SafeWind

Collaborative project funded by the European Commission under the 7th Framework Program, Theme 2007-2.3.2: Energy

"Multi-scale data assimilation, advanced wind modelling & forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration"

Grant Agreement N°: 213740

# DELIVERABLE DP-7.6

# "Description and evaluation of a new dynamical risk analysis methodology"

| | |
|---|---|
| DOCUMENT TYPE | Deliverable |
| DOCUMENT NAME: | swind.deliverable_Dp-7.6_Dynamical_Risk_v1.3.docx |
| VERSION: | V1.3(*) |
| DATE: | 2012.08.25 |
| CLASSIFICATION: | R0: General public |
| STATUS: | Approved/released |

**Abstract:** This Deliverable of SafeWind project presents new methods for enabling the effective quantification of risk for the wind power industry. It focuses primarily on the challenges stemming from the limited availability of wind speed time series of sufficiently long duration. As a solution, the report investigates the use of reanalysis data from atmospheric dynamical models as a proxy for the short actual wind speed records that are typically available. By developing an innovative approach for calibrating the reanalysis data, it is shown that superior estimates of the 50-year return period may be achieved using the reanalysis data. Based on 45 years of actual and reanalysis data collected at Schiphol airport, it is shown that this new approach is superior to using the actual data in all situations where less than twenty years of actual data are available. In addition, a method for dynamical risk analysis employing quantile regression is developed to provide a semi-parametric description of the tail of the distribution associated with extreme events. This approach deals with two of the shortcomings of traditional extreme value theory: (i) the selection of an arbitrary threshold for defining what constitutes an extreme and (ii) the assumption of independence between extremes. This is a novel prospective approach to risk management and it is shown, using an actual wind speed time series that this is more accurate than the static approach. Using a comparison of the performance for the 95% quantile, the dynamical risk estimation is shown to improve predictability by more than 26%.

| AUTHORS[1], REVIEWERS | | | | |
|---|---|---|---|---|
| MAIN AUTHOR/EDITOR: | P. E. McSharry | | | |
| AFFILIATION: | University of Oxford | | | |
| ADDRESS: | Smith School of Enterprise and the Environment, 75 George Street, Oxford OX1 2BQ, UK. | | | |
| TEL.: | +44 1865 614943 | | | |
| EMAIL: | patrick.mcsharry@smithschool.ox.ac.uk | | | |
| FURTHER AUTHORS: | G. Anastasiades | | | |
| PEER REVIEWERS: | X.G. Larsen (DTU), J. Sanz Rodrigo (CENER, WP-Leader) | | | |
| REVIEW APPROVAL: | Approved : | X | Rejected (improve as indicated below) : | |
| SUGGESTED IMPROVEMENTS: | *For a long list of remarks make reference to another document* | | | |
| APPROVER: | | | | |

| VERSION HISTORY | | | |
|---|---|---|---|
| VERSION[2]: | DATE: | COMMENTS, CHANGES, STATUS: | PERSON(S): |
| 1.0 | 13.03.2012 | Draft for Comments | P.E. MacSharry |
| 2.0 | 27.07.2012 | Final version for review | P.E. MacSharry |
| 3.0 | 25.08.2012 | Final version reviewed | J. Sanz Rodrigo |
| | | | |
| | | | |

| STATUS, CONFIDENTIALITY, ACCESSIBILITY | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| STATUS: | | | | CONFIDENTIALITY: | | | | ACCESSIBILITY: | |
| **S0** | Approved/Released | X | | **R0** | General public | **X** | | Private web site | |
| **S1** | Reviewed | | | **R1** | Restricted to project members | | | Public web site | X |
| **S2** | Pending for review | | | **R2** | Restricted to European Commission | | | Paper copy | |
| **S3** | Draft for comments | | | **R3** | Restricted to WP members + PL | | | | |
| **S4** | Under preparation | | | **R4** | Restricted to Task members +WPL+PL | | | | |

**PL:** *Project leader* **WPL:** *Work package leader* **TL:** *Task leader*

[1] **The authors of this document are solely responsible for its content, which does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of data appearing therein.**

[2] **VERSION NAMING : V0.x** draft before peer-review approval, **V1.0** at the approval, **V1.x** minor revisions, **V2.0** major revision

# Contents

# 1. Introduction

The amount of wind power being integrated into European power systems has risen rapidly over the recent decades. The wind industry has had an average annual growth of 15.6% over the last 17 years (1995-2011). In 2011, 9,616 MW of wind energy capacity was installed in the EU, making a total of 93,957 MW, which is sufficient to supply 6.3% of the EU's electricity. This figure represents 21.4% of new power capacity showing that wind energy continues to be a popular source of energy.

As demand for wind power grows, there is an increasing need to establish appropriate quality controls and adequate risk management. Selecting sites for wind farms presents a challenge in finding a balance between having sufficient wind to operate efficiently but not too much that turbines are exposed to a high risk of being damaged. Furthermore, there are many other factors such as distance from substations, location on the transmission grid, planning permission, impact on the natural environment and public acceptance.

In this report, the focus is on the specific challenge of assessing the risk that turbines in a wind farm could be damaged during a severe windstorm. A measure of this risk is an important factor when a wind farm operator considers a potential site and when financial support is being sought. The biggest challenge to performing this risk assessment is data availability. Unfortunately, potential sites for wind farms are typically in locations where there has been little reason to measure and record wind speeds. For this reason, there is a substantial technical challenge in undertaking the risk assessment using only a short record of historical wind speed data.

This report focuses on the use of information generated by numerical weather prediction (NWP) models to provide adequate risk analysis for the wind power industry. One possibility is to employ reanalysis data to facilitate the risk assessment exercise. Reanalysis data records are produced using fixed, modern versions of the data assimilation systems developed for numerical weather prediction. Reanalyses are more suitable than operational analyses for use in studies of long-term variability in climate. Reanalysis products are used increasingly in many fields that require an observational record of the state of either the atmosphere or its underlying land and ocean surfaces.

The challenge of using reanalysis data for risk analyses is that it is sampled every six hours and the spatial resolution is relatively coarse in comparison to the wind speed records available at wind farm sites. We develop a new statistical calibration technique that allows us to utilize the full reanalysis record in order to obtain more accurate estimates. Using an appropriate bootstrap analysis based on surrogate time series, we demonstrate that our calibrated estimates from the reanalysis data is superior when less than twenty years of historical wind speed data is available at the potential site.

We also develop a dynamic risk estimation approach that is particularly appropriate for wind speed time series. Our approach first uses non-parametric kernel regression to deseasonalise the time series by identifying deterministic components for the intra-annual seasonality in the level and variability. It then applies a semi-parametric technique to obtain conditional estimates of a specific quantile using an autoregressive model.

The report is structured as follows. In Section 2, we first describe the datasets employed in the analyses as these motivate the techniques that are developed. Section 3 summarises the block maxima technique that is often used for measuring return periods in risk analysis, and explains how GEV distributions are used to quantify the extreme events. Section 4 contains the main methodological contributions of the report. It describes the quantile calibration approach and the wind speed surrogates that will be used to demonstrate the utility of this approach. Section 5 introduces the technique for quantile forecasting which provide a dynamic risk assessment for providing prospective risk management. Section 6 demonstrates the advantages of the new techniques using the test case of Schiphol airport. Section 7 concludes the report and offers ideas for further investigation.

# 2. Meteorological data

## 2.1 Actual wind speed observations

The actual wind speed observations used in this report were collected at Schiphol airport in the Netherlands. This data is available from the KNMI Hydra project (http://www.knmi.nl/samenw/hydra/). This extensive time series is approximately 45 years long and ranges from 01-Sep-1957 to 31-Aug-2002. The time series is sample at hourly intervals. It has a total of 23 missing values, which were imputed using linear interpolation. As the percentage of missing values is very small (<0.0006%), the effect of the linear interpolation is practically negligible. The full time series record is shown in Figure 1.

## 2.2 Reanalysis data

Information generated by numerical weather prediction (NWP) models may be of considerable value in providing adequate risk analysis for the wind power industry. The benefits of using NWP data for estimating returns periods of extreme wind speeds are the following:

- Temporal coverage. NWP models are capable of generating long historical records. Reanalysis data can reach 58 years (at low spatial resolution), while operational forecasting models available for the project are around 5 years (at higher spatial resolution). It is worth noticing that before 1979 it is the so-called pre-satellite era, where the observation was greatly limited compared to later periods. Hence, using the whole 58 years might result to some issues regarding discontinuity in the data assimilation.
- Spatial coverage. While the measurement stations are only representative of a limited surrounding area, NWP cover wide extensions (reanalysis cover the entire globe and operational prediction models used in the framework of ANEMOS cover at least Europe).
- Consistent time series. The particular NWP is fixed over the entire model run that generates the reanalysis data and this consistency is important for obtaining robust risk analysis estimates.

