



**HAL**  
open science

## Spatio-temporal model for wind speed variability

Igor Rychlik

► **To cite this version:**

Igor Rychlik. Spatio-temporal model for wind speed variability. Annales de l'ISUP, 2015, 59 (1-2), pp.25-56. hal-03604750

**HAL Id: hal-03604750**

**<https://hal.science/hal-03604750>**

Submitted on 10 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Spatio-temporal model for wind speed variability

Igor Rychlik

*Chalmers University of Technology*

**Abstract:** Investments in wind energy harvesting facilities are often high. At the same time uncertainties for the corresponding energy gains are large. Therefore a reliable model to describe the variability of wind speed is needed to estimate the expected available wind power and other statistics of interest, e.g. coefficient of variation, expected length of the wind conditions favorable for the wind-energy harvesting etc. In this paper wind speed 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 speed in ERA Interim reanalysis data set.

### 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 speed 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  $\mathbf{p}$

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

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 [9] 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 [6].

Modeling spatial and temporal dependence of wind speed is a very complex problem. Models proposed in the literature are reviewed in [13]. Here we will use the transformed Gaussian model, which assumes that there exists a deterministic function  $G(w)$  such that  $X(\mathbf{p}, t) = G(W(\mathbf{p}, t))$  is Gaussian. (The transformations parameters can depend on location  $\mathbf{p}$  in a smooth way, see (2) and Figure 1.) 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 reanalysis one could estimate the covariance for any pair  $(\mathbf{p}_1, t_1), (\mathbf{p}_2, t_2)$ , see e.g. [8], [10]. 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. For wind energy harvesting the minimal requirements on the model are that it should provide: a correct estimates of long-term distributions of the wind speed; accurate predictions of average durations of the extreme winds conditions and reliable estimates of CDFs of top speed during storms.

We are primarily interested in modeling wind fields in offshore locations. The data are 10 minutes wind speed averages, sampled at frequencies ranging from 10 minutes to 6 hours. Our general assumptions are that the field is homogeneous in a region with radius of about 200-300 km (the assumption is likely to fail in close to coast or inland locations) and stationary for a period of 3-4 weeks. Consequently we shall first propose a temporal models valid at fixed locations. Then the velocities of wind field movement, which are functions of wind speed gradient covariances, will be used to define local spatio-temporal models. Finally the local models are employed to define a global spatio-temporal model by means of moving average field having kernels dependent on the local parameters, e.g. variance and velocities of storm movements.

The paper is organized as follows. In Section 2 a transform Gaussian model for wind speed variability in time and space is presented. Then statistical properties of storm characteristics are investigated in Section 3. The correlation structure of the time series and means to simulate the wind speed are discussed in Section 4. Next local homogeneous spatio-temporal model for wind speed is presented in Section 5. One of the most important characteristics of the model, the velocity of storm movements, is introduced and exemplified in Section 6. In Section 7 the global non-homogeneous model for wind speed field is proposed and validated by means of comparisons of the simulated wind fields and the one given by ERA Interim data. Finally in Section 8 the model is applied to study statistical properties of the encountered wind speed. Paper closes with two appendixes: Appendix A contains somewhat more technical matters while in Appendix B estimation of spatio-temporal model parameters is discussed.

## 2. Transformed Gaussian model and long-term CDFs.

In this section we shall introduce the transformed Gaussian model for the variability of wind speed. 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 location  $\mathbf{p}$ , viz.

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

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. The choice of the transformation (2) was inspired by the classical paper [7] in which the authors modeled wind speed data at offshore location. In North Atlantic the estimates of the parameter  $a$  are between 0.5 and 1 see Figure 1. In inland locations the estimates are much smaller and the Box-Cox transformation could be a better alternative to (2). The Box-Cox transformation was not used in this work.

Mean of  $X(\mathbf{p}, t)$ , denoted by  $m(\mathbf{p}, t)$ , depends both on position and time. The temporal variability of the mean is approximated by seasonal components with trends defined as follows

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

Here  $t$  has units years.

Common experience says that wind speed vary in different time scales, e.g. diurnal pattern 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 we propose to model  $X(\mathbf{p}, t)$  as a sum of four independent Gaussian processes  $X_i(\mathbf{p}, t)$  having variances  $\sigma_i^2(\mathbf{p}, t)$ ,  $i = 1, \dots, 4$ , i.e.

$$(4) \quad X(\mathbf{p}, t) = \sum_{i=1}^4 X_i(\mathbf{p}, t), \quad E[X_1(\mathbf{p}, t)] = m(\mathbf{p}, t), \quad \text{Var}(X_i(\mathbf{p}, t)) = \sigma_i^2(\mathbf{p}, t)$$

and  $E[X_i(\mathbf{p}, t)] = 0$  for  $i \geq 2$ . Similarly as  $m(\mathbf{p}, t)$  also logarithms of the variances  $\sigma_i^2(\mathbf{p}, t)$  are modeled using seasonal components with trends

$$(5) \quad \ln(\sigma_i^2(\mathbf{p}, t)) = b_{i0}(\mathbf{p}) + b_{i1}(\mathbf{p}) \cos(2\pi t) + b_{i2}(\mathbf{p}) \sin(2\pi t) + b_{i3}(\mathbf{p}) t.$$

In the following we will write  $\sigma^2(\mathbf{p}, t) = \sum_{i=1}^4 \sigma_i^2(\mathbf{p}, t)$ .

The last terms in (3) and (5) are trends which could be used to investigate possibility of slow increase of the average wind speed, or the change of variance, over years which are important issues when estimating safety of maritime operations. This type of problems are outside of the scope of this paper and have not been studied.

## 2.1. Estimates of the parameters

Ten years of data  $W(\mathbf{p}, t)$  were used to estimate parameters  $a$ ,  $m_j$  and  $b_{ij}$ ,  $i = 1, \dots, 4$  and  $j = 0, \dots, 3$ , in (3-5) in North Atlantic on a grid of 0.75 degree. Figure 1 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 speed distribution from the Gaussian one.

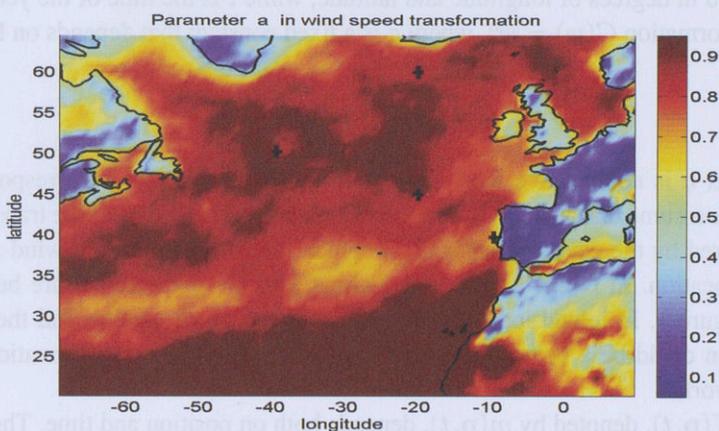


FIGURE 1. Values of parameter  $a$  in the transformation (2).

In the right plots of Figure 2, the standard deviations  $\sigma(\mathbf{p}, t)$ , defined in (4), are presented for February and August, respectively. One can see that the standard deviations change considerably with the geographical location but are less dependent on seasons. We turn next to presentation of variability of the parameter  $m(\mathbf{p}, t)$ , i.e. the mean of  $X(\mathbf{p}, t)$  defined in (2).

Since units of  $m$  are not physical we choose to show the variability of the median speed

$$(6) \quad \mu(\mathbf{p}, t) = m(\mathbf{p}, t)^{1/a(\mathbf{p})}$$

instead. Values of the median for February and August are presented in two left plots of Figure 2. As expected, wind speed are higher in winter than in summer.

## 2.2. Validation of Gaussianity of $X(\mathbf{p}, t)$ at offshore locations

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,  $X(\mathbf{p}, t)$  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.)

