

# Probabilistic Model for Wind Speed Variability Encountered by a Vessel

Igor Rychlik<sup>1</sup>, Wengang Mao<sup>2</sup>

<sup>1</sup>Mathematical Sciences, Chalmers University of Technology, Göteborg, Sweden

<sup>2</sup>Department of Shipping and Marine Technology, Chalmers University of Technology, Göteborg, Sweden

Email: [rychlik@chalmers.se](mailto:rychlik@chalmers.se), [wengang.mao@chalmers.se](mailto:wengang.mao@chalmers.se)

Received 6 August 2014; revised 12 September 2014; accepted 2 October 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

As a result of social awareness of air emission due to the use of fossil fuels, the utilization of the natural wind power resources becomes an important option to avoid the dependence on fossil resources in industrial activities. For example, the maritime industry, which is responsible for more than 90% of the world trade transport, has already started to look for solutions to use wind power as auxiliary propulsion for ships. The practical installation of the wind facilities often requires large amount of investment, while uncertainties for the corresponding energy gains are large. Therefore a reliable model to describe the variability of wind speeds is needed to estimate the expected available wind power, coefficient of the variation of the power and other statistics of interest, e.g. expected length of the wind conditions favorable for the wind-energy harvesting. In this paper, wind speeds are modeled by means of a spatio-temporal transformed Gaussian field. Its dependence structure is localized by introduction of time and space dependent parameters in the field. The model has the advantage of having a relatively small number of parameters. These parameters have natural physical interpretation and are statistically fitted to represent variability of observed wind speeds in ERA Interim reanalysis data set.

## Keywords

Wind Speeds, Wind-Energy, Spatio-Temporal Model, Gaussian Fields

---

## 1. Introduction

In the literature typically cumulative distribution function (CDF) of wind speed  $W$ , say, is understood as the long-term CDF of the wind speeds at some location or region. The distribution can be interpreted as variability of  $W$  at a randomly taken time during a year. Weibull distribution gives often a good fit. Limiting time span to,

for example, January month affects the  $W$  CDF simply because, as it is the case for many geophysical quantities, the variability of  $W$  depends on seasons. To avoid ambiguity when discussing the distribution of  $W$ , time span and region over which the observations of  $W$  are gathered need to be clearly specified. By shrinking the time span to a single moment  $t$  and geographical region to a location  $\mathbf{p}$ , one obtains (in the limit) the distribution of  $W(\mathbf{p}, t)$ . This is used as the distribution of  $W$  in this paper. Obviously the long-term CDF can be retrieved from the “local”  $W(\mathbf{p}, t)$  distributions by means of an average of the local distributions, viz for a fixed location  $\mathbf{p}$

$$\mathbb{P}(W \leq w) = \frac{1}{S} \int_s^{s+S} \mathbb{P}(W(\mathbf{p}, t) \leq w) dt, \quad (1)$$

where  $S$  can be a month, a season or a year. Similarly the long-term CDF over a region  $A$ , say, is proportional to  $\int_A \int_s^{s+S} \mathbb{P}(W(\mathbf{p}, t) \leq w) dt d\mathbf{p}$ .

In order to identify the distributions at all positions  $\mathbf{p}$  and times  $t$ , vast amount of data are needed. Here the reconstruction of  $W$  from numerical ocean-atmosphere models based on large-scale meteorological data, called also reanalysis, is utilized to fit a model. The reanalysis does not represent actual measurements of quantities but extrapolations to the grid locations based on simulations from complex dynamical models. It is defined on regular grids in time and space, and hence convenient to use. In this paper, the ERA Interim data [1] produced by European Centre for Medium-Range Weather Forecasts is used to fit the model. However the model can also be fitted to other data sets, e.g. to satellites wind measurements which has also good spatial coverage, see [2].

Modeling spatial and temporal dependence of wind speed is a very complex problem. Models proposed in the literature are reviewed in [3]. Here we propose to use the transformed Gaussian model, which assumes that there exists a deterministic function  $G(w)$ , possibly dependent on location  $\mathbf{p}$ , such that  $X(\mathbf{p}, t) = G(W(\mathbf{p}, t))$  is Gaussian. The  $X(\mathbf{p}, t)$  field is defined by the mean  $m(\mathbf{p}, t)$  and covariance structure  $\text{Cov}(X(\mathbf{p}_1, t_1), X(\mathbf{p}_2, t_2))$ . Obviously for a given transformation  $G$  and many years of hind-cast, one could estimate the covariance for any pair  $(\mathbf{p}_1, t_1)$ ,  $(\mathbf{p}_2, t_2)$ , see, e.g. [4] [5]. However such an approach is limited to relatively small grids in space. Employing the empirical covariances in time and space would result in huge matrices, which limit the applicability of such empirical approach. Consequently, a simple parametric model that catches only some aspects of the wind speed variability, important for a particular application, is of practical interest. The minimal requirements on the model are that it should provide a correct estimates of long-term distributions of the wind speeds, accurate predictions of average durations of the extreme winds conditions and reliable estimates of CDFs of top speeds during storms encountered by a vessel. In order to demonstrate the capability of the proposed model for such minimal requirements, this paper is organized as follows:

In Section 2, a general construction of non-stationary model for wind speed variability in time and space is presented. Section 3 presents probabilistic model for the velocity of storms movements. Then statistical properties of some storms characteristics are described in Section 4. The physical interpretations of the introduced parameters are also given in this section and in Appendix 1. In Section 5, on board measured wind speeds are used to validate the proposed model, where the long term CDFs of encountered wind speeds and persistence statistics are used. Total forty routes are used, see Figure 1. The time when routes were sailed are well spread over a year. Finally in Section 6, means to simulate the encountered wind speeds are briefly reviewed. Paper closes with three appendixes containing somewhat more technical matters.

## 2. Transformed Gaussian Model and Long-Term CDFs

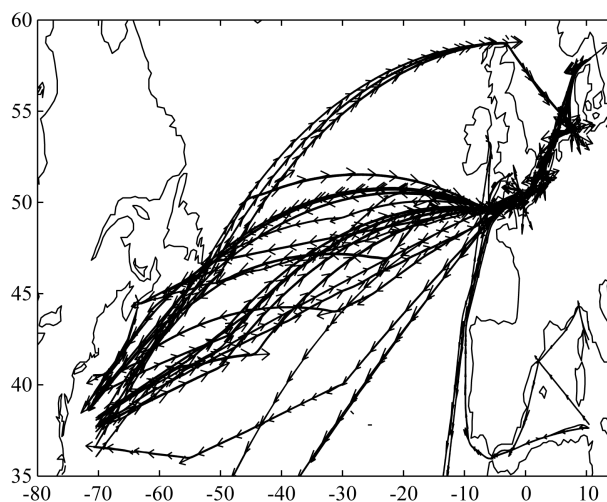
In this section we shall introduce the transformed Gaussian model for the variability of wind speeds. In particular the transformation  $G$  making the transformed wind speed data  $X = G(W)$  normally distributed will be presented. Seasonal model for the mean and variance of  $X$  is given and assumed normality of  $X$  validated.

The wind speed  $W(\mathbf{p}, t)$  is the ten minutes average of the wind speed measured at position  $\mathbf{p}$ , defined in degrees of longitude and latitude, while  $t$  is the time of the year. We will use the transformation  $G(w) = w^a$ , where  $a$  is a fixed constant that depends on the location  $\mathbf{p}$ , viz

$$X(\mathbf{p}, t) = W(\mathbf{p}, t)^{a(\mathbf{p})} \quad (2)$$

The parameter  $a$  is nonnegative with convention that the case  $a = 0$  corresponds to the logarithm. We assume that  $X(\mathbf{p}, t)$  is normally distributed.

Mean and variance of  $X(\mathbf{p}, t)$ , denoted by  $m(\mathbf{p}, t)$ ,  $\sigma^2(\mathbf{p}, t)$ , respectively, depend both on position and



**Figure 1.** The considered routes in the validation process.

time. The temporal variability of mean and variance is approximated by seasonal components with trends defined as follows

$$m(\mathbf{p}, t) = m_0(\mathbf{p}) + m_1(\mathbf{p}) \cos(2\pi t) + m_2(\mathbf{p}) \sin(2\pi t) + m_3(\mathbf{p}) t, \quad (3)$$

$$\ln(\sigma^2(\mathbf{p}, t)) = b_0(\mathbf{p}) + b_1(\mathbf{p}) \cos(2\pi t) + b_2(\mathbf{p}) \sin(2\pi t) + b_3(\mathbf{p}) t. \quad (4)$$

Here  $t$  has units years. This type of model has been used in the literature, see e.g. classical paper [6].

**Remark 1** For a fixed position  $\mathbf{p}$ , the parameters  $a$  and  $m_i$  in Equation (3) are fitted simultaneously in such a way that the distance between yearly long-term empirical CDF of  $W(\mathbf{p}, t)^a - m(\mathbf{p}, t)$  and a Gaussian distribution is minimized.

More precisely for a wind data at fixed position  $\mathbf{p}$  and parameter  $a \in (0, 1)$  we evaluate  $x(t) = w(t)^a$  and fit regression Equation (3). Then the residual  $\epsilon(t) = x(t) - m(t)$  is evaluated and its empirical CDF estimated. Next a distance between the empirical CDF and the Gaussian CDF (fitted to  $\epsilon(t)$ ) is evaluated. Finally the parameter  $a^*$  that minimizes the distance is the estimate of  $a$ .

A table of  $a$  and  $m_i$  values as function of the location  $\mathbf{p}$  is created. As additional parameters of the model will be estimated new columns with parameters estimates will be added to the table. For example, having estimated  $a$  and  $m_i$  the variance  $\sigma^2(\mathbf{p}, t)$ , defined in Equation (4), is fitted using an additional assumption that properties of the wind speed changes slowly in time. Then the parameters  $b_i$  are saved in the table. More details of model estimations are given in Appendix 3.