In the analysis presented in this report, we use the ECMWF ERA-40 reanalysis data (http://www.ecmwf.int/research/era/). The reanalysis time series is sampled at six-hourly intervals, implying that it has a coarser temporal resolution than the hourly values in the actual wind speed time series. The graph of the reanalysis data (Figure 2) and the summary statistics (Table 1) show that reanalysis data have a similar distributional pattern (shape of distributions) to the actual data (Figure 1) but a systematic underestimation may also be noted.

From the summary statistics in Table 1, we see that the mean of the reanalysis data is 4.641 m/s whereas that of the actual data is 5.124 m/s. This implies that the reanalysis underestimates the mean of the actual data by more than 9%. This underestimation is confirmed by the medians. A comparison of the interquartile ranges (IQR) and the variances show that the actual series is also more volatile than the reanalysis. The shapes of the two distributions (described by skewness and kurtosis) are roughly the same.

**Table 1:** *Summary statistics for the actual and reanalysis wind time series.*

| Variable | Actual | Reanalysis |
|---|---|---|
| Mean | 5.124 | 4.641 |
| Variance | 8.629 | 6.165 |
| Median | 4.600 | 4.260 |
| IQR | 3.800 | 3.317 |
| Skewness | 0.898 | 0.876 |
| Kurtosis | 1.101 | 0.842 |

Finally, we would like to address any issue resulting from the fact that we use reanalysis data from both the pre-satellite and satellite eras. We split the reanalysis data set into two data sets, one before 1979 (pre-satellite era) and one after 1979 (satellite era), and use a Kolmogorov-Smirnov test to assess whether or not these two sample sets are drawn from the same continuous distribution. The p-value for the test is smaller than 0.001 and hence we conclude that there exists strong statistical evidence that these two data sets are not drawn from the same probability distribution. Nevertheless, the analysis of this paper uses only data from the upper tail of the whole distribution and hence it makes sense to compare only the upper tails of the two distributions. We use again a Kolmogorov-Smirnov test on the data that exceed the (rather low) threshold speed of 11 m/s, and assess whether or not the two data sets are drawn from the same continuous distribution. The p-value of the test is now equal to 0.288 and hence the evidence that the two sample sets are not drawn from the same distribution is very week. Hence, we conclude that the analysis carried out in this paper will not be (greatly) affected by using reanalysis data taken from both the pre-satellite and satellite eras.

The main drawback of any methodology that uses reanalysis data is the accuracy of the meteorological model and the coarse spatial and temporal resolution. These challenges provide the motivation behind our development of a new statistical calibration technique, which is outlined in Section 4.

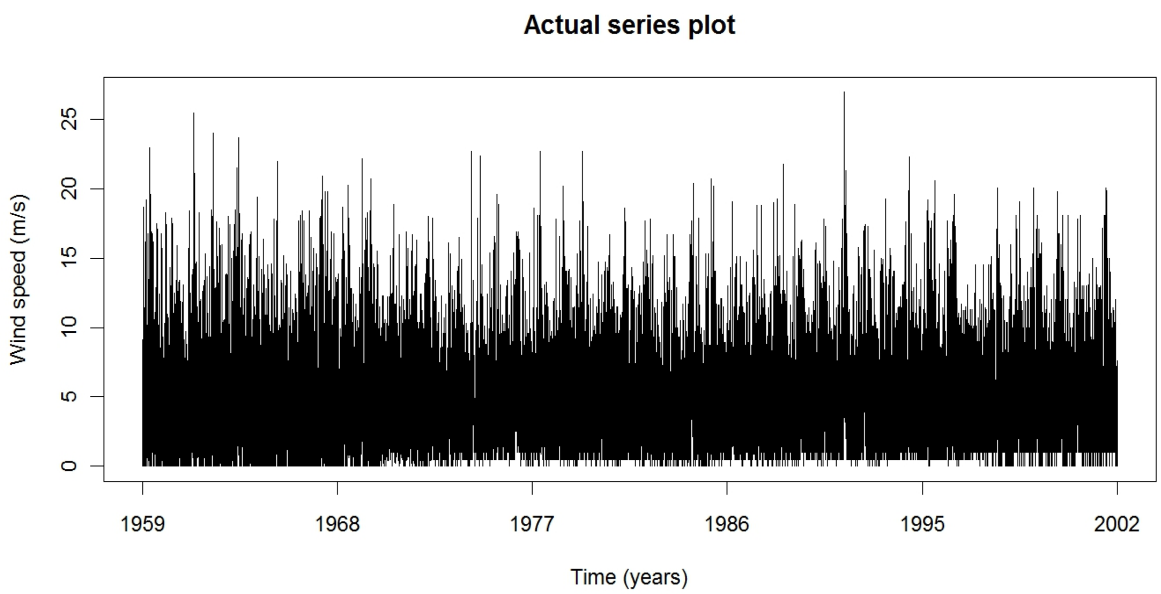


**Figure 1: Actual wind speed time series for Schiphol airport.**

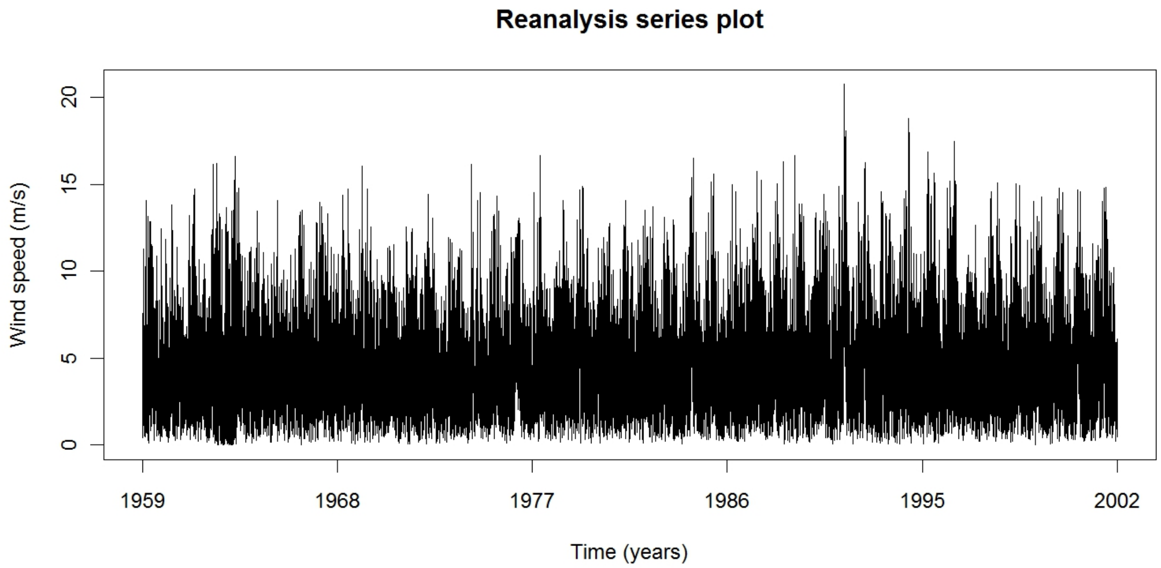


**Figure 2: ECMWF reanalysis wind speed time series for Schiphol airport.**

# 3. Extreme values

Extreme value analysis has found widespread use as a means of estimating the likelihood of extremes but it relies on a number of assumptions that are rarely met in practice. The essential idea is to fit the tail of a sample empirical cumulative distribution function. One approach relies on approximating a distribution obtained from block maxima time series, such as extracting the annual maxima. The extreme value distributions, used for fitting to the data extracted from the block maxima are the limiting distributions for the minimum or the maximum of a very large collection of independent random variables from the same arbitrary distribution.

In the following, we explain how classical extreme value theory is used to estimate a T-year return level using the block maxima technique. We define the T-year return level as the extreme wind speed value that is exceeded, on average, once a year with a probability of (100/T)%. Moreover, we denote the T-year return value of a k-year long wind speed series, of temporal resolution τ time units, by $V_T^\tau(k)$.

