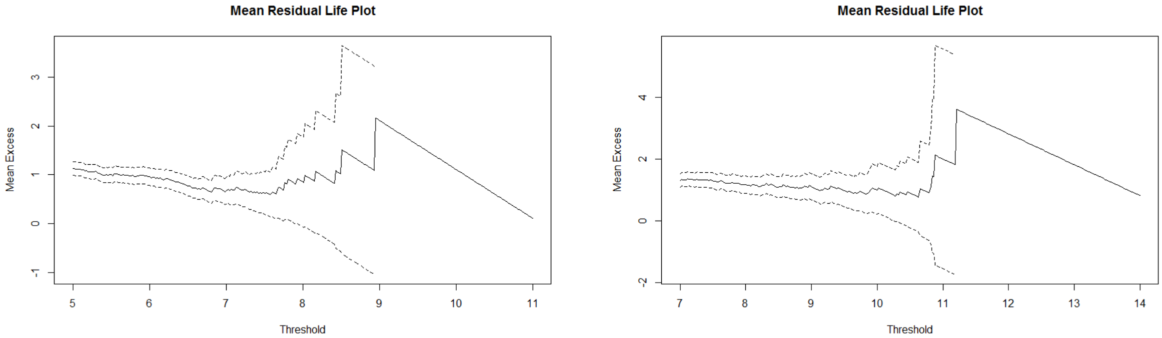

Figure 48: MRL plots of 2m and 10m moving average(two hour) wind speeds respectively after declustering

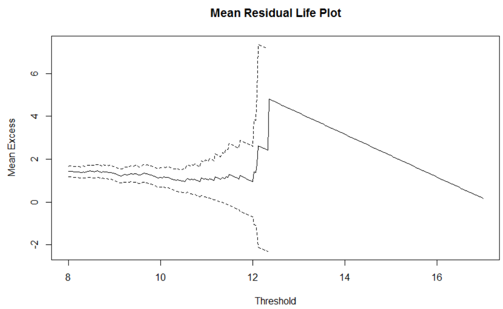

Figure 49: 16m Moving Average(two hour)

7.24 Appendix 24 - Diagnostics for the wind component moving average models

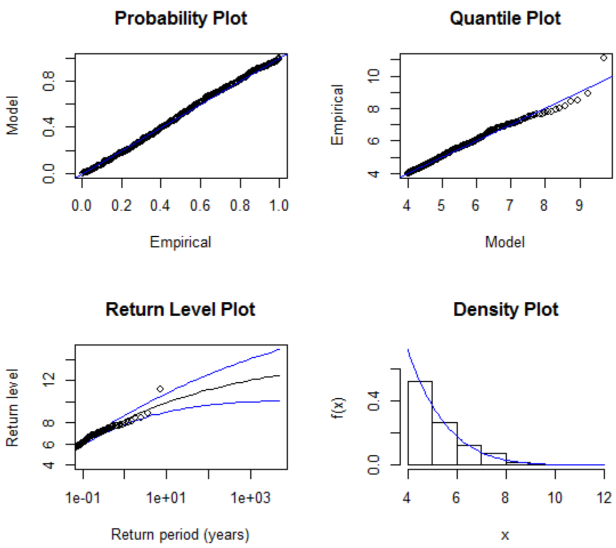

Figure 50: 2m Moving Average(one hour)

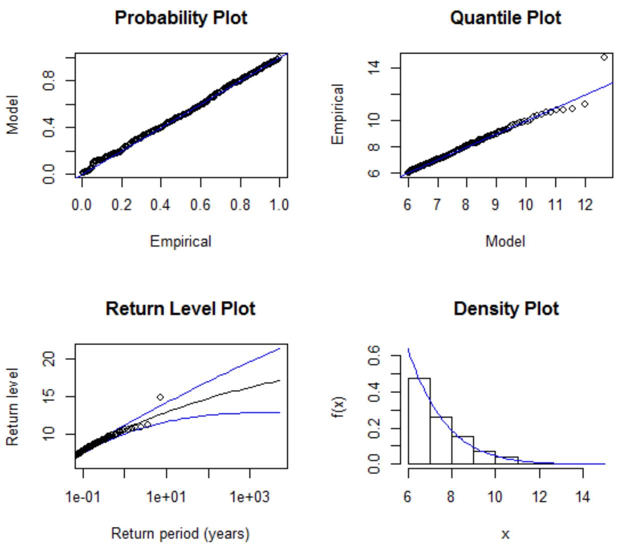

Figure 51: 10m Moving Average(one hour)