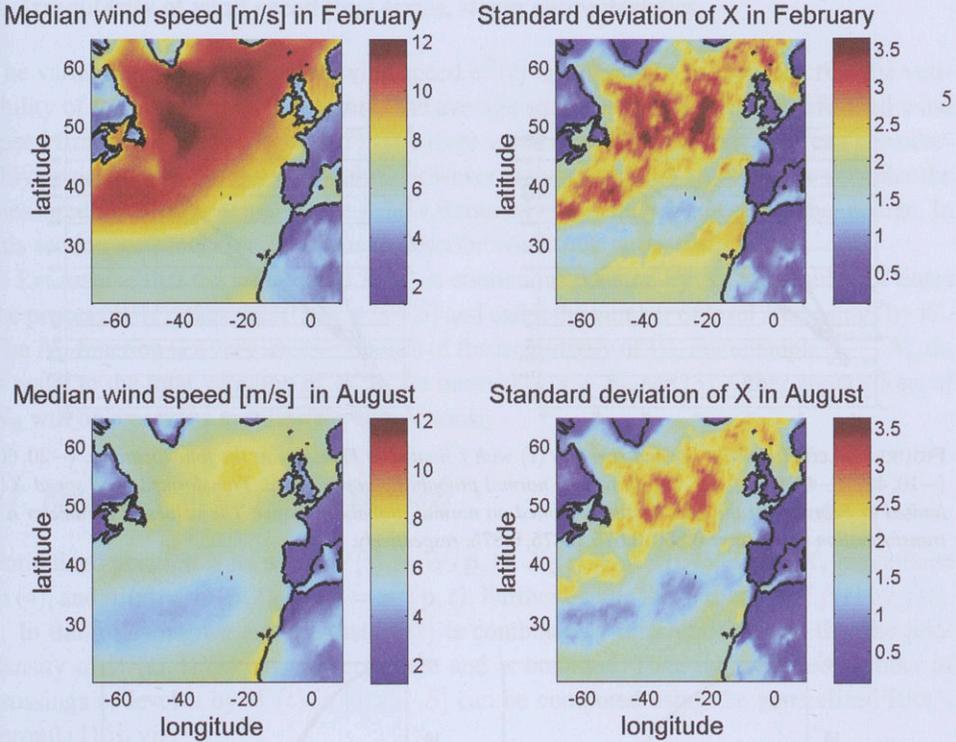


FIGURE 2. Left top - Median wind speed  $\mu$  [m/s], defined in (6), in February. Left bottom - Median wind speed in August. Right top - Standard deviation of  $X$ , computed by means of (5) in February. Right bottom - The standard deviation in August.

In the right plot of the figure the transformed data  $X = G(W)$  is plotted on the normal probability paper. One can see that, except of the region of small wind speed,  $X$  CDFs are well approximated by the Gaussian distributions.

Finally we check whether the regressions (3-5) used to model seasonal variability of  $m$  and  $\sigma^2 = \sum_{i=1}^N \sigma_i^2$  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 speed at a fixed position  $\mathbf{p}$ , defined in (1), is given by

$$(7) \quad \mathbb{P}(W \leq w) = \frac{1}{S} \int_s^{s+S} \Phi \left( \frac{w^{a(\mathbf{p})} - m(\mathbf{p}, t)}{\sigma(\mathbf{p}, t)} \right) dt,$$

where  $\Phi(x)$  is the CDF of a standard Gaussian (normal) variable. In Figure 4 the yearly probabilities for wind speed  $\mathbb{P}(W > w)$  computed using (7) at four locations in North Atlantic are compared with the empirical estimates. (The locations are marked by crosses in Figure 1.) One can see that the agreement between the estimates is excellent.

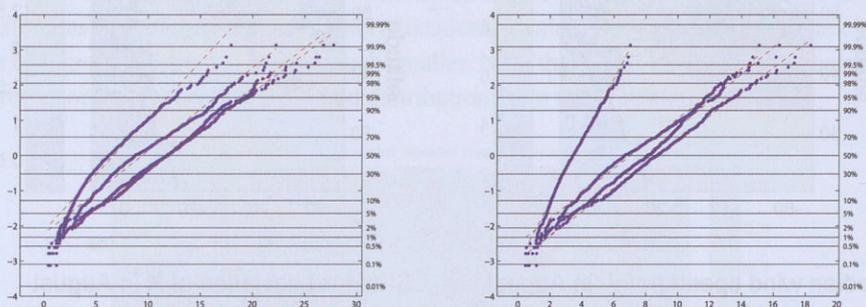


FIGURE 3. Left: Ten years of wind speed  $W(t)$  with  $t$  limited to February at the four locations.  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$ ,  $(-20, 45)$  plotted on normal probability paper. Right: Transformed wind speed  $X(t)$  limited to February at the four locations plotted on normal probability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

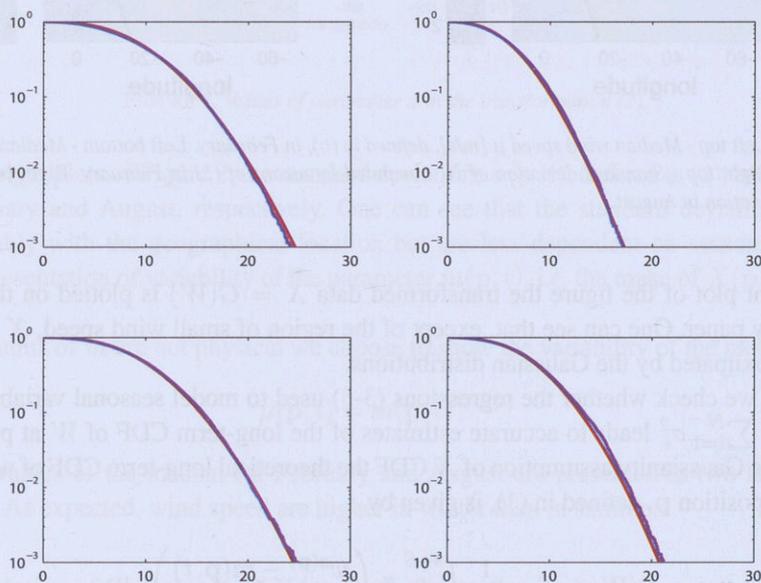


FIGURE 4. Comparisons of estimates of the long-term probability  $\mathbb{P}(W > w)$  for yearly wind speed variability (1) at four locations defined in degrees of longitude and latitude;  $(-20, 60)$ ,  $(-10, 40)$ ,  $(-40, 50)$  and  $(-20, 45)$ . The solid line is the probability computed using (7) with  $S = 1$  year. Somewhat more irregular lines are the estimated probabilities based on ten years of reanalysis data.

### 3. Irregularity of wind speed time series, storm characteristics

The variance of the transformed wind speed  $\sigma^2(t)$  is commonly used to describe the variability of the wind speed. It measures the average square distance of the transformed wind speed from the mean value  $m(\mathbf{p}, t)$ . The large variance means that the speed can considerably deviate from the expected value. However the variance does not tell how irregular the measured time series actually is, e.g. how frequently the wind speed crosses the median. In this section we present other means to describe wind time series variability.

Let assume that the wind speed  $W(t)$  is continuous. Denote by  $N_u$  the number of times the process  $W(t)$  takes value  $u$  in  $[s, s+S]$  and call it the number of level  $u$  crossings by  $W$ . The  $N_u$  function is a very useful measure of the irregularity of  $W$ . For example  $\int_{-\infty}^{+\infty} N_u du$  is equal to the total variation of  $W$  in the interval  $[s, s+S]$ , see [3]. Other applications of  $N_u$  will be presented in following subsections.

#### 3.1. Expected number of level $u$ crossings

For a fixed position  $\mathbf{p}$  let write  $W(t) = W(\mathbf{p}, t)$ ,  $X_i(t) = X_i(\mathbf{p}, t)$ , where  $X_i$  are defined in (4), and  $m(t) = m(\mathbf{p}, t)$ ,  $\sigma_i^2(t) = \sigma_i^2(\mathbf{p}, t)$ . Further denote the derivative of  $f(t)$  by  $\dot{f}(t)$ .

In the following we assume that  $W(t)$  is continuously differentiable and that the joint density of  $W(t), \dot{W}(t)$  exists everywhere and is bounded. Then the expected number of crossings of level  $u$  by  $W(t)$  in  $[s, s+S]$  can be computed using the generalized Rice's formula [15], viz.

$$(8) \quad E[N_u] = \int_s^{s+S} \int_{-\infty}^{+\infty} |z| f_{\dot{W}(t), W(t)}(z, u) dz dt.$$

Now by (2)  $N_u$  is equal to the number of times  $X(t)$  takes the value  $u^a$ , where  $a = a(\mathbf{p})$ . In Section 1 we have assumed that  $W(t)$ , and hence also  $X(t)$ , are stationary for a period of few weeks. This means that  $\text{Cov}(X(t_1), X(t_2))$  is well approximated by a function of  $t_2 - t_1$  while  $E[X(t_1)] \approx E[X(t_2)]$ . Consequently the expected value of the derivative  $\dot{X}(t)$  is approximately zero and  $X(t)$  and  $\dot{X}(t)$  are approximately uncorrelated, see [1], [2], for more details. Denote by  $\dot{\sigma}^2(t)$  the variance of  $\dot{X}(t)$ . Using these notation the integral in (8) can be written in a more explicit way, viz.

$$(9) \quad E[N_u] = \int_s^{s+S} \frac{1}{\pi} \frac{\dot{\sigma}(t)}{\sigma(t)} e^{-\frac{(u^a - m(t))^2}{2\sigma^2(t)}} dt = \int_s^{s+S} \frac{1}{\tau(t)} e^{-\frac{(u^a - m(t))^2}{2\sigma^2(t)}} dt,$$

where

$$(10) \quad \tau(t) = \pi \frac{\sigma(t)}{\dot{\sigma}(t)}.$$

Basically  $1/\tau(t)$  is the intensity of points  $t$  solving  $X(t) = m(t)$ , i.e. intensity of times the wind speed  $W(t)$  is equal to the median value. If  $W(t)$  were a stationary process then  $\tau$  would be equal to the average duration of the period when wind speed is uninterruptedly

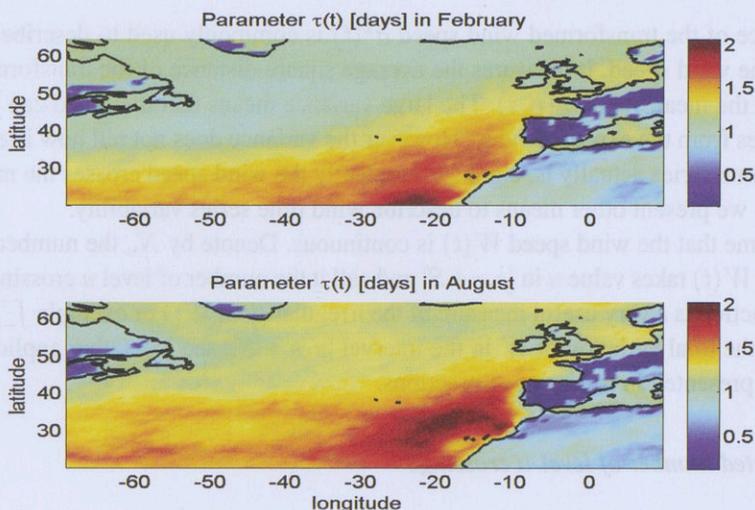


FIGURE 5. Comparison of spatial variability of  $\tau(t)$ , defined in (10) at fixed location. (Top) - February, (Bottom) August.

above the median speed. Such a period will be called the region of windy conditions. For locally stationary process  $\tau(t)$  is interpreted as the local duration of windy conditions.