### Validation of Gaussianity of $X(\mathbf{p}, t)$

Ten years of data  $W(\mathbf{p}, t)$  were used to estimate parameters  $a$ ,  $m_i$  and  $b_i$  in Equation (3), (4) in North Atlantic on a grid of 0.75 degree. **Figure 2** presents the estimates of parameter  $a$ . At offshore locations  $a$  values vary around 0.8 while close to shore or inland locations  $a$  can be much smaller. Note that small values of  $a$  indicate larger departures of the observed wind speeds distribution from the Gaussian one.

Usefulness of the proposed model relies on the accuracy of the approximation of  $X(\mathbf{p}, t)$  CDF by Gaussian distribution. The Gaussianity of the process  $X(t) = X(\mathbf{p}, t)$  has been validated for the Northern Atlantic. An example of conducted validations is shown in **Figure 3**. In the figure the left plot shows ten years of  $W$  process limited to two weeks in the middle of February, at locations  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$ ,  $(-20, 45)$ , plotted on the normal probability paper. (It is assumed that the winds are stationary for such short period of time.) In the right plot of the figure the transformed data  $X = G(W)$  is plotted on normal probability paper. One can see that  $X$  CDFs are well approximated by the Gaussian distributions.

In the right plots of **Figure 4**, the standard deviation  $\sigma(\mathbf{p}, t)$ , defined in Equation (3), is presented for February and August, respectively. One can see that the standard deviation changes considerably with the geographical location but is less dependent on season. We turn next to presentation of variability of the parameter

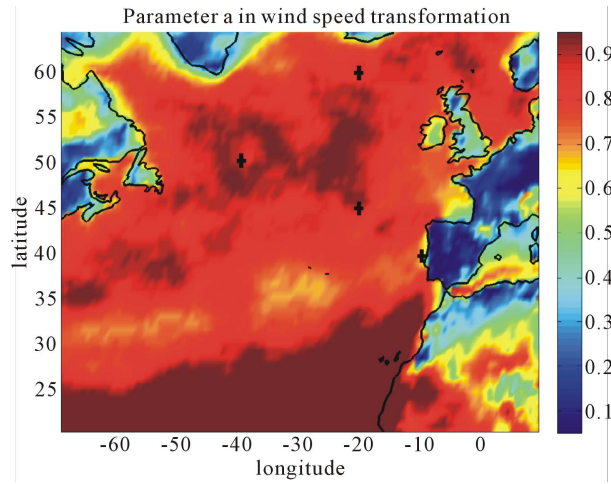


Figure 2. Values of parameter  $a$  in the transformation Equation (2).

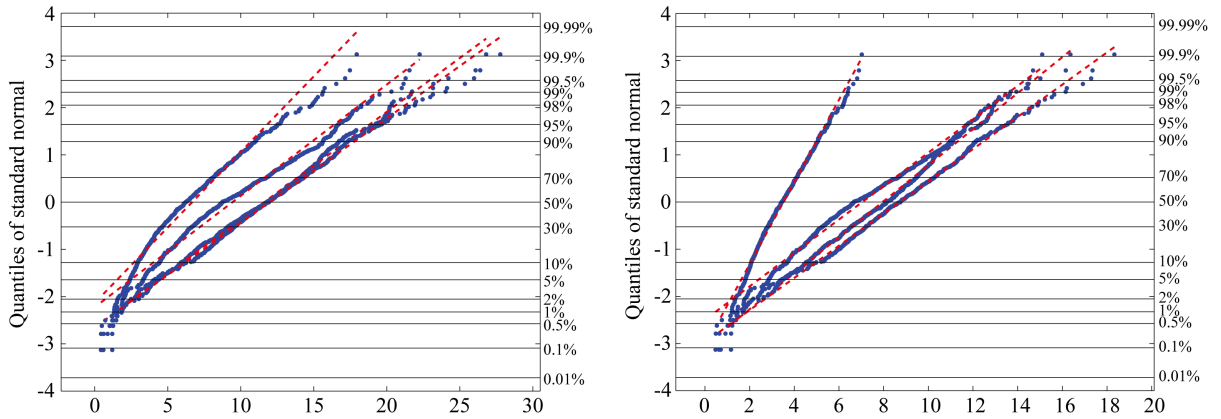


Figure 3. Left: Ten years of wind speeds  $W(t)$  with  $t$  limited to February at the four locations.  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$ ,  $(-20, 45)$  plotted on normal probability paper. Right: Transformed wind speeds  $X(t)$  limited to February at the four locations plotted on normal probability paper. The values of parameter  $a$  in transformation Equation (2) are  $a = 0.850, 0.675, 0.875, 0.875$ , respectively.

$m(\mathbf{p}, t)$ , i.e. the mean of  $X(\mathbf{p}, t)$  defined in Equation (2). However since units of  $m$  are not physical we choose to show the variability of the median speed

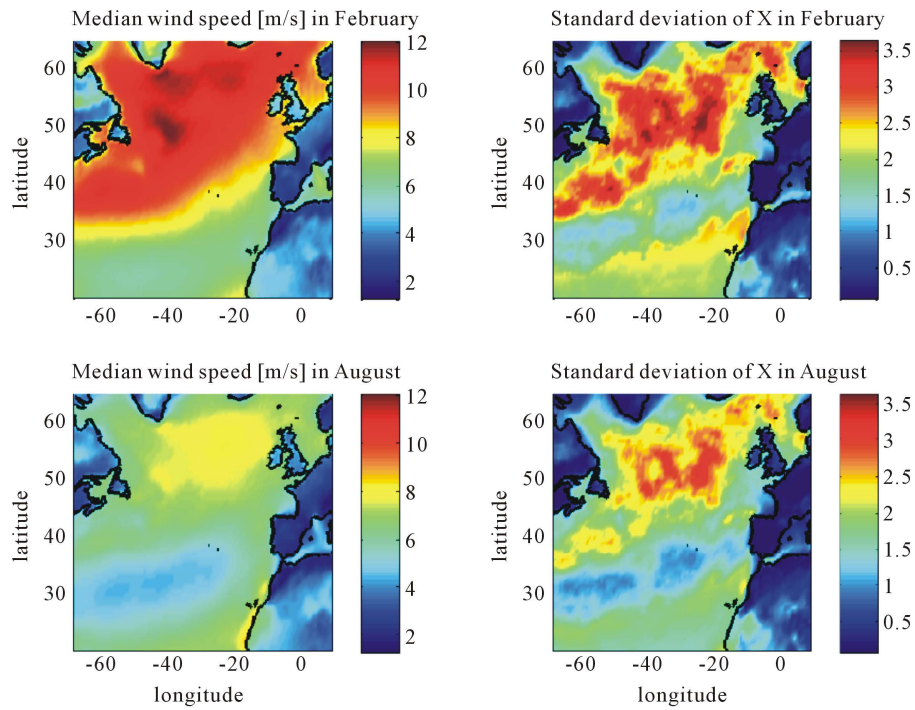
$$\mu(\mathbf{p}, t) = m(\mathbf{p}, t)^{1/\alpha(\mathbf{p})} \tag{5}$$

instead. The values of the median for February and August are presented in two left plots of Figure 4. As expected, wind speeds are higher in winter than in summer.

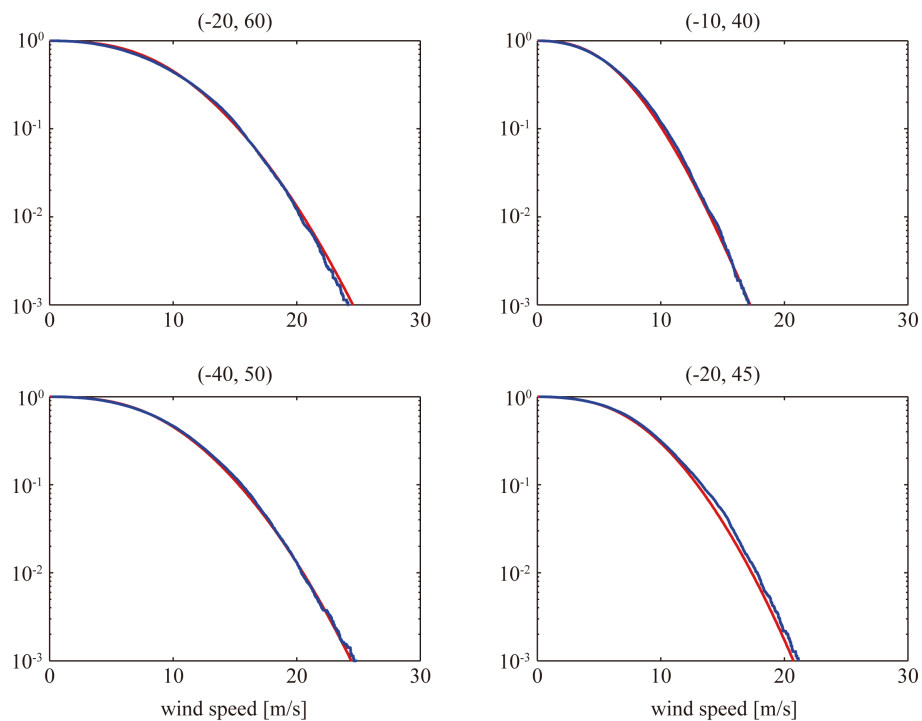
Finally we check whether the regressions Equations (3) and (4) used to model seasonal variability of  $m$  and  $\sigma^2$  leads to accurate estimates of the long-term CDF of  $W$  at position  $\mathbf{p}$ . Employing Gaussianity assumption of  $X$  CDF the theoretical long-term CDF of wind speeds at a fixed position  $\mathbf{p}$ , defined in Equation (1), is given by

$$\mathbb{P}(W \leq w) = \frac{1}{S} \int_s^{s+S} \Phi\left(\frac{w^{\alpha(\mathbf{p})} - m(\mathbf{p}, t)}{\sigma(\mathbf{p}, t)}\right) dt, \tag{6}$$

where  $\Phi(x)$  is the CDF of a standard Gaussian (normal) variable. In Figure 5, the yearly probabilities for wind speeds  $\mathbb{P}(W > w)$  computed using Equation (6) at four locations in North Atlantic are compared with the empirical estimates. (The locations are marked by crosses in Figure 2) One can see that the agreement between the estimates is excellent.