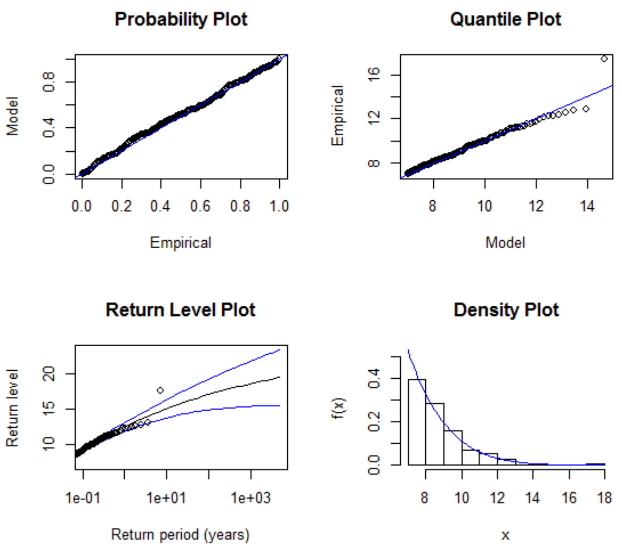

Figure 52: 16m Moving Average(one hour)

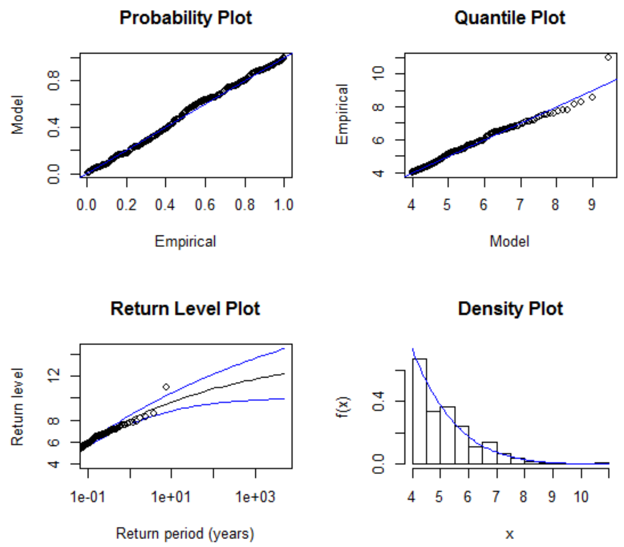

Figure 53: 2m Moving Average(two hour)

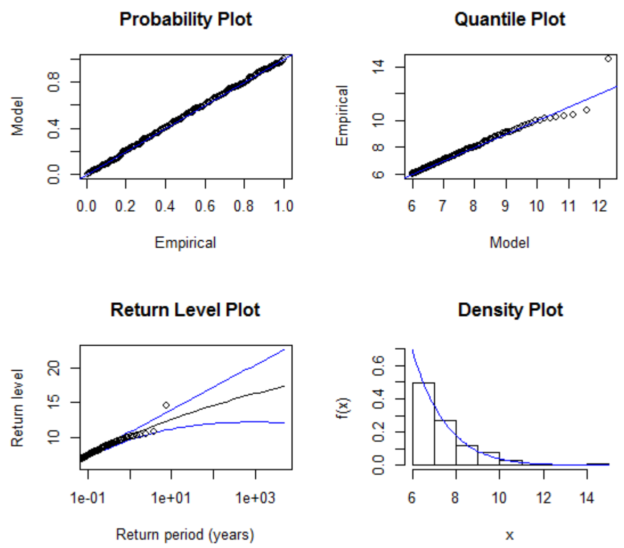

Figure 54: 10m Moving Average(two hour)

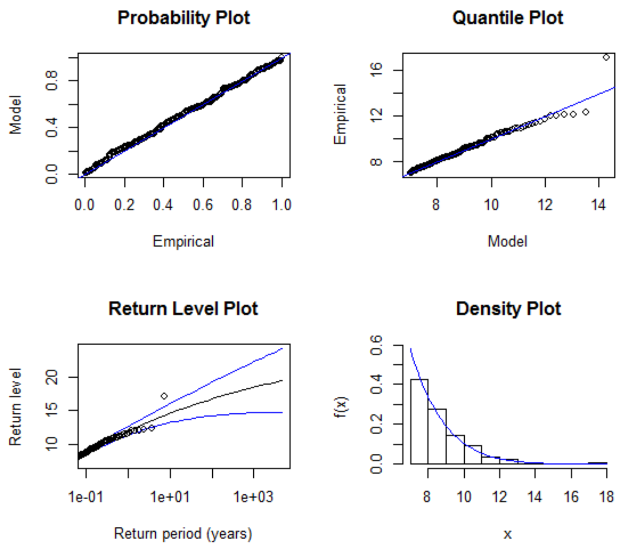

Figure 55: 16m Moving Average(two hour)