Estimates of the parameter  $\tau(t)$  for February and August are presented in Figure 5. In offshore locations  $\tau$  is less than two days, which is much shorter than the stationarity period assumed to be about 3 weeks. Hence the parameter  $\tau(t)$  is practically equal to the expected time period the wind speed exceeds the median.

### 3.2. Distributions of storm characteristics

We define stormy weather at position  $\mathbf{p}$  and time  $t$  if wind speed  $W(t)$  exceeds some fixed level  $u$ . (In the examples we will use  $u = 15$  m/s.) Similarly, we define windy weather conditions at time  $t$  if wind speed is above the median, i.e.  $W(t) > \mu(\mathbf{p}, t)$ . The region of stormy conditions consists of time intervals when wind speed are constantly above threshold  $u$ . The intervals will be called storms. Let  $N$  denote the number of storms that occurred in  $[s, s + S]$ . Durations of storms are denoted by  $T_i^{st}$ , while the highest wind speed during a storm by  $A_i^{st}$ ,  $i = 0, \dots, N$ . (Here we allow  $N$  to be zero and, for consistence of notation, we let  $T_0^{st} = 0$ ,  $A_0^{st} = 0$ .) 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 storms for which event (statement)  $A$  is true be denoted by  $N(A)$ . For example,  $N(A^{st} > w)$  is the number of storms for which wind speed exceed a threshold  $w$ ,

while  $N(T^{st} > t)$  is the number of storms that last longer than  $t$ . The empirical probability that a storm last longer than  $t$  hours can now be written as follows

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

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

$$(11) \quad \mathbb{P}(A) = \frac{E[N(A)]}{E[N]}, \text{ e.g. } \mathbb{P}(T^{st} > t) = \frac{E[N(T^{st} > t)]}{E[N]}.$$

The proposed model (2) will be validated by comparing the empirical distribution of storms strength  $A^{st}$  and the average durations of storms with theoretically computed  $\mathbb{P}(A^{st} > w)$  and  $E[T^{st}]$ . More complex storm statistics could also be used to validate the model but it would require a dedicated numerical software, see e.g. [14] and references therein, to evaluate  $E[N(A)]$ . Hence it will not be used here. In the following only a simple bound

$$(12) \quad \mathbb{P}(A^{st} > w) \leq \frac{E[N_w]}{E[N_u]}, \quad w \geq u,$$

introduced in [16], [17], will be used for validation purposes. Since  $E[N_u] = 2E[N]$  hence the expectations

$$(13) \quad E[T^{st}] = S \frac{\mathbb{P}(W > u)}{E[N_u]/2}, \quad E[T^{cl}] = S \frac{\mathbb{P}(W \leq u)}{E[N_u]/2}.$$

In (13),  $T^{cl}$  denotes time period when wind speed is uninterruptedly below the threshold  $u$ , i.e. a time period between storms. The formulas (13) are proved in Appendix A.

### 3.3. Validation of estimates (12-13) at offshore and close to coast locations

In order to further validate the proposed transformed Gaussian model for wind speed variability we will compare the theoretical statistics of the storm characteristics  $A^{st}$ ,  $T^{st}$  and  $T^{cl}$  with estimates of the statistics derived using ten years of reanalysis data. The data at four positions  $\mathbf{p}$ , marked by crosses in Figure 1, will be used.

The values  $\tau$ , presented in Figure 5, are used to evaluate expected number of crossings  $E[N_w]$  at the four positions. Then  $A^{st}$  CDF is estimated using formula (12) for the period  $S = 1$  year. In Figure 6 the theoretically derived estimates (smooth solid lines) are compared with the empirical distributions (irregular lines). One can see that the estimates of the probabilities  $\mathbb{P}(A^{st} > w)$  agree very well with empirical probabilities at the four locations.

Next by combining formulas (13) with (7) and (9) the expected duration of a storm is computed and then compared with the observed average durations. The expected duration of calmer weather, i.e. time intervals when winds speed are constantly below the threshold  $u$ , is computed in a similar way. The expectations  $E[T^{st}]$ ,  $E[T^{cl}]$  are computed for a period  $S = 1$  year and presented in Table 1. The results shown in the table and Figure 6 demonstrate a very good agreement between the observed storm characteristics at the four locations and the theoretically computed one.

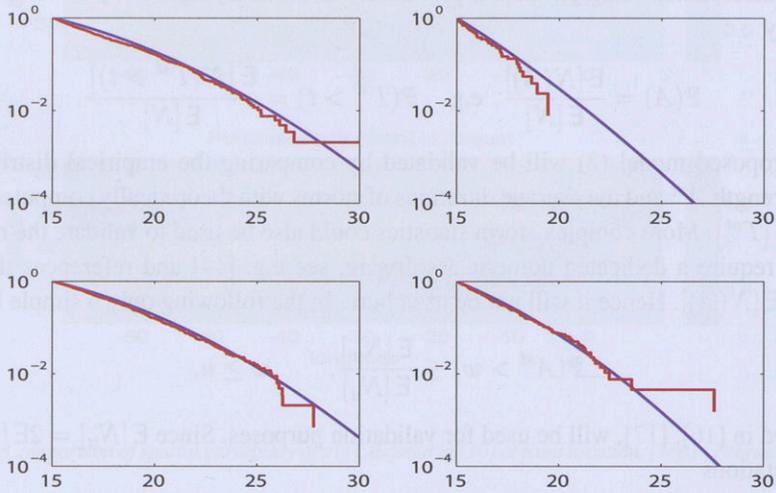


FIGURE 6. Probabilities  $\mathbb{P}(A^{st} > w)$ ,  $u = 15$  [m/s], that wind in a storm exceeds level  $w$  during one year at four locations having longitudes and latitudes;  $(-20,60)$ ,  $(-10,40)$ ,  $(-40,50)$  and  $(-20,45)$ . The solid lines are probabilities computed using (12) and (9) with  $a$ ,  $\tau(t)$  and  $\sigma(t)$  estimated at the locations. The irregular lines are the estimated probabilities using ten years of reanalysis 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

#### 4. The covariances of $X_i(t)$ - simulation of wind speed

In this paper we are primarily interested in a simple model for  $X(t)$ ,  $t \in [s, s + S]$ , having easily interpretable parameters and covariances given by explicit analytical formulas. We assume that  $X(t)$  is build up from independent components  $X_i$  and that the covariances of the components are defined by the parameters  $\sigma_i(t)$  and

$$(14) \quad \tau_i(t) = \frac{1}{\pi} \frac{\sigma_i(t)}{\dot{\sigma}_i(t)},$$

where  $\dot{\sigma}_i^2(t)$  is the variance of  $\dot{X}_i(t)$ . The number of components is four and these represent variability of wind speed in different time scales; annual pattern; frequency of depressions and anti-cyclones which usually occur with periods of about 4 days; diurnal pattern due to different temperatures at day and night; and noise. For the month February the parameters  $\tau_i$  are shown in Figure 7 while  $\sigma_i$  are presented in Figure 8. Since  $\sigma_i$  are of comparable sizes non of the components  $X_i$  can be neglected.

In Figure 7 one can see that average excursion length above the mean of the transformed wind speed component  $X_1$  is about 1-1.5 month. The processes  $X_2, X_3, X_4$  have mean zero and the parameters  $\tau_i$  measure the average duration of periods when  $X_i(t) > 0$ . One can see that  $\tau_2$  is about 5 days while  $\tau_3$  about 1 day. Means to estimate the parameters  $\tau_i$  and  $\sigma_i$  are presented in Appendix B.