**Figure 4.** *Left top*—Median wind speed  $\mu$  [m/s], defined in Equation (5), in February; *Left bottom*—Median wind speed in August; *Right top*—Standard deviation of  $X$ , computed by means of Equation (4) in February; *Right bottom*—The standard deviation in August.



**Figure 5.** Comparisons of estimates of the long-term probability  $\mathbb{P}(W > w)$  for yearly wind speeds variability Equation (1) at four locations defined in degrees of longitude and latitude:  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$  and  $(-20, 45)$ . The solid line is the probability computed using Equation (6) with  $S = 1$  year. Somewhat more irregular lines are the estimated probabilities based on ten years of hind-cast data.

### 3. Velocity of a Wind Storm

A storm occurring at time  $t$  is a region where  $W(\mathbf{p}, t) \geq u$ , e.g.  $u = 15$  m/s. The border of a storm is a  $u$ -level contour  $\{\mathbf{p} : W(\mathbf{p}, t) = u\}$ . The border changes as storms move, grow or fall. In a classical paper [7] Longuet-Higgins has introduced velocities to study movements of random surfaces. There are several definitions of velocities proposed in the literature, see [8]. Here we will use velocity in a fixed direction  $\theta$ , say. The direction  $\theta$  will be called the main azimuth of a storm. It will be defined in Remark 2, see also Example 1. As customary we use the convention that the direction south to north has azimuth  $\theta = 0$  while azimuth  $\alpha = 90^\circ$  for the direction west to east.

Following [8] the velocities in the direction  $\theta$  and  $\theta - 90^\circ$  are given by

$$\mathbf{V}_\theta = -\frac{W_t}{W_\theta}, \quad \mathbf{V}_{\theta-90^\circ} = -\frac{W_t}{W_{\theta-90^\circ}}, \quad (7)$$

where  $W_t$  is the time derivative of the wind speed,  $W_\theta$  and  $W_{\theta-90^\circ}$  are the directional derivatives having azimuths  $\theta$ ,  $\theta - 90^\circ$ , respectively. These are evaluated at a position  $\mathbf{p}$  on the  $u$ -level contour. Note that time  $t$  is fixed.

The general assumption of this paper is that parameter  $a$  does not depend on time and changes much slower in space than the wind speed  $W$  varies, see Figure 2. Hence the gradient  $\nabla W = (W_x, W_y, W_t)$  can be approximated as follows,

$$\nabla W \approx (1/a) X^{1/a-1} \nabla X, \quad (8)$$

where  $\nabla X$  is the gradient of  $X$ -field. Hence velocities defined in Equation (7) can be approximated by

$$\mathbf{V}_\theta = -\frac{X_t}{X_\theta}, \quad \mathbf{V}_{\theta-90^\circ} = -\frac{X_t}{X_{\theta-90^\circ}}. \quad (9)$$

For a homogeneous Gaussian field the velocities have median values equal to

$$\mathbf{v}_\theta = -\frac{\text{Cov}(X_\theta(\mathbf{p}, t), X_t(\mathbf{p}, t))}{\text{Var}(X_\theta)}, \quad \mathbf{v}_{\theta-90^\circ} = -\frac{\text{Cov}(X_{\theta-90^\circ}(\mathbf{p}, t), X_t(\mathbf{p}, t))}{\text{Var}(X_{\theta-90^\circ})}, \quad (10)$$

see [8] for proof. The speeds in directions  $\theta$  and  $\theta - 90^\circ$  will be denoted by  $v_\theta$ ,  $v_{\theta-90^\circ}$ , respectively. The azimuth  $\theta$  is chosen in such a way that the directional derivatives  $X_\theta$ ,  $X_{\theta-90^\circ}$  are uncorrelated, see Remark 2 for some discussions about the choice of  $\theta$ .

**Remark 2** The angle  $\theta$  depends on properties of the covariance matrix  $\Sigma$  of the gradient  $\nabla X(\mathbf{p}, t)$ . Means to estimate the matrix  $\Sigma$  are given in Appendix 3. For several reasons, see [9] for detailed discussion, it is convenient to rotate the coordinate system so that the partial derivatives  $X_x$  and  $X_y$  become uncorrelated.

Let  $A_\theta$  be the rotation by angle  $\theta$  around the  $t$ -axis matrix making covariance between  $X_x$  and  $X_y$  zero. Then let denote by  $\Sigma_\theta$  the covariance matrix of the  $\nabla X$  in the rotated coordinates viz.

$$\Sigma_\theta = A_\theta^T \Sigma A_\theta, \quad (11)$$

where  $A_\theta^T$  is the transpose of  $A_\theta$ . Now the  $\text{Var}(X_\theta)$  is the the element having index 11 in the matrix  $\Sigma_\theta$  while  $\text{Cov}(X_\theta(\mathbf{p}, t), X_t(\mathbf{p}, t))$  has index 13. Using the elements the median velocity  $\mathbf{v}_\theta$  in Equation (10) can be computed once the matrix  $\Sigma_\theta$  has been evaluated.

**Example 1** Let consider the following field

$$X(x, y, t) = \sigma_1 R_1 \cos\left(2\pi \frac{t}{T} - 2\pi \frac{x}{L} + \phi_1\right) + \sigma_2 R_2 \cos\left(2\pi \frac{t}{T} + \phi_2\right) \quad (12)$$

where  $R_1$ ,  $R_2$  and  $\phi_1$ ,  $\phi_2$  are independent variables having Rayleigh, uniform CDF, respectively, and hence  $X$  is a sum of two independent Gaussian fields. The first component is a harmonic wave moving along the  $x$ -axis with velocity  $L/T$  while the second term can be interpreted as colored noise.

Obviously  $X_x$  is independent of  $X_y$  and hence  $\theta = 90^\circ$ , see Remark 2. Further

$$\text{Cov}(X_\theta(\mathbf{p}, t), X_t(\mathbf{p}, t)) = -(2\pi)^2 \sigma_1^2 / (LT), \quad \text{Cov}(X_{\theta-90^\circ}(\mathbf{p}, t), X_t(\mathbf{p}, t)) = 0$$

while  $\text{Var}(X_\theta) = (2\pi)^2 \sigma_1^2 / L^2$  and hence the median velocities Equation (10) are given by

$$\mathbf{v}_\theta = (L/T, 0), \quad \mathbf{v}_{\theta-90^\circ} = (0, 0).$$

In this simple example the median velocities agree with the velocity of the harmonic wave moving along the  $x$ -axis.

In Figure 6, variability of the median velocities  $\mathbf{v}_\theta$  and  $\mathbf{v}_{\theta-90^\circ}$  in Equation (10) is compared. In the top plots seasonal variability of  $\mathbf{v}_\theta$  is illustrated by showing differences between the velocities in February and August. The maximal mean speed in the top plots is about 45 km/h while minimal is zero. Similar comparison for the velocity  $\mathbf{v}_{\theta-90^\circ}$  is given in the bottom plots, where the maximal speed is about 19 km/h. Generally one can say that the storms move faster in winter than in summer, and the angle  $\theta$  also changes between the seasons. For example, in the North Atlantic the storms move basically in average from west to east while in the summer months the direction is opposite in latitude of around 20 degrees.

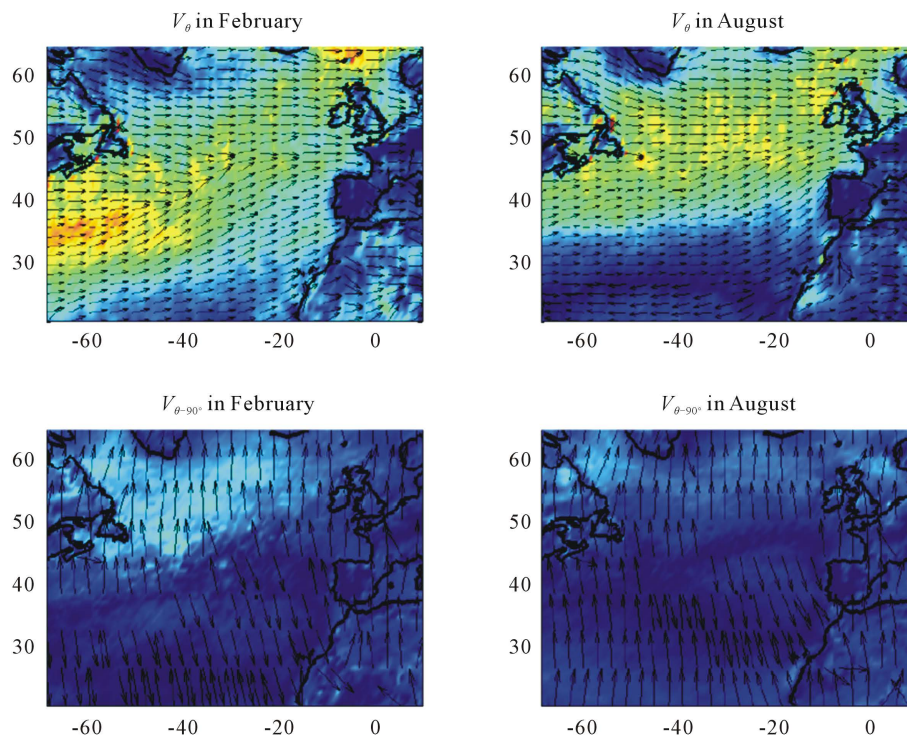
### 4. Statistics of Encountered Wind Speeds

Main subject of the paper is development of a simple model describing variability of wind speeds time series encountered by a vessel or at a fixed location. In this section we will define the model and give means to estimate the long-term CDF of encountered winds; expected duration and strength of an encountered storm.