## 3.1 GEV distributions

Let us assume that we have collected a sufficiently long record of independent and identically distributed (IID) random variables. Classical extreme value theory describes how the maxima of their samples (of sufficiently large size) can be fitted to three asymptotic extreme value distributions (Fisher and Tippett, 1928). These three distributions can be combined into a single one, called the Generalized Extreme Value (GEV) distribution (Von Mises, 1936), with cumulative distribution function of the form:

$$F(x,\mu,\sigma,\xi)=\begin{cases}\exp\left[-\left(1-\xi\dfrac{x-\mu}{\sigma}\right)^{\frac{1}{\xi}}\right], & \xi\neq 0\\[2ex] \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], & \xi=0\end{cases},$$

(1)

where μ, σ and ξ are the location, scale and shape parameters respectively. When ξ = 0, the GEV reduces to the the GEV reduces to the Gumbel distribution, also known as the Type I GEV, or Fisher-

Tippet Type I distribution. For $\xi < 0$ we get the Type II GEV (or Frechet distribution) and when $\xi > 0$ the Type III GEV (Weibull family of distributions).

In order to be more rigorous we can reformulate the above statement by firstly defining the random variable $M_n = \max\{X_1,\ldots, X_n\}$, where the $X_i$ with i=1,…,n is a sequence of independent random variables having a common distribution. $M_n$ represents the maximum of a process measured on a regular time scale, over n time units of observation. For example, if n is the number of wind records in a year, then $M_n$ corresponds to the annual wind speed maximum. Then, the extremal types theorem states that if there exist a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $P[(M_n - b_n)/a_n \leq x] \rightarrow F(x)$ as $n \rightarrow 1$ for a non-degenerate distribution function F, then F is a member of the GEV family defined above, where $\{x: 1 + \xi (x - \mu)/ \sigma > 0\}$, $-\infty < \mu; \xi < +\infty$ and $\sigma > 0$. This theorem provides the basis of classical extreme value theory. Its proof is quite technical and is out of the scope of this paper (for more details see Leadbetter et al. (1983)).

## 3.2 Block maxima

The block maxima technique consists of splitting the data into independent blocks of equal length and then fitting the GEV distribution to the collection of maxima that result from each of the blocks. The implementation of this model for any given data set relies on the choice of block size. When the size of blocks is too small, the limit approximation in the extremal types theorem may be poor, leading to bias in the estimates of the three GEV parameters, μ, σ and ξ. In contrast, selecting a large block size causes the estimation variance to increase due to the fact that there are fewer block maxima, and hence fewer points available to fit the distribution. Therefore, the choice of the block size leads to a trade-off between estimation bias and variance, and hence extra care is needed when choosing this particular parameter.

According to Coles (2001), pragmatic considerations often lead to the adoption of blocks of one year length, for a sufficiently large data series and this is often referred to as the Annual Maxima Method. Cook (2001) suggests using at least 20 years of data for reliable results, or equivalently at least 20 extremes for analysis (Palutikof et al., 1999). Hence in our analysis we will ensure that the number of extremes used for estimation is no less than 20, with a maximum block size of one year. The block size will depend on how many years of data are available for parameter estimation.

Lets denote the block maxima by $\{Z_i\}$ for i=1,…,m, and assume that these values are independent GEV distributed variables. Even if the $X_i$ form a dependent series (case not covered by the aforementioned theorem), the assumption of the $Z_i$ being independent and following a GEV distribution may still be reasonable (for more details see Coles (2001)).

## 3.3 Maximum likelihood estimation

In order to estimate the parameters of the GEV distribution, we use the Maximum Likelihood Estimation (MLE) method. If we denote by $z = (z_1,\ldots, z_m)$ the corresponding vector of realizations of the independent variables $\{Z_i\}$, then the log-likelihood of the GEV parameters is given by

$$\log L(z,\mu,\sigma,\xi)=\begin{cases} -m\log\sigma-\left(1+\dfrac{1}{\xi}\right)\sum_{i=1}^{m}\left[1+\xi\left(\dfrac{z_i-\mu}{\sigma}\right)\right]-\sum_{i=1}^{m}\left[1+\xi\left(\dfrac{z_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} & \xi\neq 0 \\ -m\log\sigma-\sum_{i=1}^{m}\left(\dfrac{z_i-\mu}{\sigma}\right)-\sum_{i=1}^{m}\exp\left[-\left(\dfrac{z_i-\mu}{\sigma}\right)\right] & \xi=0 \end{cases} \quad (2)$$

where for the case $\xi \neq 0$ we need the condition $1+ \xi[(z_i - \mu)/ \sigma]> 0$ for i=1,…, m. We maximise the likelihood with respect to the GEV parameters (μ, σ, ξ) using the Nelder and Mead (1965) optimization method. We denote the parameter values yielding this maximum as $(\hat{\mu},\hat{\sigma},\hat{\xi})$respectively. In order to check whether or not there is sufficient evidence to support the hypothesis that $\xi = 0$ (and hence use the second case of (2), we use a likelihood ratio test. We fit two distributions: one with $\xi = 0$ (Gumbel

model) and one which assumes $\xi \neq 0$. Then the test statistic, $D = -2\text{LogL}_{\xi=0} + 2\text{LogL}_{\xi\neq 0}$, is $\chi^2$ distributed with one degree of freedom. Small p-values of the test provide evidence against the null hypothesis that the Gumbel distribution is adequate to describe the underlying distribution of the sample maxima.

### 3.4 Return level estimation and goodness of fit

The T-year return level estimate, $X_T = \hat{z}_{l/T}$, from a series of m block maxima ($z_1,\ldots, z_m$) with a block length of l days is given by:

$$\hat{z}_{l/T} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\log\left(1 - \frac{l}{365.T}\right)\right\}^{-\hat{\xi}}\right] & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left[-\log\left(1 - \frac{l}{365.T}\right)\right] & \hat{\xi} = 0 \end{cases},$$

where $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$are the maximum likelihood estimates of the GEV parameters (μ, σ, ξ) given the available data. This is just the 1 – l/(365T) quantile of the distribution function (1). Note that when we use block sizes equal to one year (l = 365 days) we obtain the annual maxima method T-year return level.

It is very difficult to assess the validity of the extrapolation based on a GEV model. Nevertheless, we can check how good the fitted GEV model is (given the observed data) by using various exploratory plots, among which probability, quantile and return level plots (see for example Coles (2001), Perrin et al. (2006), Reiss and Thomas (2007)). For practical reasons (explained in the next section), we will not try to use any of these visual aids to assess the goodness of the GEV fit. Instead, we will only use the Anderson-Darling (AD) goodness of fit (GOF) test (Anderson and Darling, 1952), which tests whether there is evidence that a given sample of data did not arise from a given probability distribution. It belongs to the class of empirical distribution function (EDF) tests, since it compares the fitted distribution with the empirical one, and the model's parameters can be directly estimated form the sample data.

# 4. Calibration

### 4.1 Quantile calibration

Wind speed time series are often recorded with a specific discrete sampling interval, τ, so there is a gap of some hours between consecutive samples. Imagine we have two time series of wind speeds sampled over N years from the same location but with different discrete sampling intervals, and we would like to calculate the T-year return wind speed (T > N) using the block maxima technique.

Assume the first series has a temporal spacing between consecutive samples of $\tau_1$ time units, and the second series a temporal spacing between consecutive samples of $\tau_2$, where $\tau_2 > \tau_1$. If we represent the estimated T-year return level of the first series using only k consecutive years of data (where k = 1,…,N) by $V_1(k)$ and the corresponding estimated value of the second series by $V_1(k)$ then $V_1(k) > V_2(k)$. So, as the discrete sampling interval increases the estimated extreme wind (T-year return value) will get more underestimated (Larsen and Mann, 2006).

We propose a solution to this problem of underestimation, by assuming that $V_2(k)$ is underestimated by approximately a constant factor for all values of k, and hence the ratio $R(k) = V_1(k)/ V_2(k)$ is approximately constant. Moreover, assume we have access to the whole N years of the coarser sampled time series, but only to k years of the more frequently sampled series. If we define by $V_T$ the

theoretically most accurate T-year extreme wind we can estimate (given the data we have) for the given location, then

$$V_T \approx V_1(N) = \frac{V_1(N)}{V_2(N)} V_2(N) \approx \frac{V_1(k)}{V_2(k)} V_2(N).$$

Note that $V_1(k)$ and $V_2(k)$ were both estimated using the block maxima method and are therefore highly dependent on the goodness of fit of the GEV distribution and the independence of the underlying collection of maxima. For this reason, the estimate of this ratio R(k) is not robust when using the block maxima method. In order to reduce the uncertainty in the estimate of R(k), we propose using a non-parametric method to estimate this ratio. Recall the definition of the T-year return wind speed: it is the value that is exceeded, on average, once in a year with a probability of 100 (1/T)%. So if we have a time series, $(X_t)$, of annual maxima, then the T-year return level can also be estimated by finding the 1 – (1/T) quantile of its empirical distribution.