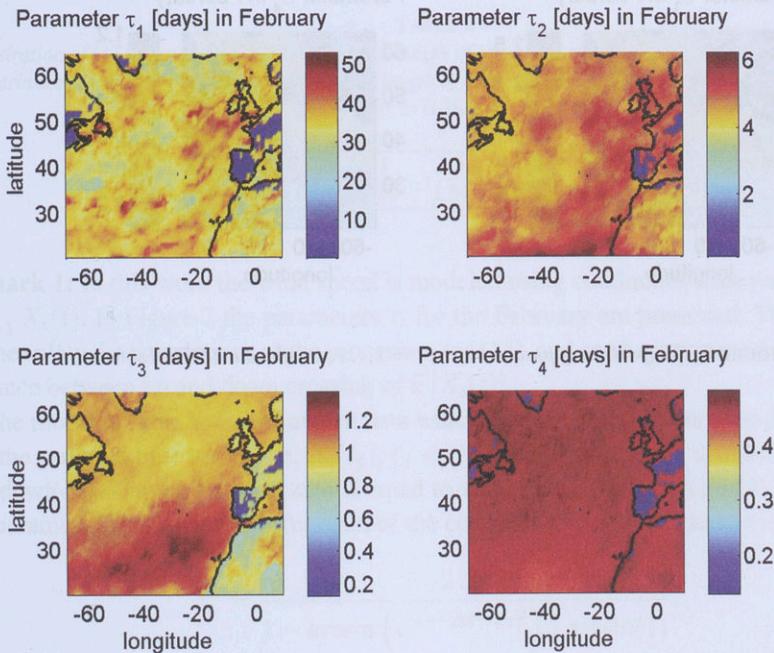


FIGURE 7. Comparison of parameters  $\tau_i$ ,  $i = 1, \dots, 4$  defined in (14), in February.

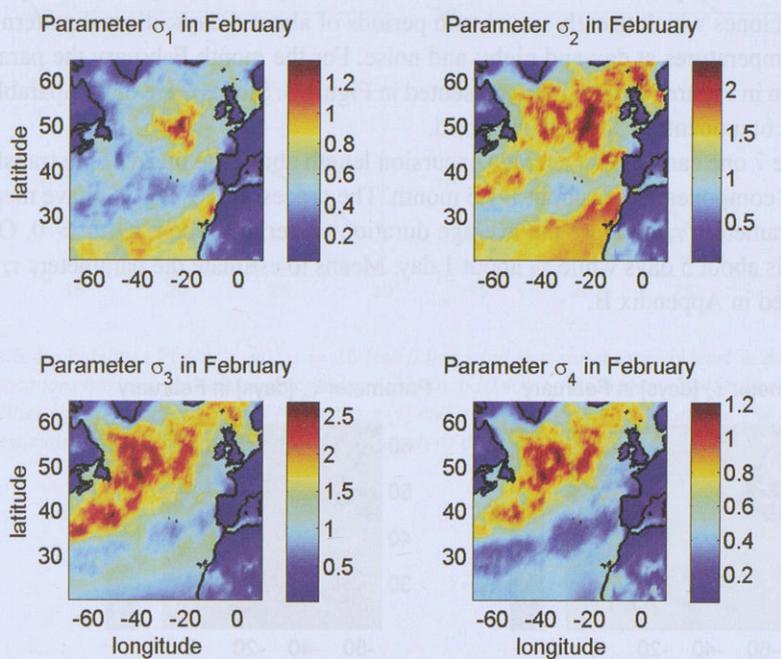


FIGURE 8. Comparison of parameters  $\sigma_i$   $i = 1, \dots, 4$  defined in (5), in February.

As mentioned in the introduction we are interested in a simple model of wind speed variability having physically interpretable parameters and as far as possible leading to explicit formulas for the non-stationary and spatio-temporal covariances, see (22) and (40), respectively. We thus propose to use Gaussian covariance functions for  $X_i$ , viz.

$$(15) \quad \text{Cov}(X_i(t), X_i(s)) = \sigma_i^2 e^{-\pi^2 (s-t)^2 / 2\tau_i^2}.$$

Consequently, by assumed independence of the processes  $X_i$  the covariance function of the process  $X$  is given by

$$(16) \quad \text{Cov}(X(t), X(s)) = \sum_{i=1}^4 \sigma_i^2 e^{-\pi^2 (s-t)^2 / 2\tau_i^2}.$$

A model having Gaussian covariance is sometimes criticized for its smoothness. Hence we introduced the component  $X_4$  which serves as an additive noise. The process has parameter  $\tau_4$  with values about 10 hours. (It should be also noted that ERA Interim surfaces are very smooth.)

In order to validate the model we have estimated the covariance of the signal  $W(\mathbf{p}, t)^{a(\mathbf{p})}$  for February month at positions  $\mathbf{p}$  marked by crosses in Figure 1. The empirical covariances are marked as dashed lines while the model based covariance as the solid lines in Figure 9. We can see that accuracy is fair.

TABLE 2

Illustration of (17). Here  $\tau$  [day] is the expected length of excursion above mean of Gaussian  $X_i(t)$  having covariance (15) while  $\tau^d$  [day] is the expected length of excursion above mean of the time series  $X_i(j\Delta t)$ ,  $\Delta t = 0.25$  day.

$\tau$	0.25	0.5	0.75	1	1.5	2	5	10	20
$\tau^d$	0.50	0.62	0.82	1.05	1.54	2.03	5.01	10.01	20.00

**Remark 1:** In this work the wind speed is modeled using continuous time process  $X(t) = \sum_{i=1}^4 X_i(t)$ . In Figure 7 the parameters  $\tau_i$  for the February are presented. The parameters define solely the correlations of the processes, see (15), and can be interpreted as the average distance between up and down crossing of  $E[X_i(t)]$ .

The model is fitted to ERA Interim data which is sampled with time step  $\Delta t$  of 6 hours. For the sampled time series, i.e.  $X_i(t_j)$ ,  $t_j = j/4$  days, the average distance between up- and down-crossing of the level zero is equal to  $\tau_i^d = \mathbb{P}(X_i(0) < 0, X_i(1/4) > 0)/2$ . Now the parameter  $\tau_i^d$  is an explicit function of the correlation which for model (15) is given by

$$(17) \quad \tau_i^d = \frac{2\Delta t}{1 - \arcsin\left(e^{-\pi^2 \Delta t^2 / 2\tau_i^2}\right) / \arcsin(1)}.$$

The relation (17) is illustrated in Table 2. By comparing the estimates of  $\tau_i$  presented in Figure 7 and values given in Table 2 one can conclude that process  $X_4$  is a model of a noise

which is hard to estimate if the sampling frequency is 6 hours. For the process  $X_3$  errors seems to be about 5% which is acceptable value and hence continuous time model seems to be an option. Finally processes  $X_3$  and  $X_4$  are very smooth and the sample step  $\Delta t$  is adequate.

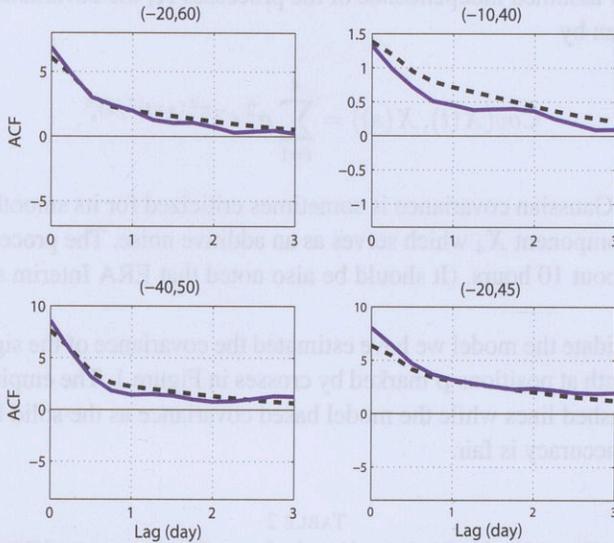


FIGURE 9. Comparisons between the covariances of  $X(t)$ ,  $t$  in February, given in (16) (dashed lines) and the covariances fitted to the data (solid lines) at four locations having longitudes and latitudes; (-20,60), (-10,40), (-40,50) and (-20,45).

The process  $X_i$  can then be defined as a moving average as follows

$$(18) \quad X_i(t) = \int_{-\infty}^{+\infty} \sigma_i f_{\tau_i}(t-r) dB_i(r).$$

Here  $B_i$  are independent Brownian motions while the kernels  $f_\tau$  are given by

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

#### 4.1. Non-stationary covariances

When the process  $X(t)$  is considered for a period  $S$  longer than one month the assumption of stationarity may not hold. This is manifested in the dependence of parameters  $\tau_i$  and  $\sigma_i$  on  $t$ . In the case when  $S$  is longer than one month we propose to model  $X_i$  as a moving average (18) with variable kernel, viz.

$$(19) \quad X_i(t) = \int_{-\infty}^{+\infty} \sigma_i(t) f_{\tau_i(t)}(t-r) dB_i(r).$$

Obviously the integrals in (19) has to be computed numerically. This is carried out using the following approximation

$$(20) \quad \int_{-\infty}^{+\infty} f_{\tau_i}(t)(t-s) dB_i(s) \approx \sum_j f_{\tau_i}(t)(t-s_j) \sqrt{ds} Z_{ij},$$

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 kernels  $f_{\tau_i}$ . (In the case when wind speed are simulated on very dense grid then it is recommended to slightly smooth the parameters  $a$ ,  $\sigma_i(t)$  and  $\tau_i(t)$ .)