A ship route is a sequence of positions  $\mathbf{p}_i$ , say, a ship intends to follow. We assume that a ship will follow straight lines between the positions having azimuth  $\alpha_i$ , say. A voyage starts at time  $s$  and will last for  $S$  days. Initial position  $\mathbf{p}(s)$ , azimuths  $\alpha(t)$  and ship speeds  $v^{sh}(t)$ ,  $t \in [s, s+S]$ , define its position  $\mathbf{p}(t)$  at any time  $t$  during a voyage. Then the encountered wind speeds are given by

$$W^e(t) = W(\mathbf{p}(t), t), \quad s \leq t \leq s+S. \tag{13}$$

A ship sailing along a route  $(\mathbf{p}(t), t) = (x(t), y(t), t)$ ,  $t \in [s, s+S]$ , has velocity



**Figure 6.** *Top*—Estimates of the median velocities, km/h, the windy field moves in direction  $\theta$  in February and August. The color corresponds to speed. The highest speed (orange) is 45.1 km/h while the lowest (blue) is 0 km/h. *Bottom*—Comparisons of the median velocities  $\mathbf{v}_{\theta-90^\circ}$  in February and August. The highest speed is 18.6 km/h.

$$\mathbf{v}^{sh}(t) = (\dot{x}(t), \dot{y}(t)) = v^{sh}(t)(\sin \alpha(t), \cos \alpha(t)), \tag{14}$$

where  $v^{sh}(t)$  is the ship speed at time  $t$ . (Recall that the  $x$  axis has azimuth  $90^\circ$  while the  $y$ -axis has azimuth  $0^\circ$ .) In the following we will use the transformed Gaussian field Equation (2) to model the encountered wind speeds  $W^e(t)$ , viz

$$W^e(t) = X(\mathbf{p}(t), t)^{1/a(\mathbf{p}(t))} = X^e(t)^{1/a(t)}. \tag{15}$$

The process  $X^e(t)$  is Gaussian with mean  $m(t) = m(\mathbf{p}(t), t)$  and variance  $\sigma^2(t) = \sigma^2(\mathbf{p}(t), t)$ , respectively.

The long-term CDF of encountered wind speeds is defined by

$$\mathbb{P}(W^e \leq w) = \frac{1}{S} \int_s^{s+S} \mathbb{P}(W^e(t) \leq w) dt, \quad w \geq 0. \tag{16}$$

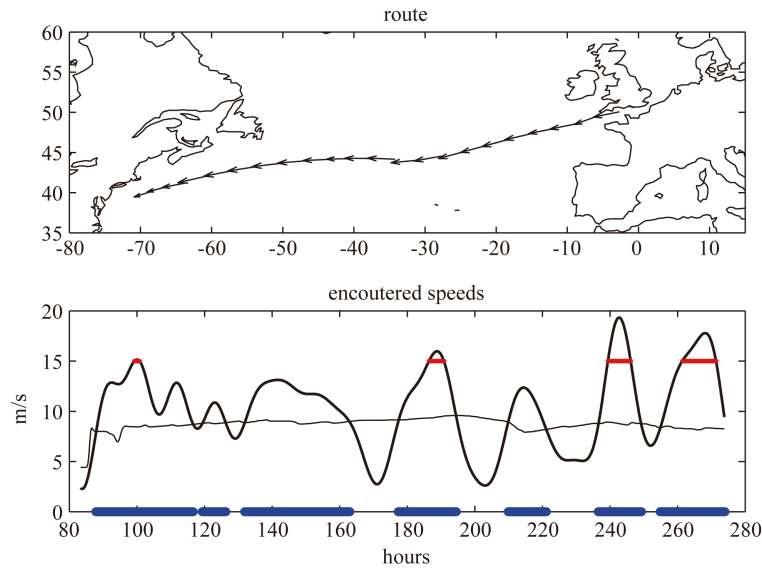
The CDF given in Equation (16) could be estimated by fitting an appropriate distribution to available data. (Weibull distribution is often used.) Alternative approach is to compute the theoretical CDF, viz.

$$\mathbb{P}(W^e \leq w) = \frac{1}{S} \int_s^{s+S} \Phi\left(\frac{w^{a(t)} - m(t)}{\sigma(t)}\right) dt. \tag{17}$$

### 4.1. Distributions of Storm Characteristics

Similarly as in Section 3 we will say that a ship encounters stormy conditions at time  $t$  if wind speed  $W^e(t)$  exceeds some fixed level  $u$ . (In the examples we will use  $u = 15$  m/s.) Similarly, it encounters windy weather conditions at time  $t$  if wind speed is above the median, i.e.  $W^e(t) > \mu(\mathbf{p}(t), t)$ . In **Figure 7** wind speeds encountered along a route in October are presented. The upper intervals mark times in storms while the lower intervals show the periods of windy weather encountered by a vessel. The thin line, shown in the lower plot, illustrates variability of the encountered median wind speed.

The region of stormy conditions consists of time intervals when the wind speed is constantly above threshold  $u$ . The intervals will be called storms. Then let  $N_u$  denote the number of encountered storms. For example,



**Figure 7.** Illustration of the definition of stormy, windy weather regions. *Top*—A route taken in October; *Bottom*—Solid thick line shows the on-board measured wind speeds. The thin solid line presents variability of the median wind speed along the route. The intervals plotted at level 15 m/s represent times when ship encounters the stormy weather while intervals plotted at level zero marks the encountered windy weather regions.



$N_u = 3$  in **Figure 7**. The durations of storms are denoted by  $T_i^{st}$ , while the highest wind speed during a storm by  $A_i^{st}$ ,  $i = 0, \dots, N_u$ . The probability distributions of the characteristics will be defined next. In order to efficiently write down the formulas for the CDFs we need some additional notation introduced next.

Let the number of encountered storms for which event (statement)  $A$  is true be denoted by  $N_u(A)$ . For example,  $N_u(A^{st} > w)$  is the number of encountered storms for which wind speeds exceed a threshold  $w$ , while  $N_u(T^{st} > t)$  is the number of storms that last longer than  $t$ . Obviously  $N_u = N_u(T^{st} > 0)$ , since all  $T_i^{st} > 0$ , is the number of upcrossings of level  $u$  by the encountered wind speeds. The empirical probability that a storm last longer than  $t$  hours can now be written as follows

$$\mathbb{P}^{emp}(T^{st} > t) = \frac{N_u(T^{st} > t)}{N_u}$$

Next the theoretical, based on model, probability of event  $A$ , e.g.  $A = "T^{st} > t"$ , will be defined by

$$\mathbb{P}^u(A) = \frac{\mathbb{E}[N_u(A)]}{\mathbb{E}[N_u]}, \quad \text{e.g. } \mathbb{P}^u(T^{st} > t) = \frac{\mathbb{E}[N_u(T^{st} > t)]}{\mathbb{E}[N_u]}. \quad (18)$$

The proposed model Equation (15) will be validated by comparing the empirical distribution of storms strength  $A^{st}$  and the average durations of storms with theoretically computed  $\mathbb{P}^u(A^{st} > w)$  and  $\mathbb{E}[T^{st}]$ . More complex storms statistics could also be used to validate the model but it would require a dedicated numerical software, see e.g. [10] and references therein, to evaluate  $\mathbb{E}[N_u(A)]$ . Hence it will not be used here. In the following only a simple bound

$$\mathbb{P}(A^{st} > w) \leq \frac{\mathbb{E}[N_w]}{\mathbb{E}[N_u]}, \quad w \geq u, \quad (19)$$

introduced in [11], and the expectations

$$\mathbb{E}[T^{st}] = S \frac{\mathbb{P}(W^e > u)}{\mathbb{E}[N_u]}, \quad \mathbb{E}[T^{cl}] = S \frac{\mathbb{P}(W^e \leq u)}{\mathbb{E}[N_u]}, \quad (20)$$

will be used for validation purposes. In Equation (20),  $T^{cl}$  denotes time period when wind speed is uninterrupted below the threshold  $u$ , i.e. a time period between storms. The Equation (20) will be proved in Appendix 2.

In order to evaluate Equation (19) and Equation (20), the formula for  $\mathbb{E}[N_w]$  is needed. The expected number of upcrossings of level  $w$  by  $W^e$  can be computed using the generalized Rice's formula [12], viz.

$$\mathbb{E}[N_w] = \int_s^{s+S} \int_0^{+\infty} z f_{\dot{W}^e(t), W^e(t)}(z, w) dz dt, \quad (21)$$

see also [13]. Here  $\dot{W}^e(t)$  is the time derivative of  $W^e(t)$ .