Therefore, from the two series sampled over a period of k years with temporal resolutions $\tau_1$ and $\tau_2$, and empirical cumulative distribution functions $F_{1,k}$ and $F_{2,k}$ (respectively), their approximate estimates of the T-year return level are given by $V_i(k) = F_i^{-1}(1- \tau_i/(24.365.T))$ for i = 1 and 2 where the sampling times $\tau_i$ are measured (for simplicity) in hours. These two empirically estimated T-year return values will be neither robust nor unbiased, but their ratio,

$$\hat{R}(k) = \frac{F_{1,k}^{-1}\left(1 - \frac{\tau_1}{24.365.T}\right)}{F_{2,k}^{-1}\left(1 - \frac{\tau_2}{24.365.T}\right)},$$

should be robust because we removed the parametric uncertainty associated with the estimation of $V_i(k)$ by fitting appropriate GEV distributions to their underlying series.

## 4.2 Wind speed surrogates

Having only a single long (actual) wind speed series of $\tau_1$ time units of temporal spacing between consecutive samples, we can produce only a single trustworthy T-year return value by applying the annual maxima method to the whole length of the time series. In order to test the validity of the block maxima method it will be worthwhile to produce a number of sufficiently long wind speed series, that have the same characteristics as the original one: specifically the same autocorrelation (seasonality of our actual data) and the same unconditional distribution.

Using the Amplitude Adjusted Fourier Transform (AAFT) algorithm (Theiler et al. 1992) we can produce a number of surrogate series with the same spectrum as the original series and simultaneously preserving the amplitude of its distribution. Let us denote the original series by $(X_t)$ for t=1,…,M. The production of the surrogate series is achieved in the following four steps, as explained by Theiler et al. (1992).

Firstly, we create a Gaussian time series, $Y_t$ (t=1,…,M), where each element is randomly generated from a Gaussian random number generator. Secondly, we re-order the time sequence of $Y_t$ so the ranks of the two series agree and hence $Y_t$ will follow $X_t$ in time, but also have a Gaussian distribution. Third step is to create a surrogate series, $Y_t$' of $Y_t$ using the Fourier Transform (FT) algorithm. This surrogate is computed by randomizing the phases of the Fourier coefficients of $Y_t$. Finally the amplitude adjusted surrogate series, $X_t$' is created by re-ordering the series $X_t$ so its ranks agree with the ranks of the series $Y_t$'. Then the time-re-ordered series, $X_t$', provides a surrogate of the original series that preserves both the power spectrum and the amplitude of the original distribution.

Now assume, as in Section 4.1, that we have two series: a series of actual wind speed data and a proxy series (reanalysis), with both spanning the same time range but with different sampling periods. Assume the first series has a temporal spacing between consecutive samples of $\tau_1$ time units, and the second series a temporal spacing between consecutive samples of $\tau_2$, where $\tau_2 > \tau_1$.

We have to deal with two challenges. Firstly, we would like to produce surrogates of the actual wind speed series and also simultaneously manage to retrieve the exact time correspondence between the surrogates and the proxy series' values. Secondly, we would like to create surrogates that are as independent as possible from the original series, in the sense that the surrogates are not just rearranged values of the original series. This will prevent us from producing T-year return values that are almost the same for each surrogate, given that we use the whole length of the series for the calculation.

The first challenge is addressed by producing the amplitude preserving surrogates, hence the proxy distribution can be re-ordered in time to match the surrogates. To address the second challenge we proceed as follows. Assuming the two series are sampled over N years, we create 12 'monthly' series (spanning the entire N years) by simply taking all the values of the original series that correspond to each calendar month. For example, we create a 'January' series by just taking all the N blocks that correspond to the month of January in the original series. Then we produce N AAFT surrogates of these twelve series and keep only the first j points that correspond to the length of the corresponding month. Therefore, by this method we create N different AAFT surrogates of each calendar month, which can be joined together to form the new N-year long surrogate of the original series.

## 4.3 Simulation

It has been shown that the annual maxima method (block maxima method with block sizes of one year) can produce reliable results when we have at least 20 years of data (Cook, 1985). But what should be recommended if one only has access to a smaller actual wind speed data sample of a couple of years? This is the familiar situation that many wind farm developers find themselves. We will demonstrate that it is unwise to blindly assume that the return level estimate from short time series is accurate. Furthermore, we offer a thorough analysis of the statistical properties of different estimators as a function of the length of the available time series.

Our aim in the following simulation study is to assume data availability is limited to a record of actual hourly wind speeds of k years long where k =1,…,45. The challenge is then to understand under which conditions the reanalysis data can help us produce 50-year return wind speeds comparable to what could have been achieved using the whole 45 years of actual data (which is not available in practice).
We define the 50-year return value of a k-year long actual wind speed series, by $V^A(k)$, and the corresponding value of the reanalysis data by $V^R(k)$. Moreover, denote the 50-year return value obtained from k-years of actual wind time series, but using the whole 45-years reanalysis series (estimated using the quantile calibration method of Section 4.1) by $\hat{V}^R(k)$.

The procedure followed in order to test the contribution of the reanalysis series to the production of trustworthy 50-year return values, using the block maxima method, is the following:

(1) Produce M surrogates of the 45 years actual wind speed series using the methodology explained in Section 4.2. For each of the M surrogates find the corresponding M reanalysis series.

(2) For each pair (actual, reanalysis data) of surrogates, estimate $V^A(45)$ and $V^R(45)$, using the whole length of the series and the annual maxima method. $V^A(45)$ will represent the most accurate estimate of the 50-year return we can find.

(3) For each surrogate of the actual data randomly (block) sample 45 series of length k = 1,…, 45, and find the corresponding reanalysis series. Then, for each new pair of series corresponding to different k, calculate:

- The corresponding estimate $V^A(k)$ for each k, using the block maxima method. We would like to keep the number of fitted maxima not less than 20 (Palutikof et al., 1999). Hence for k = 1,…,19 we use block sizes of length (k.365/20) days, and for 20 ≤ k ≤ 45 block sizes of one year.
- The quantile calibration ratio $\hat{R}(k)$ as explained in Section 4.1.
- The corresponding 50-year return level using the whole length of reanalysis data and the calibration ratio, given by $\hat{V}^R(k) = \hat{R}(k)V^R(45)$.

In the above procedure, whenever we are using the block maxima method we have to firstly decide whether a Gumbel distribution is the most appropriate one (hence set $\xi = 0$). To do that we use a likelihood ratio test (Section 3.3), and reject the null hypothesis (Gumbel distribution is adequate) for p-values < 0.05. Secondly, in order to test the goodness of fit of the GEV distributions, we use the Anderson-Darling test (Section 3.4) with a 5% significance level.

# 5. Dynamic risk estimation

From the outset, wind speed is challenging because it displays multiple seasonalities such as the intra-annual fluctuations and intra-day variability that may lead to a diurnal cycle. Furthermore, temporal dependence from one observation to the next implies that the time series cannot be viewed as a collection of independent observations. Extreme events may occur as part of a particular windstorm and as such, multiple extremes (observations above a specific threshold) may be registered and cannot be deemed independent, as they have been generated by the same windstorm.

In contrast, our dynamic approach first uses non-parametric kernel regression to deseasonalise the time series and then applies a semi-parametric technique to obtain conditional estimates of a specific quantile. Rather than ignore the time series dependency, we explicitly take this into account when modelling the temporal evolution of the tail of the distribution. The approach utilizes an autoregressive model for the quantile after first identifying deterministic components for the intra-annual seasonality in both the level and variability.

Following Haslet and Raftery (1989), we investigate the square root of wind speed instead of modeling wind speed directly. This gives a time series which is more amenable to time series modelling and has also been employed by Taylor et al. (2009).

## 5.1 Intra-annual seasonality

We first define a time of year variable such that zero equates with the 1$^{st}$ January and one with the 31$^{st}$ December. A graph of the wind speed against this time of year variable demonstrates a number of interesting characteristics (Figure 3). Firstly the wind speed is, on average higher during the winter than the summer. Secondly, the wind speed variability is highest in the wintertime. There are a number of techniques that could be used to extract estimates of the level and variability throughout the year. A simple method would be to employ dummy variables for each month of the year but such a piecewise constant approach requires estimation of twelve parameters and would not give a smooth representation of the intra-annual seasonality. Other parametric models that could be used include a polynomial model (McSharry et al., 2005) and a trigonometric expansion using appropriate sine and cosine terms to describe the harmonics of the annual cycle (Taylor et al., 2009). In order to obtain a parsimonious model, we prefer a non-parametric approach known as kernel regression with the time of year as an explanatory variable, requiring estimation of a single bandwidth parameter. The solid line shows the estimated mean and the dashed lines indicate one standard deviation above and below the mean.