The process  $X_i$  defined by (19) has an explicit analytical covariance function given by

$$(21) \quad \text{Cov}(X_i(t), X_i(s)) = \sigma_i(t)\sigma_i(s) \sqrt{\frac{2\tau_i(t)\tau_i(s)}{\tau_i^2(t) + \tau_i^2(s)}} e^{-\pi^2(s-t)^2/(\tau_i^2(s)+\tau_i^2(t))},$$

and hence

$$(22) \quad \text{Cov}(X(t), X(s)) = \sum_{i=1}^4 \sigma_i(t)\sigma_i(s) \sqrt{\frac{2\tau_i(t)\tau_i(s)}{\tau_i^2(t) + \tau_i^2(s)}} e^{-\pi^2(s-t)^2/(\tau_i^2(s)+\tau_i^2(t))}.$$

#### 4.2. Simulation of wind speed

The proposed model gives means for efficient simulation of wind speed over any time interval. Given parameters  $a(t)$ ,  $m(t)$ ,  $\sigma_i^2(t)$  and  $\tau_i(t)$  the kernels  $f_{\tau_i}$  can be evaluated and processes  $X_i$  simulated using (20). Alternatively one can simulate  $X(t)$  employing covariances defined in (22) and some of many methods to simulate Gaussian vectors. The algorithm based on (20) is preferable when densely sampled wind speed for long period of time are needed. In Figure 10 (top plots) simulated one year of wind speed are compared with the ERA interim data at two locations. The lower plots show the the same winds speed as the top ones limited to the February. The signals look similar.

### 5. Local spatio-temporal models for wind speed field

In the previous section we presented a model for wind speed variability at a fixed location  $W(t) = W(\mathbf{p}, t)$ . The parameters in the models depend on geographical location and season. Consequently having established locally temporal models over a globe one could evaluate long term distribution of wind speed at any fixed location. We have shown means to estimate some storm characteristics. However there are problems which solutions requires knowledge of simultaneous wind variability at several locations, e.g. the storm characteristics encountered by a vessel. Another example is the prediction of energy production from a system of wind mill farms. Such problems call for statistical spatio-temporal model for wind speed  $W(\mathbf{p}, t)$ . By assumed Gaussianity of the transformed wind speed it means that one need to find the covariances between  $X(\mathbf{p}_1, t_1)$  and  $X(\mathbf{p}_2, t_2)$  which will be denoted by  $\mathbb{C}(\mathbf{p}_1, t_1, \mathbf{p}_2, t_2)$ .

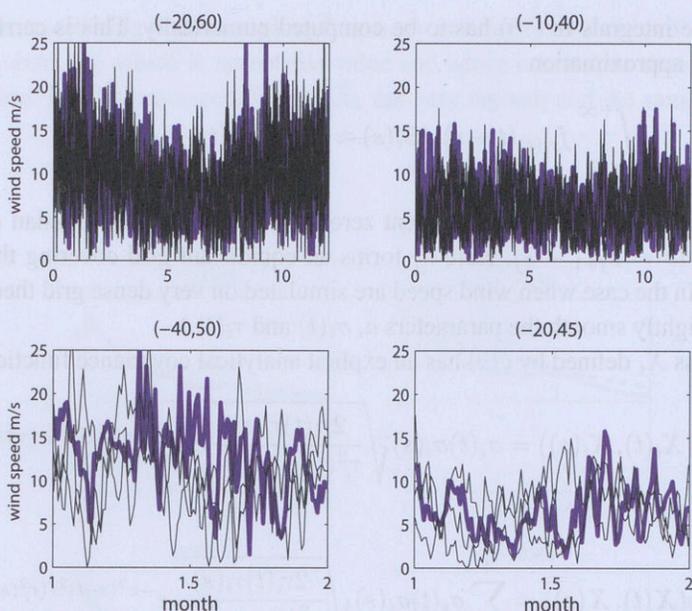


FIGURE 10. Top - Comparison of one year of ERA Interim data at two locations (solid thick line) with two simulation of the speed. Bottom - Zoomed top plots when wind speed are limited to February.

Similarly as for  $X(t)$  we will consider different time scales of wind speed variability and define

$$(23) \quad X(\mathbf{p}, t) = \sum_{i=1}^4 X_i(\mathbf{p}, t),$$

where where  $X_1(\mathbf{p}, t)$  describes the seasonal variability,  $X_4(\mathbf{p}, t)$  is modeling the observed high frequency colored noise, respectively. The field  $X_2(\mathbf{p}, t)$  will model wind speed variability related to frequency of depressions and anti-cyclones occurrences while  $X_3(\mathbf{p}, t)$  will model diurnal pattern. (The fields  $X_2, X_3, X_4$  have mean zero.) Since  $X_i$  are independent

$$\mathbb{C}(\mathbf{p}_1, t_1, \mathbf{p}_2, t_2) = \sum_{i=1}^4 \mathbb{C}_i(\mathbf{p}_1, t_1, \mathbf{p}_2, t_2).$$

The proposed model will be locally homogeneous and stationary in the sense described in following sections. Presentation starts with a definition of local covariance functions, which are homogeneous and stationary. The covariance

$$\mathbb{C}_i(\mathbf{p}_1, t_1, \mathbf{p}_2, t_2) = \mathbb{C}_i(\mathbf{p}_2 - \mathbf{p}_1, t_2 - t_1)$$

will be Gaussian and will reduce to (15) if  $\mathbf{p}_2 = \mathbf{p}_1$ .

Since components  $X_i$  are independent hence, for simplicity of presentation only, we drop the subscript  $i$  and write  $X$  for  $X_i$ . Further we shall denote by  $X_x, X_y, X_t$  the partial

derivatives of  $X(x, y, t)$  on  $x, y$  and  $t$ , respectively. In addition we will denote  $(\mathbf{p}, t)$  by  $\mathbf{x}$  when writing formulas for the covariances. We turn now to define the homogeneous covariance describing the local correlation structure.

Let  $\Lambda$ ,

$$(24) \quad \Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix}$$

be a symmetrical matrix having positive eigenvalues then

$$(25) \quad C(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \exp\left(-\frac{\pi^2}{2}(\mathbf{x}_2 - \mathbf{x}_1)\Lambda(\mathbf{x}_2 - \mathbf{x}_1)^T\right) = C(\mathbf{x}_2 - \mathbf{x}_1)$$

is a homogeneous covariance function. Obviously  $\text{Var}(X) = \sigma^2$  while some simple algebra show that

$$(26) \quad \lambda_{11} = \frac{1}{\pi^2} \frac{\text{Var}(X_x)}{\text{Var}(X)}, \quad \lambda_{22} = \frac{1}{\pi^2} \frac{\text{Var}(X_y)}{\text{Var}(X)}, \quad \lambda_{33} = \frac{1}{\pi^2} \frac{\text{Var}(X_t)}{\text{Var}(X)}.$$

Similarly one can show that

$$(27) \quad \lambda_{12} = \frac{1}{\pi^2} \frac{\text{Cov}(X_x, X_y)}{\text{Var}(X)}, \quad \lambda_{13} = \frac{1}{\pi^2} \frac{\text{Cov}(X_x, X_t)}{\text{Var}(X)}, \quad \lambda_{23} = \frac{1}{\pi^2} \frac{\text{Cov}(X_y, X_t)}{\text{Var}(X)}.$$

Now it is easy to see that  $\pi^2 \sigma^2 \cdot \Lambda$  is the covariance matrix of the gradient vector  $(X_x, X_y, X_t)$ .

Note that parameter  $\tau$  defined in (10) is equal to  $\lambda_{33}^{-1/2}$ . Similarly

$$(28) \quad L_x = \lambda_{11}^{-1/2} \quad L_y = \lambda_{22}^{-1/2},$$

are the average length of the windy region in  $x, y$ , direction, respectively. As before we say that the windy conditions are present at position  $\mathbf{p}$  at time  $t$  if wind speed exceeds the median speed, i.e.  $W(\mathbf{p}, t) \geq \mu(\mathbf{p}, t)$ .

Summarizing the diagonal elements of  $\Lambda$  are equal to reciprocals of  $L_x^2, L_y^2$  and  $\tau^2$ . In the following section we shall demonstrate that the off-diagonal elements  $\lambda_{13}, \lambda_{23}$  define velocity of storm (windy weather region) movements. Obviously if  $\lambda_{13} = \lambda_{23} = 0$  then the covariance is separable and there is no organized movement of the wind field. Finally the element  $\lambda_{12}$  will define the main direction of the storm movement.

## 6. 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 [11] Longuet-Higgins has introduced velocities to study movements of random surfaces. There are several definitions of velocities proposed in the literature, see [4]. Here we will use velocity in a fixed direction  $\theta$ , say, the direction  $\theta$ ,

called the main azimuth of a storm, will be defined in Remark 2, see also Example 1. As customary we use the convention that the direction south to north has azimuth  $\theta = 0^\circ$  while azimuth  $\theta = 90^\circ$  for the direction west to east.

Following [4] the velocities in the direction  $\theta$  and  $\theta - 90^\circ$  are given by

$$(29) \quad \vec{V}_\theta = -\frac{W_t}{W_\theta}, \quad \vec{V}_{\theta-90^\circ} = -\frac{W_t}{W_{\theta-90^\circ}},$$

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 and fixed time  $t$ .