**Remark 3** Consider a stationary Gaussian process  $X$  with mean  $m$  and variance  $\sigma^2$ . Let  $N_w$  be the number of upcrossings of level  $w$  by  $X$  in time interval of length  $S$  then classical result of Rice [12] gives

$$\mathbb{E}[N_w] = \frac{S}{2\pi\sigma} \sqrt{\text{Var}(\dot{X}(0))} \exp(-(w-m)^2/2\sigma^2). \quad (22)$$

Consequently the average distance between upcrossing of the mean level  $m$  by  $X$  is

$$\mathbb{E}[T] = 2\pi \frac{\sigma}{\sqrt{\text{Var}(\dot{X}(0))}} \quad (23)$$

## 4.2. Evaluation of $\mathbb{E}[N_w]$

From definition of the encountered wind speed process  $W^e$  it follows that the number of upcrossings of the level  $w$  by  $W^e(t)$  in the interval  $[s, s+S]$  is equal to the number of upcrossings of the level  $w^{a(t)}$  by  $X^e(t)$ . Since we are primarily interested in modeling wind fields in offshore locations we assume that the field is homogeneous in a region with radius of about 100 km and stationary for a period of couple of weeks. (The as-

sumptions are likely to fail in close to coast or inland locations.) Under the assumption  $\dot{m}(t)=0$  and  $X^e(t)$  and  $\dot{X}^e(t)$  are independent. Consequently the integral in Equation (21) can be written in a more explicit way, viz.

$$\mathbb{E}[N_w] = \int_s^{s+S} \frac{1}{2\pi} \frac{\sqrt{\text{Var}(\dot{X}^e(t))}}{\sigma(t)} e^{-\frac{(w^a(t)-m(t))^2}{2\sigma^2(t)}} dt. \quad (24)$$

In the following we shall use an additional parameter  $\tau^e(t)$  defined by

$$\tau^e(t) = \pi \frac{\sigma(t)}{\sqrt{\text{Var}(\dot{X}^e(t))}} \quad (25)$$

and write Equation (24) in an alternative form

$$\mathbb{E}[N_w] = \int_s^{s+S} \frac{1}{2\tau^e(t)} e^{-\frac{(w^a(t)-m(t))^2}{2\sigma^2(t)}} dt. \quad (26)$$

Note that if  $X^e$  is stationary, then  $\tau^e = \mathbb{E}[T]/2$ , see Equation (23). Hence  $\tau^e$  is the average time period that windy conditions last for. Properties and means to evaluate  $\tau^e(t)$  using physically interpretable parameters are discussed in Appendix 1.

## 5. Validation of the Model

The proposed model is validated by investigating the accuracy of the theoretically computed distributions with the empirical distributions estimated from data. Firstly at fixed positions  $p$  the theoretical statistics of the storm characteristics  $A^{st}$ ,  $T^{st}$  and  $T^{cl}$  will be compared with estimates of the statistics derived using ten years of hind-cast data. Secondly, the long-term wind speed distributions encountered by vessels are compared with the theoretically computed distributions using the model and the estimates derived from the hind-cast. The expected number of encountered upcrossing will also be used in the validations. However statistics of encountered storm characteristics will not be used in the validation process. This is because the wind speeds measured on-board ships are biased by captains' decisions to avoid sailing in heavy storms, reported also in [14]. Some validations of the model at inland locations was presented in [15].

### 5.1. Distributions of Storm Characteristics $A^{st}$ , $T^{st}$ and $T^{cl}$ at a Fixed Position

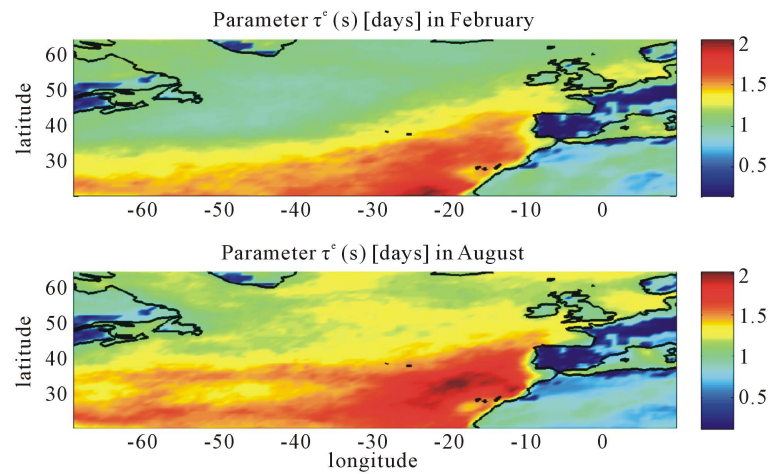
Consider a buoy at position  $\mathbf{p}$  then  $X^e(t) = X(\mathbf{p}, t)$ . The parameter  $\tau^e$ , see Equation (25), is then given by

$$\tau^e(t) = \pi \frac{\sigma(t)}{\sqrt{\text{Var}(X_t(\mathbf{p}, t))}}. \quad (27)$$

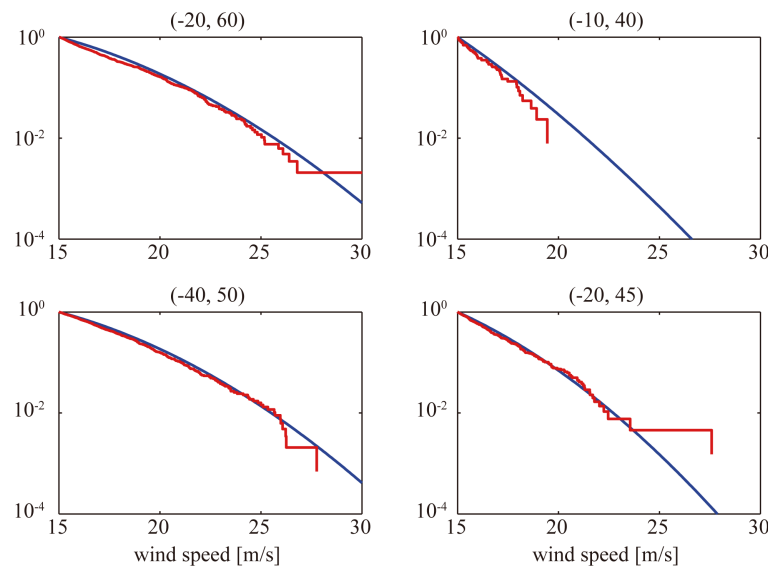
In **Figure 8** values of the parameter  $\tau^e$  evaluated using Equation (27) for February and August are presented. In offshore locations  $\tau^e$  is less than two days, which is much shorter than the stationarity period assumed to be about 3 weeks. Hence the parameter  $\tau^e(t)$  is the expected time period the wind speeds exceeds the median and that  $\tau^e(t)$  is approximately constant for about a month.

The values  $\tau^e$ , presented in **Figure 8**, will be first used to validate the approximation of probability that a storm observed at position  $\mathbf{p}$  will have wind speeds exceeding a level  $w > u = 15$  m/s, *i.e.* formula Equation (19). Next by combining formulas Equation (20) with Equation (16) and Equation (24) the expected duration of a storm will be computed and then compared with the observed average durations. The expected duration of calmer weather, *i.e.* time intervals when winds speeds are constantly below the threshold  $u$ , will be computed in a similar way and then used in the validation

The probabilities  $\mathbb{P}^u(A^{st} > w)$  and expectations  $\mathbb{E}[T^{st}]$ ,  $\mathbb{E}[T^{cl}]$  are computed for a period  $S=1$  year and positions  $\mathbf{p}$  marked by crosses in **Figure 2**. The results presented in **Figure 9** and **Table 1** show very good agreement between the observed storm characteristics at the four locations and the theoretically computed characteristics.



**Figure 8.** Comparison of spatial variability of  $\tau^e(s)$ , defined in Equation (25) for a buoy. (Top) February, (Bottom) August.



**Figure 9.** Probabilities  $\mathbb{P}^u(A^{st} > w)$ ,  $u = 15$  [m/s], that wind in a storm exceeds level  $w$  during one year at four locations having longitudes and latitudes;  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$  and  $(-20, 45)$ . The solid lines are probabilities computed using Equation (19) and Equation (25) with  $a$ ,  $\tau^e(t)$  and  $\sigma(t)$  estimated at the locations. The irregular lines are the estimated probabilities using ten years of hind-cast data.

**Table 1.** Long-term (one year) expected storm/calm durations in days.

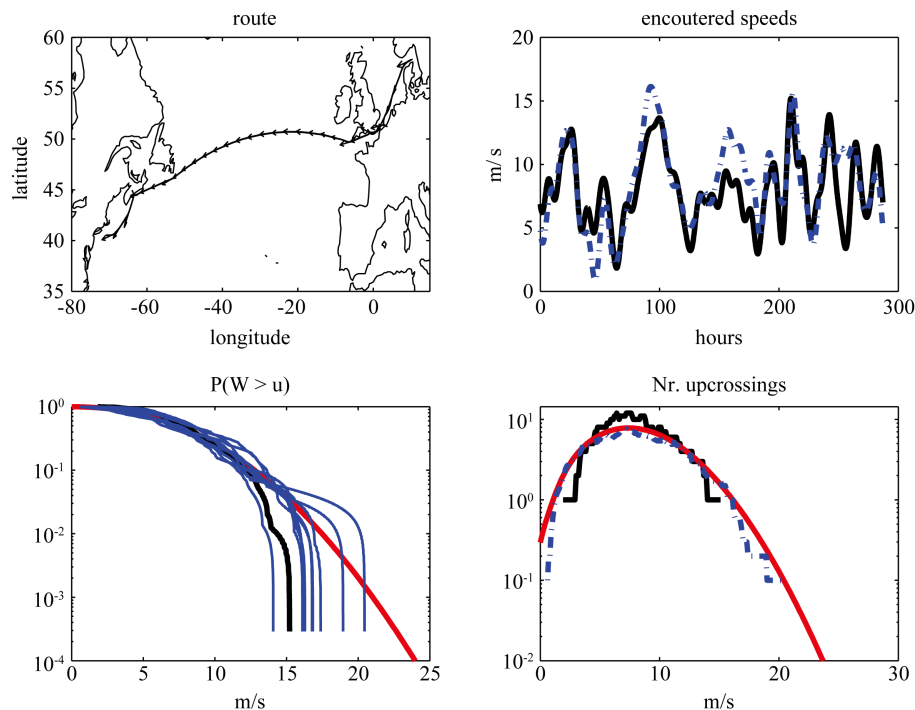
Position	$u = 15$ m/s				$u = 18$ m/s			
	$E[T^{st}]$	$\bar{T}^{st}$	$E[T^{cl}]$	$\bar{T}^{cl}$	$E[T^{st}]$	$\bar{T}^{st}$	$E[T^{cl}]$	$\bar{T}^{cl}$
$(-20, 60)$	0.6	0.5	4.4	4.2	0.5	0.4	13.	11
$(-10, 40)$	0.3	0.4	56	69	0.3	0.3	514	525
$(-40, 50)$	0.6	0.5	4.4	4.2	0.5	0.4	12	11
$(-20, 45)$	0.6	0.5	11	13	0.4	0.4	46	57