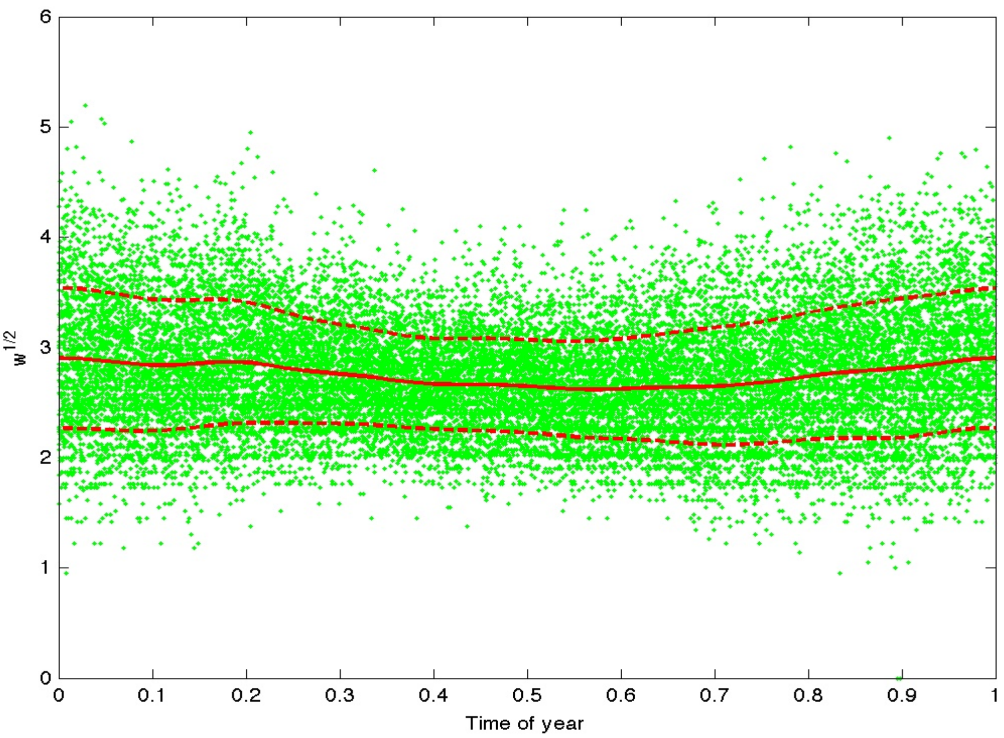


**Figure 3: Modal graph of the square root of wind speed at Schiphol airport between 1950 and 2010. The horizontal axis reflects the time of year with 1st January at zero and 31st December at 1.**

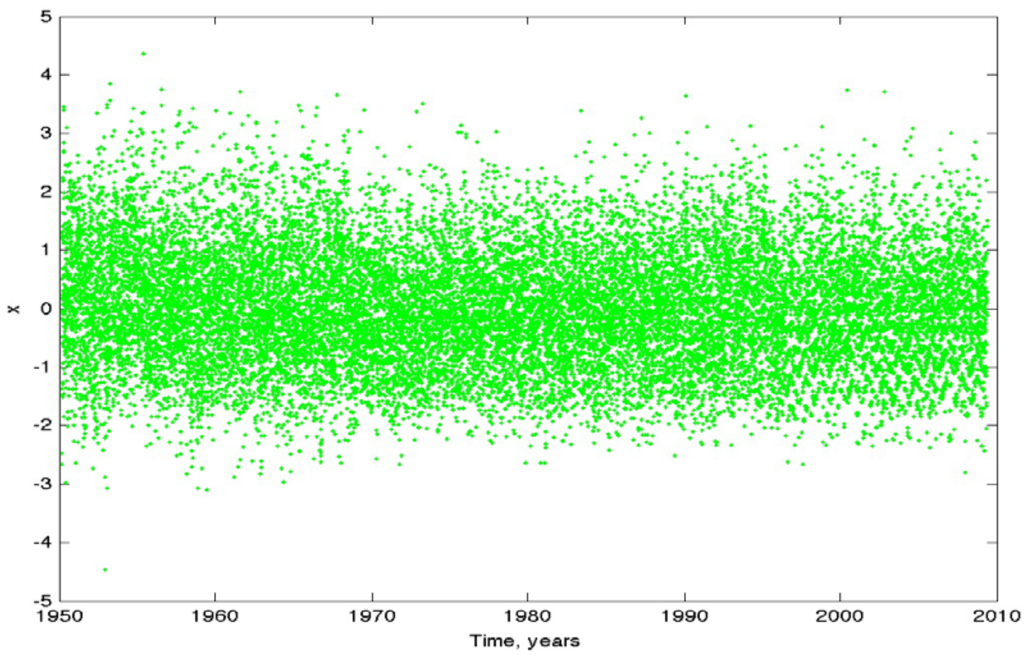


**Figure 4: Deaseasonalised wind speed time series obtained by subtracting the intra-annual mean and dividing by the intra-annual standard deviation.**

## 5.2 Conditional autoregressive risk estimation

We can define the quantile loss function, also known as the check function, for a given proportion, θ, defined in the interval [0, 1], as:

$$\rho(Q_t,\theta) = \sum_{t|y_t \geq Q} \theta \left| y_t - Q_t \right| + \sum_{t|y_t < Q} (1-\theta) \left| y_t - Q_t \right|.$$

We model the time-varying quantile, $Q_t(\theta)$ as an autoregressive process (Engle and Manganelli, 2004)

$$Q_t(\theta) = \omega + \alpha Q_{t-1}(\theta) + \beta x_{t-1}.$$

This model for the quantile is similar in spirit to an autoregressive model and has an intercept ω, a dependence on previous quantiles via parameter α, and a dependence on the previous value of the time series through the parameter β. The quantile regression check-function is used as an objective function for estimating the three parameters of the model (ω, α and β). The parameter estimation was carried out by sampling the parameters in a grid 10,000 times where each parameter spans the interval [0, 1]. Nelder-Mead simplex optimisation was then used to fine tune the ten estimates with the lowest quantile loss function values in order to arrive at a single best estimate.

# 6. Results

## 6.1 Surrogates

Before moving on to the actual results, we may convince ourselves that the surrogates produced indeed look and behave like real wind speed time series. Figure 6 shows how one example of a surrogate time series compares to the actual data in Figure 5. Note that both the seasonality and the amplitude of the actual data are preserved. The unconditional distribution of the actual wind speed time series is maintained. Moreover, Figure 7 shows the Autocorrelation Function (ACF) of the two series for a maximum lag of seven days. These two plots are almost identical, which shows that the surrogate method has successfully preserved the linear temporal relationships (and therefore the power spectrum of the original series).