The general assumption of this paper is that parameter  $a$ , see Figure 1, does not depend on time and changes slowly in space. Hence the gradient  $\nabla W = (W_x, W_y, W_t)$  can be approximated by

$$(30) \quad \nabla W \approx (1/a) X^{1/a-1} \nabla X,$$

where  $\nabla X$  is the gradient of  $X$ -field. Hence velocities defined in (29) can be approximated by

$$(31) \quad \vec{v}_\theta = -\frac{X_t}{X_\theta}, \quad \vec{v}_{\theta-90^\circ} = -\frac{X_t}{X_{\theta-90^\circ}},$$

where  $X_t$  is the time derivative of the transformed wind speed,  $X_\theta$  and  $X_{\theta-90^\circ}$  are the directional derivatives having azimuths  $\theta$ ,  $\theta - 90^\circ$ , respectively. For a homogeneous Gaussian field median velocities are given by

$$(32) \quad \vec{v}_\theta = -\frac{\text{Cov}(X_\theta(\mathbf{p}, t), X_t(\mathbf{p}, t))}{\text{Var}(X_\theta)}, \quad \vec{v}_{\theta-90^\circ} = -\frac{\text{Cov}(X_{\theta-90^\circ}(\mathbf{p}, t), X_t(\mathbf{p}, t))}{\text{Var}(X_{\theta-90^\circ})},$$

see [4] for a proof. The speed in directions  $\theta$  and  $\theta - 90^\circ$  will be denoted by  $v_\theta$ ,  $v_{\theta-90^\circ}$ , respectively.

In the special case when  $\theta = 90^\circ$  or  $\theta = 0^\circ$ , i.e. one is interested in velocities of movements along the  $x$ -axis and  $y$ -axis, we have that

$$(33) \quad \vec{v}_x = -\frac{\lambda_{13}}{\lambda_{11}}, \quad \vec{v}_y = -\frac{\lambda_{23}}{\lambda_{22}}.$$

In general 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:** For several reasons, see [5] 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  $\Lambda(\theta)$  the matrix  $\Lambda$  in the rotated coordinates viz.

$$(34) \quad \Lambda(\theta) = A_\theta^T \Lambda A_\theta,$$

where  $A_\theta^T$  is the transpose of  $A_\theta$ . Now Equations (26-27) can be written as follows

$$(35) \quad \lambda_{11}(\theta) = \frac{1}{\pi^2} \frac{\text{Var}(X_\theta)}{\text{Var}(X)}, \quad \lambda_{22}(\theta) = \frac{1}{\pi^2} \frac{\text{Var}(X_{\theta-90^\circ})}{\text{Var}(X)}, \quad \lambda_{33}(\theta) = \frac{1}{\pi^2} \frac{\text{Var}(X_t)}{\text{Var}(X)} = \lambda_{33},$$

and

$$(36) \quad \lambda_{12}(\theta) = 0, \quad \lambda_{13}(\theta) = \frac{1}{\pi^2} \frac{\text{Cov}(X_\theta, X_t)}{\text{Var}(X)}, \quad \lambda_{23}(\theta) = \frac{1}{\pi^2} \frac{\text{Cov}(X_{\theta-90^\circ}, X_t)}{\text{Var}(X)}.$$

Further note that parameter  $\tau$  defined in (10) is equal to  $\lambda_{33}^{-1/2}$ . Similarly

$$(37) \quad L_\theta = \lambda_{11}(\theta)^{-1/2} \quad L_{\theta-90^\circ} = \lambda_{22}(\theta)^{-1/2},$$

are the average length of the windy region in  $\theta$ ,  $\theta - 90^\circ$ , direction, respectively. Obviously the median speed  $\vec{v}_\theta$  and  $\vec{v}_{\theta-90^\circ}$  can be evaluated using (33) once the matrix  $\Lambda(\theta)$  has been evaluated.

**Example 1:** Let consider the following field

$$(38) \quad X(x, y, t) = \sigma_1 R_1 \cos\left(2\pi \frac{t}{T} - 2\pi \frac{x}{L} + \phi_1\right) + \sigma_2 R_2 \cos\left(2\pi \frac{t}{T} + \phi_2\right)$$

where  $R_1, R_2$  and  $\phi_1, \phi_2$  are independent variables having Rayleigh, uniform CDF, respectively, i.e.  $X$  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  $\text{Var}(X_\theta) = (2\pi)^2 \sigma_1^2 / L^2$ . Hence the median velocities (32) are given by

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

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

In Figure 11 variability of the median velocities  $\vec{v}_\theta$  and  $\vec{v}_{\theta-90^\circ}$  (32) are compared. In the top plots seasonal variability of  $\vec{v}_\theta$  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  $\vec{v}_{\theta-90^\circ}$  is given in the bottom plots, where the maximal speed is about 19 km/h. Generally one can say that 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.

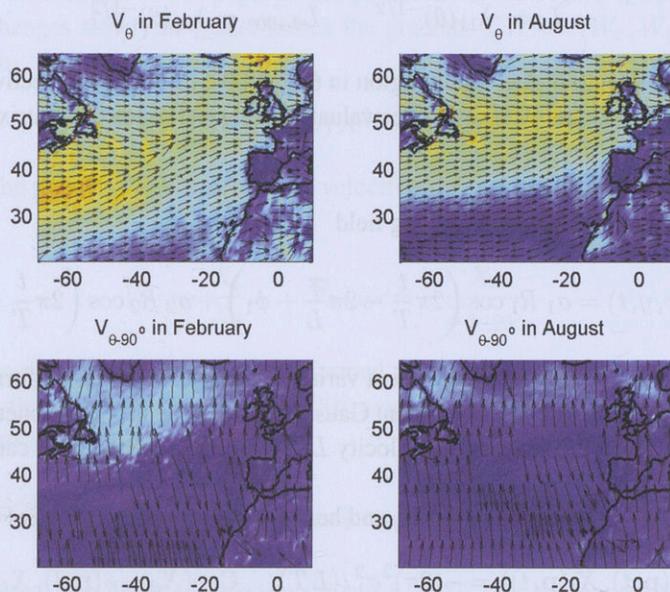


FIGURE 11. Top - Estimates of the median velocity, km/h, of a storm (windy region) movement in direction  $\theta$  in February and August. The color corresponds to speed. The highest speed (orange) is 45.1 km/h while the lowest (blue) is 0 km/h. Bottom - Comparisons of the median velocities  $\vec{v}_{\theta-90}$  in February and August. The highest speed is 18.6 km/h.

We turn now to illustration of the dynamics of the fields  $X_i$ . Only the winter month February and velocities  $\vec{v}_\theta$  will be presented here, see Figure 12. In the figure, left upper plot, one can see that for the component  $X_1$ , which model the seasonal variability, the velocity  $\vec{v}_\theta \approx 0$ , i.e. the component has locally separable covariance ( $\lambda_{13}$  and  $\lambda_{23}$  are practically zero). Similarly the component  $X_3$ , modeling diurnal pattern, moves slowly with speed about 10 km/h. The component  $X_2(\mathbf{p}, t)$ , modeling wind speed variability related to frequencies of depressions and anti-cyclones occurrences, moves relatively fast with speeds about 50 km/h. Further the "noise" field  $X_4$  moves with speed which are not negligible however the estimation error of the speed can be large because of relatively long sampling step  $\Delta t = 6$  hours.

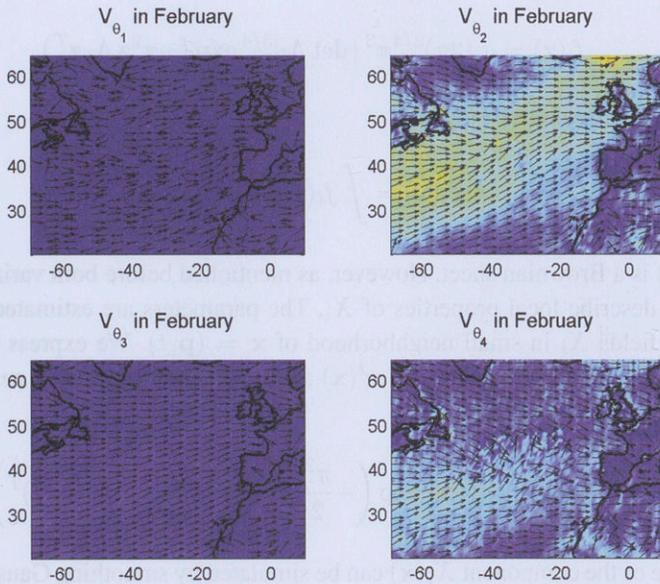


FIGURE 12. Estimates of the median velocities, km/h, the fields  $X_i$  moves in direction  $\theta$  in February. The highest speed  $X_1$  moves is 3 km/h,  $X_2$  maximal speed is 61 km/h,  $X_3$  maximal speed is 12 km/h while  $X_4$  maximal speed is 37 km/h.