## 5.2. Validation-Wind Speeds Encountered by Vessels

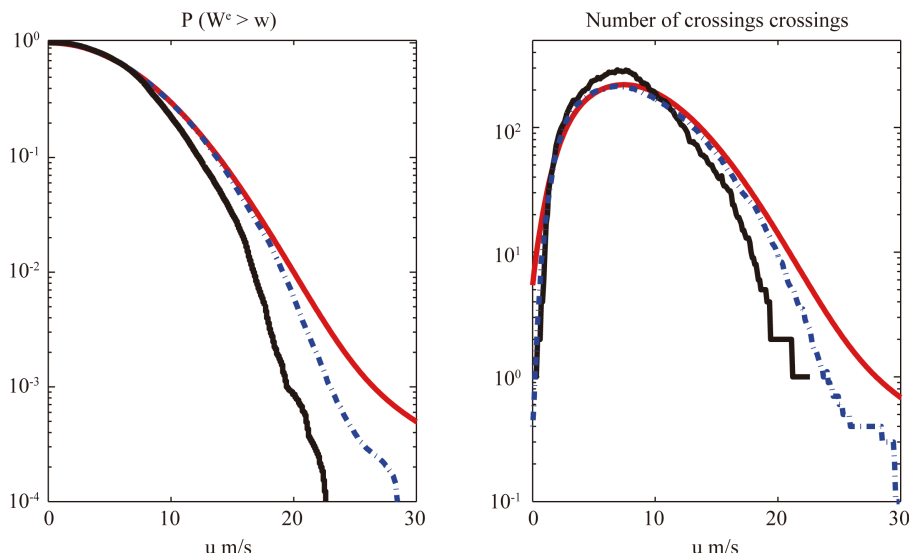
Measurements of the wind speed over ground, *i.e.* ten minutes averages, recorded each ten minutes on-board some ships, are used to validate the proposed model. Since the data are recorded much denser than the hind-cast we have removed high frequencies from the signals (periods above 1.5 hour were removed using FFT). The data used in this study is limited to the North Atlantic and western region of Mediterranean Sea. The accuracy of the theoretically computed long-term distribution of encountered wind speed will be investigated.

First a single voyage operated in late August, shown in the top left plot of **Figure 10**, is considered. In right top plot of the figure, the measured wind speeds, shown as solid line, are compared with the estimated wind speeds using hind-cast, dashed dotted line. One can see that the two signals are reasonably close.

In the left bottom plot of **Figure 10**, ten thin lines show the empirical long term probabilities  $\mathbb{P}(W^e > w)$  computed for hind-cased based estimates of wind-speeds that would be encountered if the ship were sailing the same route every year. One of the estimates is not visible since it is very close to the  $\mathbb{P}(W^e > w)$  estimated using the on-board measured wind speed, the thick solid line. The ten estimates show large variability between years. The regular solid line is the theoretically computed  $\mathbb{P}(W^e > w)$ . It is close to the average of the ten estimates derived from the hind-cast (not shown in the figure). We conclude that for the considered route the theoretical long-term distribution of wind describes well long-term variability of winds along the route. Similar conclusions can be drawn from **Figure 11** left plot were the combined long-term distributions for all 40 voyages are



**Figure 10.** *Top left*—A route sailed from Europe to America in late August. *Top right*—Wind speeds measured on-board a vessel (solid irregular line) compared with their estimates derived from the hind-cast data (dashed dotted line); *Bottom left*—Comparisons of estimates of the long-term probability  $\mathbb{P}(W^e > w)$ , plotted on the logarithmic scale, for the voyage. The thick smooth line is the probability computed by using Equation (6). The less smooth thick line is the probability estimated using the on-board measured wind speeds. The thin irregular lines are the probabilities estimated from the hind-cast data for ten different years; *Bottom right*—Comparisons of the estimates of  $E[N_w]$ , plotted on the logarithmic scale, for the voyage. The thick smooth line is  $E[N_w]$  computed by using Equation (21). The thick irregular line is the  $N_w$  evaluated from the on-board measured wind speeds. The dashed dotted line is the estimate of  $E[N_w]$  using hind-cast derived wind speeds for the route sailed in ten years.



**Figure 11.** *Left*—Comparisons of the estimates of the long-term probability  $\mathbb{P}(W^e > w)$  for the forty voyages. The solid smooth line is the probability computed using Equation (6). The dashed-dotted line is the estimate of  $\mathbb{P}(W^e > w)$  using ten years of hind-cast while the irregular line is the estimate of the wind speeds encountered by vessels. *Right*—Comparisons of the estimates of  $E[N_w]$  for the forty voyages. The solid smooth line is the  $E[N_w]$  computed using Equation (21). The dashed dotted line is an estimate of  $E[N_w]$  using ten years hind-cast while the irregular line is the on-board observed  $N_w$ .

shown. Based on the results presented in **Figure 10** and **Figure 11**, we conclude that the theoretical long term distribution of wind speeds encountered by a sailing vessels agrees well with the distribution derived using hind-cast; and secondly that the routing systems used in planning a route is successful in selecting routes with calmer wind conditions than average one.

In **Figure 10**, bottom right plot, and in **Figure 11**, right plot, estimates of  $E[N_w]$  based on hind-cast (dashed dotted line) and the observed  $N_w$  (the solid irregular line) are compared with the theoretical  $E[N_w]$  computed using Equation (24) for routes shown in **Figure 1** and **Figure 10** top left plot. One can see that the lines are close except for the high wind speeds. The observed crossings of high wind speeds (solid irregular line) are fewer than theoretically predicted. This we attribute to use of routing programs that successfully choose calmer roots than the average one. This claim is also supported by studies of the estimate of  $E[N_w]$  derived from 10 years of hind-cast, shown as the dashed line. One can see that these estimates are higher than on-board observed  $N_w$  for wind speeds above 12 m/s.

### 6. Simulation of the Encountered Wind Speeds

Common experience says that wind speeds vary in different time scales, e.g. diurnal patten due to different temperatures at day and night; frequency of depressions and anti-cyclones which usually occur with periods of about 4 days and annual pattern. To follow the claim the transformed observed wind speed field  $x(\mathbf{p}, t)$  is decomposed into four parts which contain periods above 40 days, between 40 and 5 days, between 5 and 1 day and the noise. For each field variances  $\sigma_i^2$  and the covariance matrices of the gradient vector  $\Sigma_i$  were estimated, i.e. the transformed Gaussian model fitted.

Now for any voyage one can compute parameters  $\tau_i^e(t)$ , then independently simulate the encountered four components along a ship route. Adding the components gives simulation of  $X^e(t)$  and finally transformation  $W^e(t) = X^e(t)^{\frac{1}{a(t)}}$  gives simulated wind speeds that could be encountered along a route.

More precisely, for a ship route  $(\mathbf{p}(t), t)$ ,  $s \leq t \leq s+S$ , one finds parameters  $a(t)$ ,  $m(t)$ ,  $\sigma_i(t)$  and  $\tau_i^e(t)$ ,  $i = 1, \dots, 4$ , then  $X^e(t)$  is simulated by

$$X^e(t) = m(t) + \sum_{i=1}^4 \sigma_i(t) \int_{-\infty}^{+\infty} f_{\tau_i(t)}(t-s) dB_i(s). \tag{28}$$

Here  $B_i$  are independent Brownian motions while the kernels  $f_{\tau_i}$  are given by

$$f_{\tau}(t) = (2/\pi)^{1/4} \frac{1}{\sqrt{\tau}} \exp\left(-\pi^2 \left(\frac{t}{\tau}\right)^2\right).$$

The process  $X^e(t)$  is Gaussian with mean  $m(t)$  and the covariance function given by

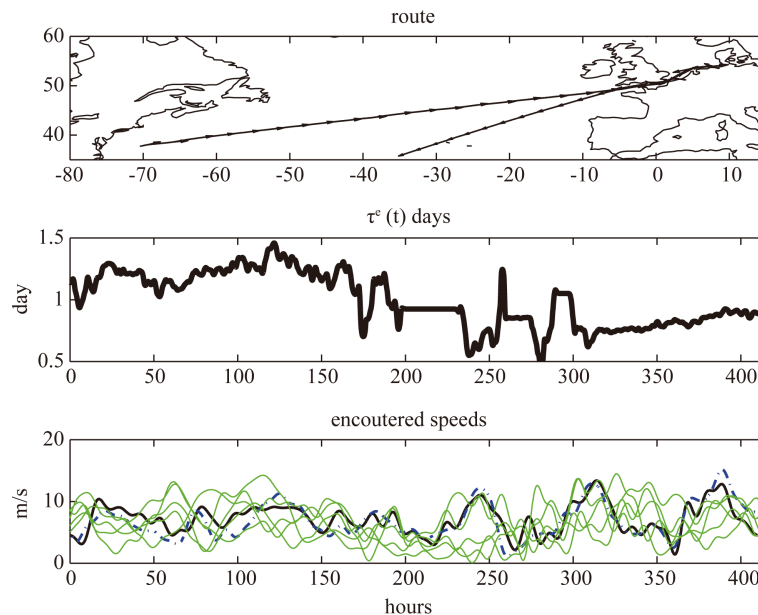
$$\text{Cov}(X^e(t), X^e(s)) = \sum_{i=1}^4 \sigma_i(t) \sigma_i(s) \sqrt{\frac{2\tau_i(t)\tau_i(s)}{\tau_i^2(t) + \tau_i^2(s)}} e^{-\pi^2(t-s)^2 / (\tau_i^2(s) + \tau_i^2(t))}. \tag{29}$$