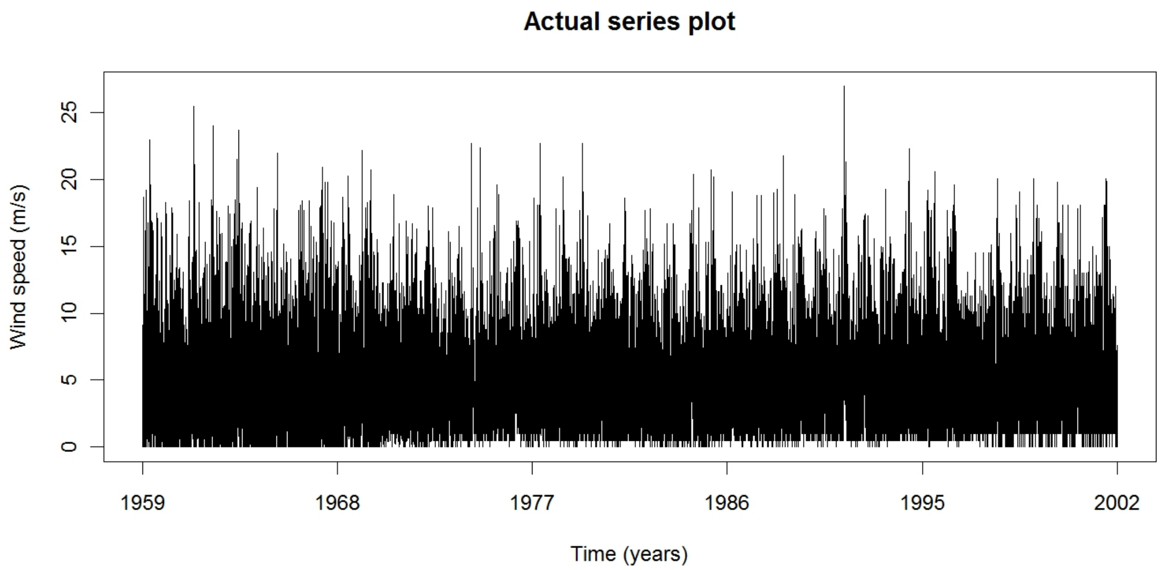


**Figure 5: Actual wind speed time series over a two-year period.**

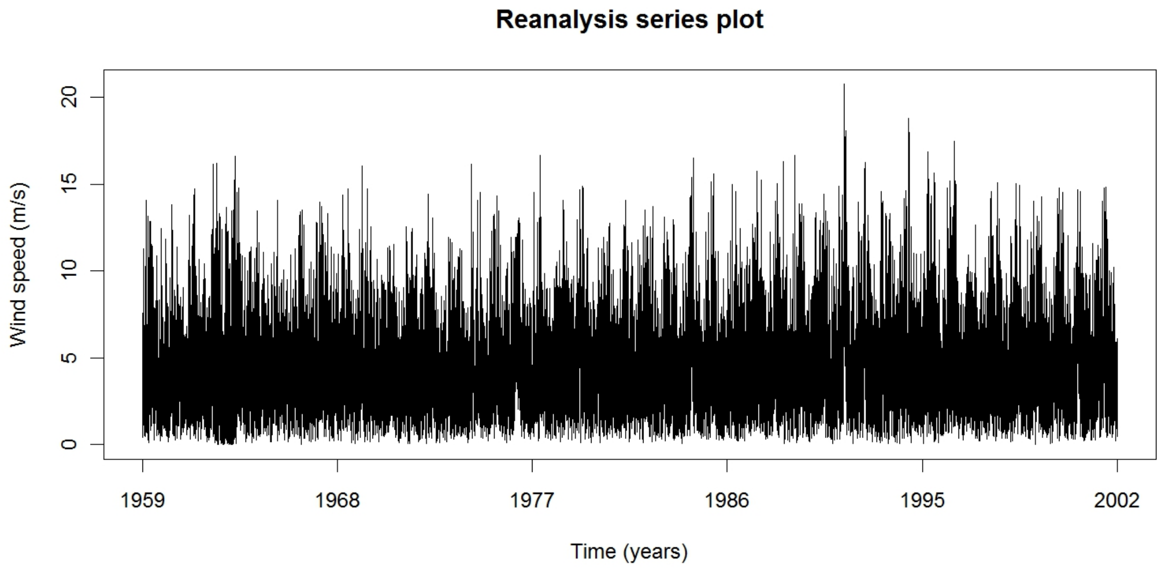


**Figure 6: Surrogate wind time series over a two year period.**

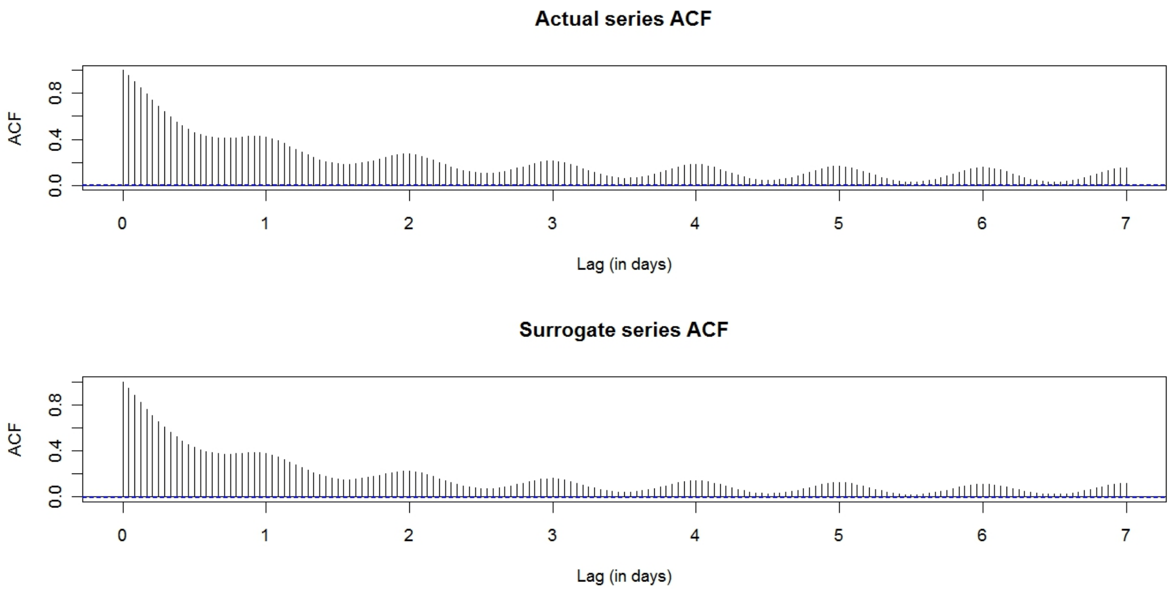


**Figure 7: Autocorrelation for the actual and surrogate wind speed time series for lags up to seven days.**

## 6.2 Return level estimation

By following the procedure explained above, we produce 2,500 50-year return speeds, $V^A(45)$, using the whole 45 years of surrogates of the actual data. These are the most accurate 50-year return speeds we can estimate, and hence all results will be based on how well the reanalysis data can match these reference values. From these 2,500 values 326 are removed because the GEV fit of their corresponding surrogates was not good enough according to the AD test at a 5% level of significance. For this reason, all results ($V^A(k)$ and $\hat{V}^R(k)$ values) acquired using these particular surrogate series will also be ignored.

Examing the $V^A(k)$ for k = 1,…,45 we notice that some of these values are extremely high (over 100 m/s) and hence should be considered as outliers as these values are physically unrealistic. Therefore,

it makes sense to eliminate those values before checking the GEV goodness of fit for each of the estimates. We achieve this by setting a wind speed upper bound of 100 m/s. After first removing these outliers, we then remove the estimates for each k that were produced from cases where the GEV goodness of fit was inferior, quantified by the AD test with a 5% level of significance. Finally we select M = 2,000 estimates for each k that are a result of a good distributional fit, and find the corresponding $\hat{R}(k)$ and $\hat{V}^R(k)$values.

From Figure 4 we can see that the quantile calibration ratio, $\hat{R}(k)$, is correct and sufficiently close to the optimum ratio, R(45) = $V^A(45)/V^R(45)$, for all k = 1,…,45. The median of the $\hat{R}(k)$estimates is very close to the median of the optimum ratio, and the magnitude of the difference decreases on average as we move towards using the whole 45 years of actual and reanalysis data for the calibration. From the variation in the 5% and 95% quantiles for each k, we observe that the distribution of the estimated ratio is also converging as we use more data for the calibration procedure.

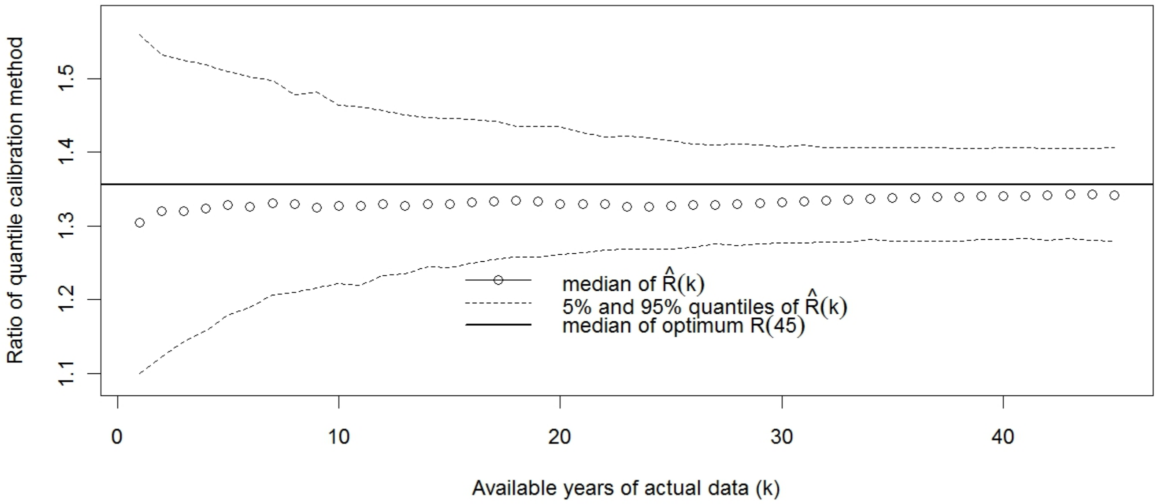


**Figure 8: Median, 5% and 95% quantiles of the quantile calibration ratio.**

In order to evaluate the accuracy of the $\hat{V}^R(k)$ estimates, compared to the most accurate 50-year return estimates available, the $V^A(45)$ estimates, we calculate the Mean Absolute Error (MAE) of the quantile calibrated reanalysis estimates for each k, defined by:

$$MAE(k) = \frac{1}{M}\sum_{i=1}^{M}\left|\hat{V}_i^R(k) - V_i^A(45)\right|.$$

The smaller the MAE is the more accurate are the corresponding estimates. As a benchmark, we use the analogous MAE produced from the M = 2,000 estimates $V^A(k)$ for k = 1,…,45. Note that the $V^A(k)$ estimates were produced the using the block maxima method applied to the k-year surrogates of the actual wind speed series.