**Remark 3:** Following irregularity factors can be useful

$$(39) \quad \alpha_\theta^2 = \frac{\lambda_{13}(\theta)}{\sqrt{\lambda_{11}(\theta)\lambda_{33}(\theta)}}, \quad \alpha_{\theta-90^\circ}^2 = \frac{\lambda_{23}(\theta)}{\sqrt{\lambda_{22}(\theta)\lambda_{33}(\theta)}}.$$

Roughly, smaller values of the factors higher risks of extreme storms, see [5] for more details. Further, if  $\alpha_\theta^2 + \alpha_{\theta-90^\circ}^2 = 1$  then the surface  $X$  drifts, viz.

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

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

## 7. Non-homogeneous spatio-temporal model for the field $X$

Derivation of the inhomogeneous covariance function is a simple extension of the one dimensional construction, used in Section 4.1, of the non-stationary covariances (22). More precisely a zero mean Gaussian field  $X_i$ , having covariance function

$$\mathbb{C}_i(\mathbf{x}_1, \mathbf{x}_2) = \sigma_i^2 \exp\left(-\frac{\pi^2}{2}(\mathbf{x}_2 - \mathbf{x}_1)\Lambda_i(\mathbf{x}_2 - \mathbf{x}_1)^T\right),$$

will be defined by means of a moving average process, i.e. a convolution of infinitesimal "white noise" with the following kernel

$$f_i(\mathbf{z}) = \sigma_i(2\pi)^{3/4}\pi^3 |\det \Lambda_i|^{3/4} \exp(-\pi^2 \mathbf{z} \Lambda_i \mathbf{z}^T),$$

which one can write more formally as

$$X_i(\mathbf{x}) = \int f_i(\mathbf{x} - \mathbf{z}) dB_i(\mathbf{z}),$$

where  $B_i(\mathbf{z})$  is a Brownian sheet. However, as mentioned before both variances  $\sigma_i^2$  and the matrices  $\Lambda_i$  describe local properties of  $X_i$ . The parameters are estimated using observations of the fields  $X_i$  in small neighborhood of  $\mathbf{x} = (\mathbf{p}, t)$ . We express this dependence explicitly in the notation by means of  $\sigma_i^2(\mathbf{x})$  and  $\Lambda_i(\mathbf{x})$ . For  $\mathbf{x}_1, \mathbf{x}_2$  in a neighborhood of  $\mathbf{x}$  a covariance (25) can be defined, viz.

$$\mathbb{C}_i(\mathbf{x}_1, \mathbf{x}_2) = \sigma_i^2(\mathbf{x}) \exp\left(-\frac{\pi^2}{2}(\mathbf{x}_2 - \mathbf{x}_1)\Lambda_i(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1)^T\right),$$

and the value of the component  $X_i(\mathbf{x})$  can be simulated by smoothing Gaussian white noise by means of a kernel

$$f_i(\mathbf{z}; \mathbf{x}) = \sigma_i(\mathbf{x})(2\pi)^{3/4}\pi^3 |\det \Lambda_i(\mathbf{x})|^{3/4} \exp(-\pi^2 \mathbf{z} \Lambda_i(\mathbf{x}) \mathbf{z}^T),$$

giving

$$X_i(\mathbf{x}) = \int f_i(\mathbf{x} - \mathbf{z}; \mathbf{x}) dB_i(\mathbf{z}).$$

The so defined non-homogeneous field is zero mean Gaussian with covariance function given by

$$\mathbb{C}_i(\mathbf{x}, \mathbf{x}') = \int f_i(\mathbf{x} - \mathbf{z}; \mathbf{x}) f_i(\mathbf{x}' - \mathbf{z}; \mathbf{x}') dz.$$

The integral can be evaluated leading the following explicit formula for the covariance

$$(40) \quad \mathbb{C}_i(\mathbf{x}, \mathbf{x}') = \sigma_i(\mathbf{x})\sigma_i(\mathbf{x}') \sqrt{\frac{2^3 |\Lambda_i(\mathbf{x})|^{-1/2} |\Lambda_i(\mathbf{x}')|^{-1/2}}{|\Lambda_i(\mathbf{x})^{-1} + \Lambda_i(\mathbf{x}')^{-1}|}} e^{-\pi^2 \mathbf{z}(\Lambda_i(\mathbf{x})^{-1} + \Lambda_i(\mathbf{x}')^{-1})^{-1} \mathbf{z}^T},$$

where  $\mathbf{z} = \mathbf{x}' - \mathbf{x} = (x' - x, y' - y, t' - t)$ . Finally, by assumed independence of the components  $X_i$  the covariance structure of the  $X$  field is given by

$$(41) \quad \mathbb{C}(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^4 \mathbb{C}_i(\mathbf{x}, \mathbf{x}').$$

**Example 2: Simulations of wind fields.** For any finite collection of points  $\mathbf{x}_i = (x_i, y_i, t_i)$  the random variables  $X(\mathbf{x}_i)$ ,  $i = 1, \dots, n$  can be simulated using covariances  $\mathbb{C}(\mathbf{x}_i, \mathbf{x}_j)$  evaluated using (40-41) and the average values of  $X_1(\mathbf{x}_i)$ ,  $m(\mathbf{x}_i)$  defined in (3). In Figure 13 two simulations of the wind speed field for  $t$  around 15th February and six hours later are presented. The simulated fields are compared with the two fields extracted from ERA Interim data base. The fields look similarly. The comparison is by no means validation of the model. It just shows that the model is not obviously wrong. The model will be validated in the following section where statistics of the encountered wind speed measured on-board of vessels will be compared with the theoretically, i.e. from the model, derived distributions. The material is based on [18].

## 8. Statistics of encountered wind speed

In this section we shall apply the proposed spatio-temporal covariances (40-41) to describe variability of wind speed time series encountered by a vessel. We will give means to derive the long-term distribution of encountered wind speed and the expected number of crossings of a level  $u$  that are needed to estimate expected duration and strength of encountered storms. More detailed study of the data can be found in [18].

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 speed  $v^{sh}(t)$ ,  $t \in [s, s + S]$ , define its position  $\mathbf{p}(t)$  at any time  $t$  during a voyage. Then the encountered wind speed are given by

$$(42) \quad W(t) = W(\mathbf{p}(t), t), \quad s \leq t \leq s + S.$$

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

$$(43) \quad \vec{v}^{sh}(t) = (\dot{x}(t), \dot{y}(t)) = v^{sh}(t) (\sin \alpha(t), \cos \alpha(t)),$$

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  $\theta = 0^\circ$ .) In the following we will use the transformed Gaussian field (2) to model the encountered wind speed  $W(t)$ , viz.

$$(44) \quad W(t) = X(\mathbf{p}(t), t)^{1/a(\mathbf{p}(t))} = X(t)^{1/a(t)}.$$

The process  $X(t)$  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 distribution of  $W(t)$  can be derived by means of the model viz.

$$(45) \quad \mathbb{P}(W \leq w) = \frac{1}{S} \int_s^{s+S} \Phi \left( \frac{w^{a(t)} - m(t)}{\sigma(t)} \right) dt.$$