Obviously the integrals in Equation (28) have to be computed numerically. This is carried out using the following approximation

$$\int_{-\infty}^{+\infty} f_{\tau_i(t)}(t-s) dB_i(s) \approx \sum_j f_{\tau_i(t)}(t-s_j) \sqrt{ds} Z_{ij}, \tag{30}$$

where  $Z_{ij}$ ,  $i = 1, \dots, 4$ , are independent zero mean variance one Gaussian random variables, while  $ds = s_{j+1} - s_j$ . Here  $s_j$  forms an equidistant grid covering the domain of the kernel  $f_{\tau_i}$ . (In the case when wind speeds are simulated on very dense grid then it is recommended to slightly smooth the parameters  $a(t)$ ,  $\sigma_i(t)$  and  $\tau_i^e(t)$ .)

The proposed model gives means for efficient simulation of wind speeds along any ship routes. The parameters  $a(t)$ ,  $m(t)$ ,  $\sigma_i^2(t)$  and  $\tau_i^e(t)$  are specified by means of Equation (28) and Equation (30). Alternatively one can simulate  $X^e(t)$  using covariances defined in Equation (29) and some of many methods to simulate Gaussian vectors. The algorithm based on Equation (30) is preferable when densely sampled wind speeds along a long ship route are needed. For example, for a route defined in Figure 12 top plot that was sailed for 400 hours giving 2400 recorded wind speeds, it took less than 100 seconds on laptop to simulate 100 wind speed profiles along the route. Five of the simulated profiles of  $W^e(t)$  are presented as thin solid lines in



**Figure 12.** *Top*—A route sailed in Northern Atlantic in April; *Middle*—The expected length of encountered windy weather period  $\tau^e(t)$ ; *Bottom*—Wind speeds measured on-board a vessel (solid thick line) hind-cast prediction (dashed dotted line) and five simulations of the wind speeds by means of Equation (15) and Equation (28) (thin solid lines).

**Figure 12** bottom plot. The measured wind speeds are presented as the solid thick line while dashed dotted line is the hind-cast based estimate of the speeds.

Note that parameters  $\sigma^2$  and  $\tau^e$  are simply computable from the parameters  $\sigma_i^2$ ,  $\tau_i^e$  alone. Hence the theoretical long-term distributions and statistics of storm characteristics can be computed by means of methods discussed in previous sections.

## 7. Conclusion

A statistical model for the wind speed field variability in time and over large geographical region has been proposed. The model was fitted to ERA Interim reanalyzed data. Validation tests show very good match between the distributions estimated from the data and the theoretical computed one from the model. The model was also used to estimate risk of encountering extreme winds and the theoretical estimates agree well with the empirical one. Realistic wind profiles can be simulated using the model.

## Acknowledgements

Support of Chalmers Energy Area of Advance is acknowledged. Research was also supported by Swedish Research Council Grant 340-2012-6004 and by Knut and Alice Wallenberg Stiftelse. The authors also would like to thank Wallenius Lines AB for providing us with onboard wind measurement data.

## References

- [1] Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-INTERIM Reanalysis: Configuration and Performance of the Data Assimilation System. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553-597. <http://dx.doi.org/10.1002/qj.828>
- [2] Baxevani, A., Caires, S. and Rychlik, I. (2008) Spatio-Temporal Statistical Modelling of Significant Wave Height. *Environmetrics*, **20**, 14-31. <http://dx.doi.org/10.1002/env.908>
- [3] Monbet, V., Ailliot, P. and Prevosto, M. (2007) Survey of Stochastic Models for Wind and Sea State Time Series. *Probabilistic Engineering Mechanics*, **22**, 113-126. <http://dx.doi.org/10.1016/j.probengmech.2006.08.003>
- [4] Caralis, G., Rados, K. and Zervos, A. (2010) The Effect of Spatial Dispersion of Wind Power Plants on the Curtailment of Wind Power in the Greek Power Supply System. *Wind Energy*, **13**, 339-355. <http://dx.doi.org/10.1002/we.355>
- [5] Kiss, P. and János, I.M. (2008) Limitations of Wind Power Availability over EUROPE: A Conceptual Study. *Nonlinear Processes in Geophysics*, **15**, 803-813. <http://dx.doi.org/10.5194/npg-15-803-2008>
- [6] Brown, B.G., Katz, R.W. and Murphy, A.H. (1984) Time Series Models to Simulate and Forecast Wind Speed and Wind Power. *Journal of Climate and Applied Meteorology*, **23**, 1184-1195. [http://dx.doi.org/10.1175/1520-0450\(1984\)023<1184:TSMSTA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1984)023<1184:TSMSTA>2.0.CO;2)
- [7] Longuet-Higgins, M.S. (1957) The Statistical Analysis of a Random, Moving Surface. *Philosophical Transactions of the Royal Society A*, **249**, 321-387. <http://dx.doi.org/10.1098/rsta.1957.0002>
- [8] Baxevani, A., Podgórski, K. and Rychlik, I. (2003) Velocities for Moving Random Surfaces. *Probabilistic Engineering Mechanics*, **18**, 251-271. [http://dx.doi.org/10.1016/S0266-8920\(03\)00029-8](http://dx.doi.org/10.1016/S0266-8920(03)00029-8)
- [9] Baxevani, A. and Rychlik, I. (2006) Maxima for Gaussian Seas. *Ocean Engineering*, **33**, 895-911. <http://dx.doi.org/10.1016/j.oceaneng.2005.06.006>
- [10] Podgórski, K., Rychlik, I. and Machado, U.E.B. (2000) Exact Distributions for Apparent Waves in Irregular Seas. *Ocean Engineering*, **27**, 979-1016. [http://dx.doi.org/10.1016/S0029-8018\(99\)00030-X](http://dx.doi.org/10.1016/S0029-8018(99)00030-X)
- [11] Rychlik, I. and Leadbetter, M.R. (2000) Analysis of Ocean Waves by Crossing and Oscillation Intensities. *International Journal of Offshore and Polar Engineering*, **10**, 282-289.
- [12] Rice, S.O. (1944) The Mathematical Analysis of Random Noise Part I and II. *Bell System Technical Journal*, **23**, 282-332.
- [13] Rychlik, I. (2000) On Some Reliability Applications of Rice Formula for Intensity of Level Crossings. *Extremes*, **3**, 331-348. <http://dx.doi.org/10.1023/A:1017942408501>
- [14] Mao, W., Ringsberg, J.W., Rychlik, I. and Storhaug, G. (2010) Development of a Fatigue Model Useful in Ship

Routing Design. *Journal of Ship Research*, **54**, 281-293.

- [15] Rychlik, I. and Mustedanagic, A. (2013) A Spatial-Temporal Model for Wind Speeds Variability. Department of Mathematical Sciences, Division of Mathematical Statistics, Chalmers University of Technology and University of Gothenburg, Gothenburg, 1-18. [www.math.chalmers.se/Math/Research/Preprints/2013/8.pdf](http://www.math.chalmers.se/Math/Research/Preprints/2013/8.pdf)



## Appendix 1: Computation of $\mathbb{V}ar(\dot{X}^e(t))$

The parameter  $\tau^e(t)$  was defined in Equation (25), viz.  $\tau^e(t) = \pi\sigma(t)/\sqrt{\mathbb{V}ar(\dot{X}^e(t))}$ . In order to evaluate  $\mathbb{V}ar(\dot{X}^e(t))$  one needs to introduce a time dependent gradient vector  $\nabla X(t) = (X_x, X_y, X_t)(p(t), t)$  and the vector of derivatives

$$\mathbf{v} = (\dot{x}(t), \dot{y}(t), 1). \quad (31)$$

Obviously  $\mathbf{v}^{sh}(t) = (\dot{x}(t), \dot{y}(t))$  and  $\dot{X}^e(t) = \mathbf{v}(t) \cdot \nabla X(t)$ , where  $\cdot$  is the scalar product. Hence

$$\mathbb{V}ar(\dot{X}^e(t)) = \mathbf{v}(t)\Sigma(t)\mathbf{v}(t)^T, \quad (32)$$

where  $\Sigma(t)$  is the covariance matrix of the gradient vector  $\nabla X(t)$ . The matrix  $\Sigma$  has to be estimated in the region of interest. In Appendix 3 a sketch of the estimation procedure is given. In the following we shall give an alternative formula for  $\mathbb{V}ar(\dot{X}^e(t))$  which employs a physically interpretable parameters which could be useful for comparison of suitability of a trade for use of wind sails or other means to harvest wind energy.