Figure 9 shows the two MAEs as a function of the available years of actual data used, k. As we can see, the 50-year return estimates produced from the quantile calibration method applied to the reanalysis data are outperforming the corresponding estimates of the block maxima method applied to the actual data, for up to 20 years of available data. The benchmark method is better when we have more than 25 years of available actual data.

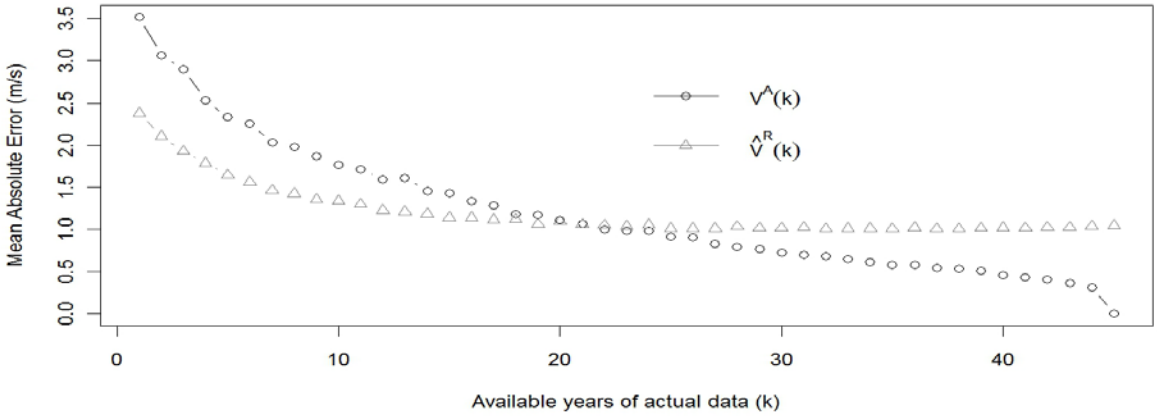


**Figure 9: Mean absolute error of the calibrated reanalysis $V^R(k)$ estiamtes and the $V^A(k)$ estimates produced using the block maxima method to k-year surrogates.**

Figure 10 shows the median the 5% and 95% quantiles of the $\hat{V}^R(k)$ and $V^A(k)$ and estimates for k years of actual data. The median, 5% and 95% quantiles of the $V^A(45)$ estimates are the most accurate estimates available. These serve as our reference and are highlighted by three straight lines on the plot. The quantile calibration method produces 50-year return estimates whose median underestimates the optimum median of $V^A(45)$ for almost all k, but not more than 1.1 m/s (for k = 1). On the other hand the median of $V^A(k)$ estimates are overestimating the optimum median for k < 14, and the magnitude of this overestimation is getting as large as 1.8 m/s (for k = 1) and remains larger than 1 m/s for k ≤ 8. Moreover, the distribution of the $\hat{V}^R(k)$ estimates is more concentrated than the distribution of the $V^A(k)$ estimates. This shows that the results of the quantile calibration method are less variable and hence more reliable than the results of the block maxima method applied to k years of actual data.

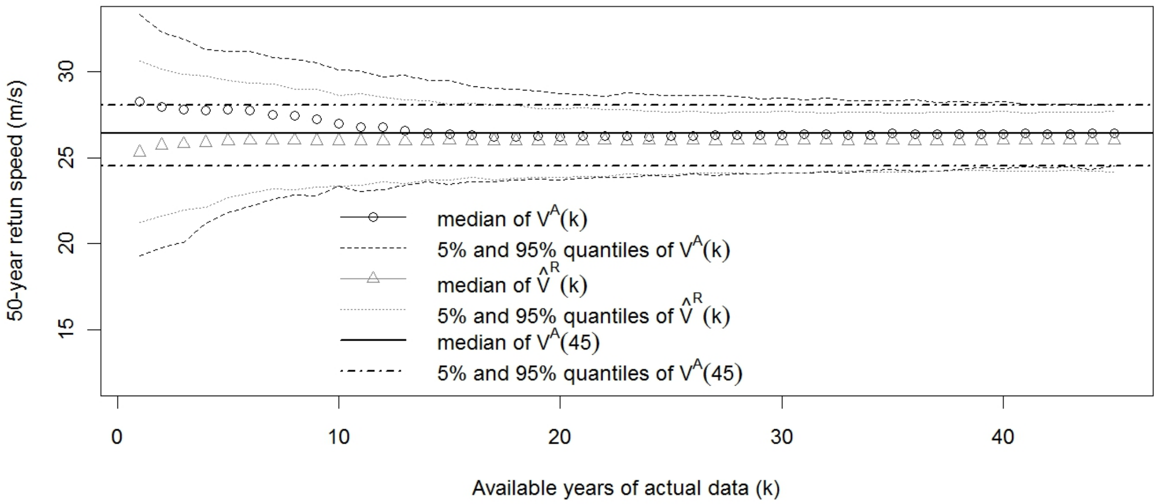


**Figure 10: Median, 5% and 95% quantiles of the estimate for k=1,...,45 years.**

The variability of the two sets of 50-year return estimates can also be compared using the Median Absolute Deviation statistic. This is a measure of variability, and is more resilient to outliers (hence more robust) than the standard deviation. Figure 11 shows that the quantile-calibrated reanalysis

estimates are less variable than the corresponding 50-year return estimates produced using the block maxima method applied to k-years of actual data, for all k < 41.

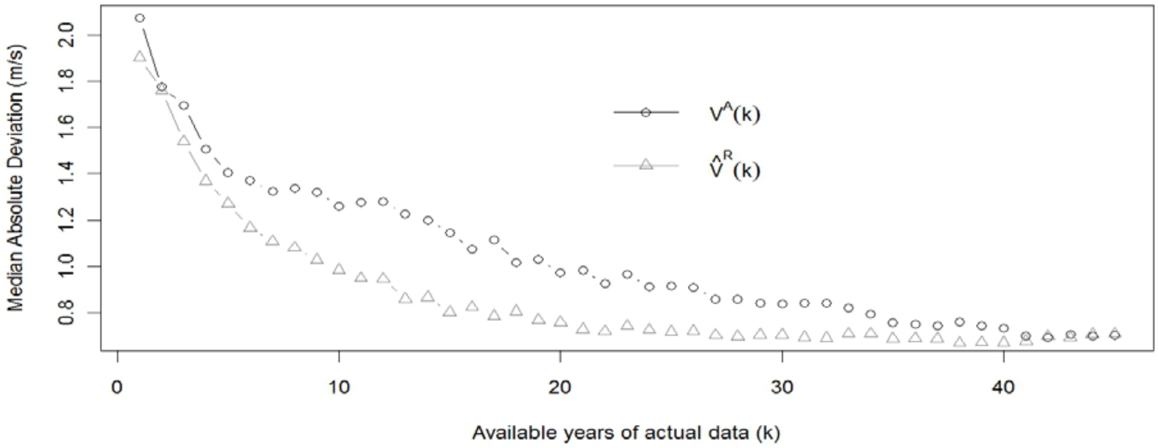


**Figure 11: Median absolute deviation of the estimates for k=1,...,45 years.**

## 6.3 Risk forecasting

We divided the available time series into training and testing datasets of approximately equal parts. These had time periods of 01-Mar-1950 to 31-Dec-1980 and 01-Jan1981 to 14-May-2009 respectively. The out-of-sample 95% quantile forecasts for the year of 1981 are shown in Figure 12.