### Parameter $\tau^e(t)$ as a Function of Wind, Ship Velocities and Geometrical Sizes of Windy Regions

The variance  $\mathbb{V}ar(\dot{X}^e(t))$  is independent of the choice of coordinate system. Here we will use the rotated coordinate system by azimuth  $\theta$  called in Section 3 main azimuth of a storm. The matrix  $\Sigma$  in the rotated coordinate system will be denoted by  $\Sigma_\theta(t)$  and has the following diagonal elements

$$\sigma_{200} = \mathbb{V}ar(X_\theta(\mathbf{p}, t)), \quad \sigma_{020} = \mathbb{V}ar(X_{\theta-90^\circ}(\mathbf{p}, t)), \quad \sigma_{002} = \mathbb{V}ar(X_t(\mathbf{p}, t)), \quad (33)$$

and following off-diagonal elements

$$\sigma_{110} = 0, \quad \sigma_{101} = \mathbb{C}(X_\theta(\mathbf{p}, t), X_t(\mathbf{p}, t)), \quad \sigma_{011} = \mathbb{C}(X_{\theta-90^\circ}(\mathbf{p}, t), X_t(\mathbf{p}, t)). \quad (34)$$

The ships velocity  $\mathbf{v}^{sh}(t) = v^{sh}(t)(\sin \alpha(t), \cos \alpha(t))$  is in the rotated coordinates given by

$$\mathbf{v}_\theta^{sh}(t) = v^{sh}(t)(\sin(\alpha(t) - \theta(t)), \cos(\alpha(t) - \theta(t))).$$

Obviously  $\mathbb{V}ar(\dot{X}^e(t)) = [\mathbf{v}_\theta^{sh}(t), 1]\Sigma_\theta(t)[\mathbf{v}_\theta^{sh}(t), 1]^T$  and after some algebra

$$\mathbb{V}ar(\dot{X}^e(t)) = \sigma_{200}(v_\theta^e)^2 + \sigma_{020}(v_{\theta-90^\circ}^e)^2 + \sigma_{002}\left(1 - \frac{\sigma_{200}}{\sigma_{002}}v_\theta^2 - \frac{\sigma_{020}}{\sigma_{002}}v_{\theta-90^\circ}^2\right), \quad (35)$$

where the encountered velocity, e.g. the difference between the ship velocity and the wind field velocity is, in the rotated coordinates, given by

$$(v_\theta^e, v_{\theta-90^\circ}^e) = (v_{\theta(t)} - v^{sh}(t)\sin(\alpha(t) - \theta(t)), v_{\theta-90^\circ}(t) - v^{sp}(t)\cos(\alpha(t) - \theta(t))). \quad (36)$$

In order to interpret components in Equation (35), we need to introduce some additional parameters that describe average size of windy weather regions and some irregularity factors.

Recall that windy weather conditions region at time  $t$  is the region consists all  $\mathbf{p}$  where wind speeds exceeds the median  $\mu(\mathbf{p}, t)$ . Now we shall introduce parameters related to average size of windy region in directions  $\theta$  and  $\theta - 90^\circ$ . The parameters will be denoted by  $L_\theta$  and  $L_{\theta-90^\circ}$ , respectively. The third parameter  $T$  is the average period the windy weather last lasts at a fixed position  $\mathbf{p}$ . The parameters are defined by

$$L_\theta = \pi \frac{\sigma}{\sqrt{\sigma_{200}}}, \quad L_{\theta-90^\circ} = \pi \frac{\sigma}{\sqrt{\sigma_{020}}}, \quad T = \pi \frac{\sigma}{\sqrt{\sigma_{002}}}, \quad (37)$$

see Equation (22) and Remark 3. Obviously the values of parameters are slowly changing functions of position and time and that why we call them local sizes of windy regions. However if the field  $X(\mathbf{p}, t)$  were homoge-

neous and stationary then the parameters would be equal to the average length between upcrossings and downcrossing of the median by wind speed  $W$  in direction  $\theta$ ,  $\theta + 90^\circ$  and in time, see [9] for details.

Now by multiplying both sides of the Equation (35) by  $(2\pi)^2$  and dividing by  $\sigma^2$ , we obtain that

$$\frac{1}{(\tau^e)^2} = \left(v_\theta^e/L_\theta\right)^2 + \left(v_{\theta-90^\circ}^e/L_{\theta-90^\circ}\right)^2 + (1/T)^2 \cdot (1 - \alpha_\theta^2 - \alpha_{\theta-90^\circ}^2),$$

where

$$\alpha_\theta^2 = \frac{\sigma_{101}^2}{\sigma_{200}\sigma_{002}}, \quad \alpha_{\theta-90^\circ}^2 = \frac{\sigma_{011}^2}{\sigma_{020}\sigma_{002}} \quad (38)$$

are useful irregularity factors. Roughly, smaller values of the factors higher risks of extreme storms, see [9] for more details. Further, if  $\alpha_\theta^2 + \alpha_{\theta-90^\circ}^2 = 1$  then the surface  $X$  drifts, viz

$$X(\mathbf{p}, t) = X(\mathbf{p} - \mathbf{v}^{dr} t, 0).$$

If  $\mathbf{p}$  has rotated coordinates then  $\mathbf{v}^{dr}$  in rotated coordinates is equal to  $(v_\theta^e, v_{\theta-90^\circ}^e)$ . Finally

$$\tau^e(t) = \frac{1}{\sqrt{\left(v_\theta^e/L_\theta\right)^2 + \left(v_{\theta-90^\circ}^e/L_{\theta-90^\circ}\right)^2 + (1/T)^2 \cdot (1 - \alpha_\theta^2 - \alpha_{\theta-90^\circ}^2)}}. \quad (39)$$

For a homogeneous wind field  $v_\theta^e/L_\theta$  is the reciprocal of the average time spend in windy weather region when sailing with azimuth  $\theta$ , similar interpretation can be given to  $v_{\theta-90^\circ}^e/L_{\theta-90^\circ}$  while  $T$  is the average time the windy weather is observed by a ship at rest or a buoy. These parameters can be estimated from the on-board measured signals or given subjective values based on experience. Usefulness of Equation (39) and Equation (24) lies in possibility of predicting risks for encountering extreme storms using easily available parameters which have clear physical meaning.

## Appendix 2: Proof of Equation (20)

Let assume that  $W^e(s)$  is a smooth process. Using Fubini's theorem

$$\mathbb{E}[T^{st}] = \frac{\int_0^{+\infty} \mathbb{E}[N_u(T^{st} > t)] dt}{\mathbb{E}[N_u]} = \frac{\mathbb{E}\left[\int_0^{+\infty} N_u(T^{st} > t) dt\right]}{\mathbb{E}[N_u]}.$$

Since  $\int_0^{+\infty} N_u(T^{st} > t) dt = \int_s^{s+T} \mathbf{1}_{\{W^e(t) \geq u\}} dt$ , where  $\mathbf{1}_A(x)$  is the indicator function of the set  $A$  taking value 1 if  $x \in A$  and zero otherwise. Again by Fubini's theorem

$$\mathbb{E}\left[\int_s^{s+T} \mathbf{1}_{\{W^e(t) \geq u\}} dt\right] = \int_s^{s+T} P(W^e(t) > u) dt$$

and hence

$$\mathbb{E}[T^{st}] = T \frac{P(W^e > u)}{\mathbb{E}[N_u]}.$$

## Appendix 3: Estimation of Parameters

The parameters of the model have been fitted for the North Atlantic. Here the ERA Interim data has been used, although in future work we plan to also use data from satellite based sensors. A moment's method and regression fit were employed to estimate the parameters. In this section we give a short description of the applied estimation procedure. In the following the measured wind speeds at a location will be denoted by  $w(t)$ .

Step 1: For a fixed geographical location and  $0 < a < 1$  the transformed wind speed  $w(t)^a$  is computed and the mean Equation (3) fitted using LS regression. Empirical cumulative distribution function (CDF) and Gaus-

sian (CDF) are fitted to the residual  $w(t)^a - m(t)$ . Parameter  $a^*$  minimizing the distance between the two distributions is selected as an estimate of  $a$ . The corresponding mean  $m^*(t)$  is an estimate of  $m(t)$ . Further the residual  $x(t) = w(t)^{a^*} - m^*(t)$  is evaluated and then used to estimate parameters  $\sigma_i^2(t)$  in the following steps.

Step 2: Estimation of signals  $x_i(t)$ ,  $i = 1, \dots, 4$ . The signal  $x_1$  is estimated as follows; first one filters out from  $x(t)$  (see Step 1) the harmonics with periods shorter than 40 days. The resulting signal is an observation of  $x_1(t)$ . The signal  $x_2(t)$  is derived by filtering out harmonics with periods below 5 days from the signal  $x(t) - x_1(t)$ . The signal  $x_3(t)$  is derived by filtering out harmonics with periods below 1 day from the signal  $x(t) - x_1(t) - x_2(t)$ . Finally,  $x_4(t) = x(t) - x_1(t) - x_2(t) - x_3(t)$ .

Step 3: For a signal  $x_i(t)$  the parameters  $\sigma_i^2(t)$  are estimated as follows. For a sequence of times  $t_j$ , assuming stationarity of  $x_i(s)$  for  $s$  in a neighborhood of 10 days around  $t_j$ , estimates of  $\sigma_i(t_j)$  are found. Then  $\sigma_i^2(t)$  are estimated by fitting seasonal components, similar to Equation (3), to sequences of observations  $(t_j, \sigma_i(t_j)^2)$ .

Step 4: Estimation of  $\Sigma_i(\mathbf{p}, t)$ , *i.e.* the covariance matrix of the gradient vector evaluated at  $(\mathbf{p}, t)$ . The covariance matrix is defined by six covariances between the partial derivatives of  $X_i$ . The functions are changing slowly with season but spatial variability can be high, particularly at coastal and inland locations. Consequently we fit six seasonal components to the covariances for each of positions  $\mathbf{p}$  on a grid with mesh 0.75 degree. The components are estimated in a similar way as discussed in Step 3.