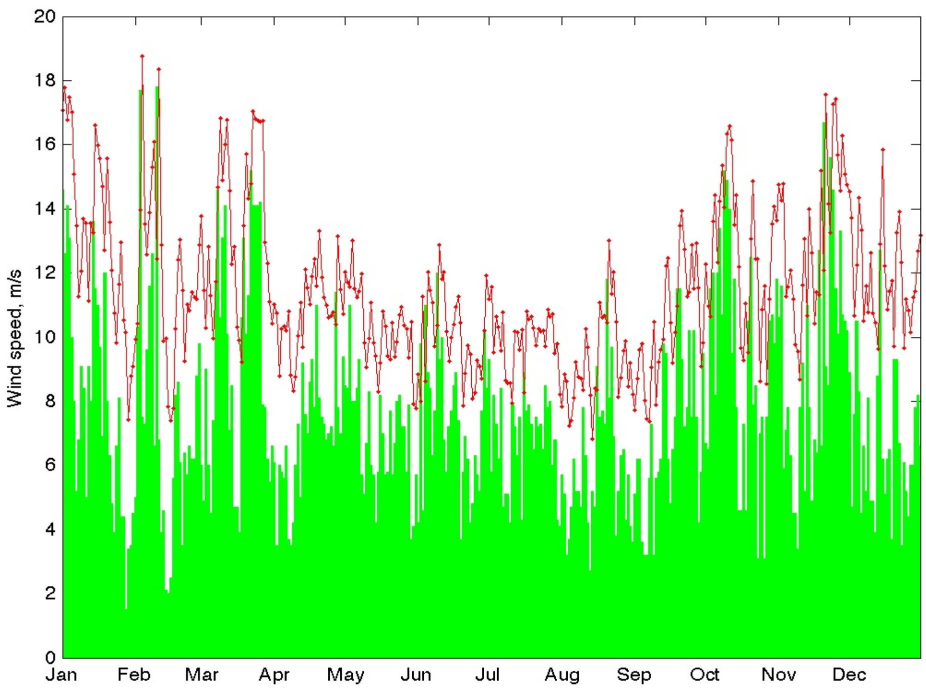


**Figure 12: Quantile forecasts of wind speed based on dynamic risk estimation approach for the year of 1981. The actual wind speeds (green) are compared with the one day ahead 95% quantile forecasts (red)**

We use the quantile regression check function score as a means of evaluating the performance of the quantile forecasts. Our forecast benchmark is the static unconditional quantile estimated using the training data set. To facilitate comparison, we also provide the skill score, which indicates the forecast gain as a percentage when using our dynamic risk estimation approach instead of the static unconditional estimate. Following McSharry et al. (2009), this gain, measured as a percentage improvement relative to the unconditional reference, is known as a skill Score and is defined as:

$$Skill = 1 - \frac{Score}{Score_{ref}}.$$

Table 2 gives the breakdown of results for the forecast comparison exercise. The results demonstrate an improvement of 26% from using the dynamic risk estimation approach over the static unconditional estimate.

**Table 2:** *Performance of the quantile forecasts measured using the quantile loss function.*

| Method\Data | Training | Testing |
|---|---|---|
| Unconditonal | 0.4139 | 0.3876 |
| Dynamic | 0.3353 | 0.2865 |
| Skill score | 19.00% | 26.09% |

# 7. Conclusions

This report was motivated by the practical need for an innovative way of producing accurate 50-year return wind speeds, when we have only a limited length of actual wind speed data. It showed that using the longer record of reanalysis data as a proxy has numerous advantages. We demonstrated the efficacy of our new method by using a long record of actual wind speed time series and corresponding reanalysis data from Schiphol airport.

One of our aims was to answer a number of questions:

(1) What if we have a small time series (k-years long) of actual wind speed data?
(2) Can we use a long enough series of reanalysis data to produce as reliable 50-year return speeds as by using a long enough series of actual data?
(3) How can we overcome the problem of different temporal resolution of the actual and reanalysis series?

In order to answer these questions we applied the block maxima method to 2,000 surrogates of the actual and reanalysis series, by assuming we only had access to k years of actual data (for k = 1,…,45) and 45 years of reanalysis data. We introduced the quantile calibration method in order to overcome the challenge of different sampling frequencies of the two series and systematic underestimation by the reanalysis time series.

The results show that for 1 ≤ k ≤ 20, using all 45 years of reanalysis data and the quantile calibration method applied to the k years of actual data, yields 50-year return estimates that are less biased than the estimates produced by just applying the block maxima method directly to the k years of actual data. Moreover, the 50-year return levels produced by the quantile calibration method are less variable and hence more reliable (especially for k years where 2 < k < 35) than the estimates produced by the block maxima method applied directly to k years of actual data.

Of course, there are other more suitable methods than the block maxima technique that can be used to produce 50-year return estimates given k < 20 years of actual available data. Most popular are the peaks over threshold (POT) method (Cook, 1985; Rootzen and Tajvidi, 1995; McNeil, 1997), and the method of independent storms (MIS) (Cook 1982, 1985). The POT method increases the number of

samples used by selecting inter-annual storms exceeding a certain threshold. The resulting maxima are then fitted to a generalized Pareto distribution (GPD). The MIS imposes a minimum separation between maxima and then these maxima are fitted to a GEV distribution. In contrast to the POT method, the MIS guarantees that the samples are independent. For both of these methods we need to identify specific threshold (or minimum separation) parameters by mainly using exploratory plots, which is practically infeasible when undertaking a simulation study as we would have to apply it to a large number of surrogate series as has been used in this paper.

In our analysis of quantile forecasting, we have shown that an autoregressive time series model can be deployed to provide real-time risk assessment. This approach could be easily applied to a large number of sites and could help to manage risk on a daily basis.

# 8. References

Anderson T, Darling DA. (1952). Asymptotic theory of certain ”goodness-of-fit” criteria based on stochastic processes. Annals of Mathematical Statistics 1952; 23:193212.

Coles S. (2001). An introduction to statistical modeling of extreme values. Springer Series in Statistics, Springer-Verlag: London.

Cook NJ. (1982) Towards better estimation of wind speeds. Wind Eng. Ind. Aerodyn. 9:295323.

Cook N. (1985). The Designer’s Guide to Wind Loading of Building Structures: Background, damage survey, wind data, and structural classfication. The Designer’s Guide to Wind Loading of Building Structures, Building Research Establishment, Dept. of the Environment.

Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics, 22 (4): 367-381.

Fisher R, Tippett L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 1928; 7:180–190.

Haslett, J. and Raftery, A.E. (1989). Space-time modelling with long memory dependence: assessing Ireland’s wind power resource. J. Roy. Stat. Soc. C, 38(1): 1-50.

Larsen XG, Mann J. (2006). The effects of disjunct sampling and averaging time on maximum mean wind speeds. Journal of Wind Engineering and Industrial Aerodynamics 2006; 94(8):581 – 602.

Leadbetter M, Lindgren G, Rootz´en H. Extremes and related properties of random sequences and processes. Springer series in statistics, Springer-Verlag, 1983.

McNeil AJ. (1997). Estimating the tails of loss severity distributions using extreme value theory. Astin Bulletin, 27(1):117–138.

McSharry, P.E. Bouwman, S. and Bloemhof, G. (2005). Probabilistic forecasts of the magnitude and timing of peak electricity demand. IEEE Trans. Power Systems, 20(2): 1166- 1172.

Nelder JA, Mead R. (1965). A simplex method for function minimization 1965; 7(4):308–313.

Palutikof JP, Brabson BB, Lister DH, Adcock ST. (1999). A review of methods to calculate extreme wind speeds. Meteorological Applications, 6(02):119–132.

Perrin O, Rootzn H, Taesler R. (2006). A discussion of statistical methods used to estimate extreme wind speeds. Theoretical and Applied Climatology 2006; 85:203–215.

Reiss R, Thomas M. (2007). Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields. Birkhauser.

Rootzen H, Tajvidi N. (1995). Extreme value statistics and wind storm losses: A case study. Scand. Actuarial J 1995; 1:70–94.

Taylor, J.W., McSharry, P.E. and Buizza, R. (2009). Wind power density forecasting using wind ensemble predictions and time series models. IEEE Transactions on Energy Conversion, 24, 775-782.

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. (1992). Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena 1992; 58(14):77 – 94.

Von Mises R. (1936). La distribution de la plus grande de n valeurs. Revue de l’Union Interbalkanique 1936; 1:1–20.