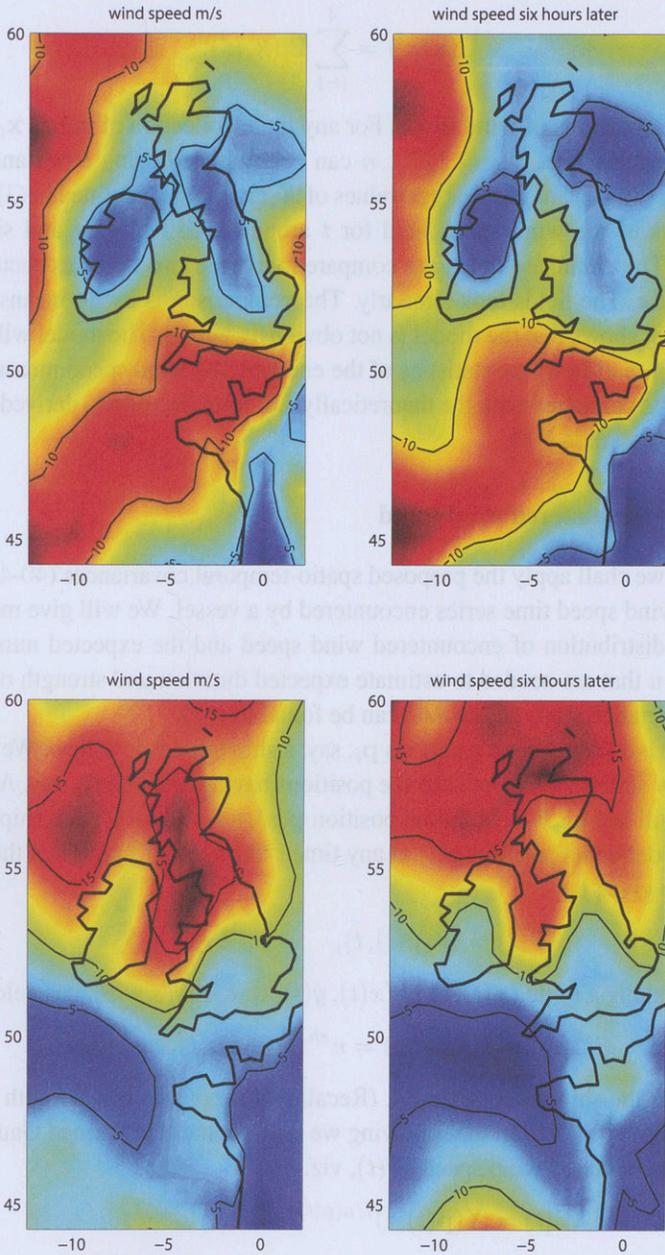


FIGURE 13. Two simulations of wind fields around 15th of February

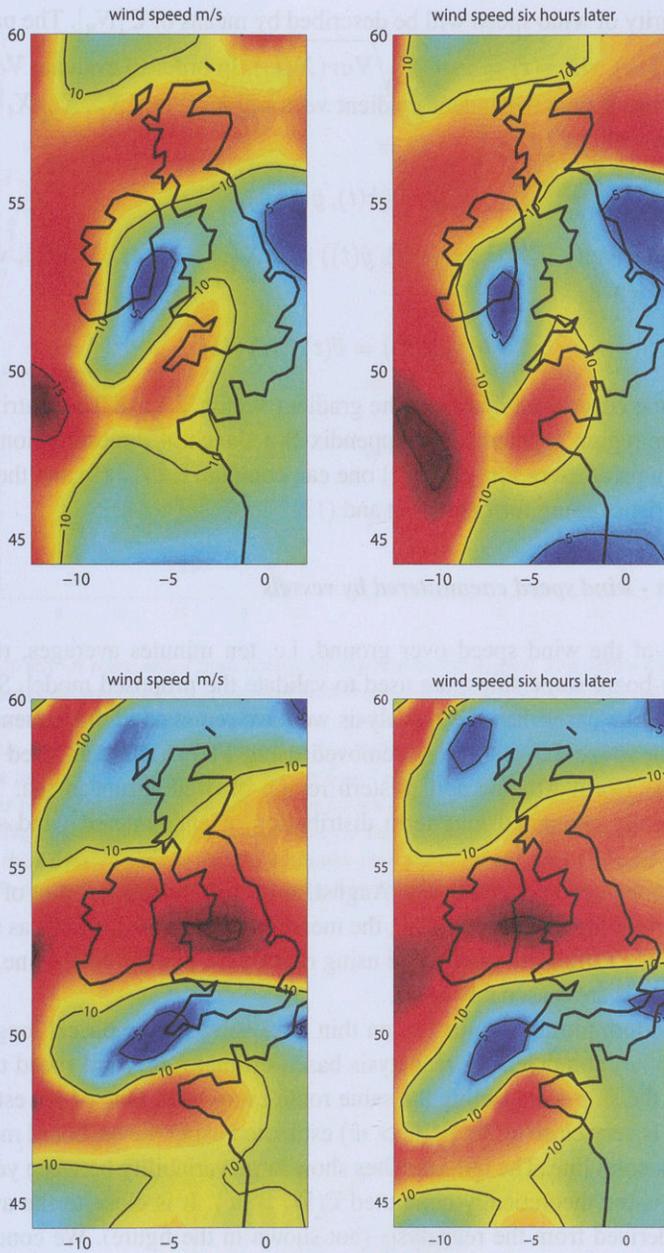


FIGURE 14. ERA Interim wind fields around 15th of February

The irregularity of wind speed will be described by means of  $E[N_u]$ . The parameter  $\tau(t)$  was defined in (10), viz.  $\tau(t) = \pi\sigma(t)/\sqrt{\text{Var}(\dot{X}(t))}$ . In order to evaluate  $\text{Var}(\dot{X}(t))$  one needs to introduce a time dependent gradient vector  $\nabla X(t) = (X_x, X_y, X_t)(\mathbf{p}(t), t)$  and the vector of derivatives

$$(46) \quad \vec{v} = (\dot{x}(t), \dot{y}(t), 1).$$

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

$$(47) \quad \text{Var}(\dot{X}(t)) = \vec{v}(t) \Lambda(t) \vec{v}(t)^T,$$

where  $\Lambda(t)$  is the covariance matrix of the gradient vector  $\nabla X(t)$ . The matrix  $\Lambda$  has to be estimated in the region of interest. In Appendix B a sketch of the estimation procedure is given. Knowing parameters  $\tau(t)$  and  $\sigma(t)$  one can compute  $E[N_w]$  (9) and the encountered storm characteristics using formulas (12) and (13)

### 8.1. Validation - wind speed 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 reanalysis 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 15, is considered. In right top plot of the figure, the measured wind speed, shown as solid line, are compared with the estimated wind speed using reanalysis, dashed dotted line. One can see that the two signals are reasonably close.

In the left bottom plot of Figure 15, ten thin lines show the empirical long term probabilities  $\mathbb{P}(W > w)$  computed for reanalysis based estimates of wind speed 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  $\mathbb{P}(W > w)$  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  $\mathbb{P}(W > w)$ . It is close to the average of the ten estimates derived from the reanalysis (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 rout. Similar conclusions can be drawn from Figure 17 left plot where the combined long-term distributions for all 40 voyages are shown. Based on the results presented in Figures 15 - 17, we conclude that the theoretical long term distribution of wind speed encountered by a sailing vessels agrees well with the distribution derived using reanalysis; and secondly that the routing systems used in planning a route is successful in selecting routs with calmer wind conditions than average one, see also [12].

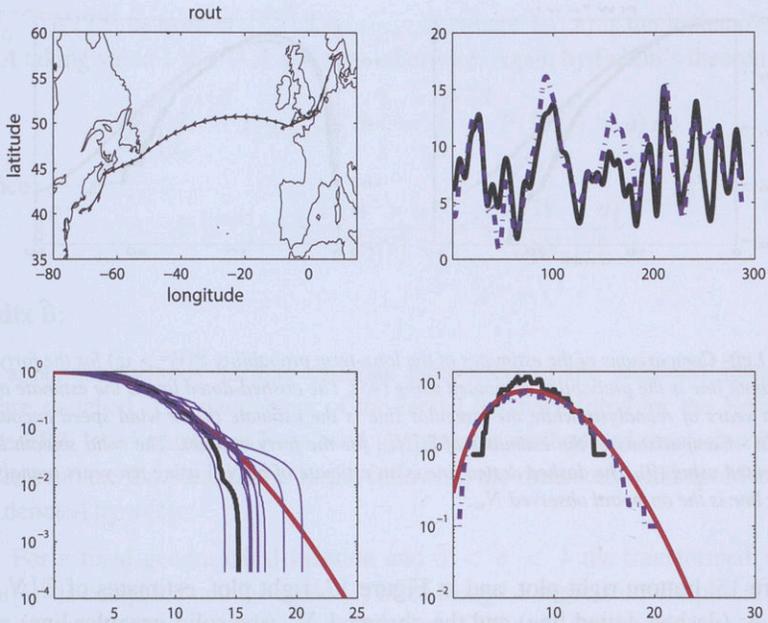


FIGURE 15. Top left - A route sailed from Europa to America in late August. Top right - Wind speed measured on-board a vessel (solid irregular line) compared with their estimates derived from the reanalysis data (dashed dotted line). Bottom left - Comparisons of estimates of the long-term probability  $\mathbb{P}(W > w)$ , plotted on the logarithmic scale, for the voyage. The thick smooth line is the probability computed using (45). The less smooth thick line is the probability estimated using the on-board measured wind speed. The thin irregular lines are the probabilities estimated from the reanalysis 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 using (9). The thick irregular line is the  $N_w$  evaluated from the on-board measured wind speed. The dashed dotted line is the estimate of  $E[N_w]$  using reanalysis derived wind speed for the route sailed in ten years.

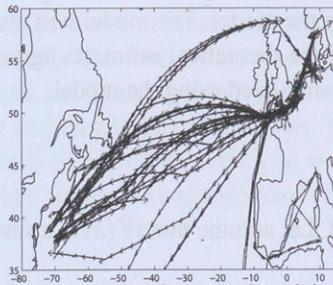


FIGURE 16. The considered routes in the validation process.

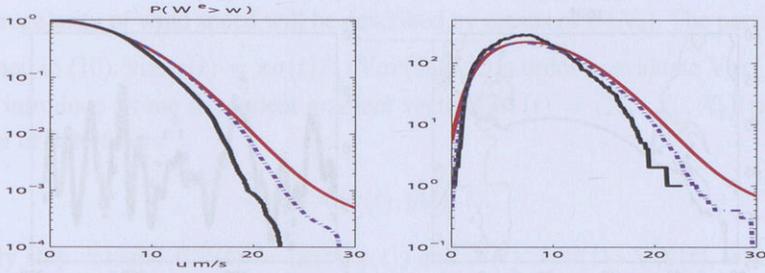


FIGURE 17. Left- Comparisons of the estimates of the long-term probability  $\mathbb{P}(W > w)$  for the forty voyages. The solid smooth line is the probability computed using (45). The dashed-dotted line is the estimate of  $\mathbb{P}(W > w)$  using ten years of reanalysis while the irregular line is the estimate of the wind speed 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 (9). The dashed dotted line is an estimate of  $E[N_w]$  using ten years reanalysis while the irregular line is the on-board observed  $N_w$ .

In Figure 15, bottom right plot, and in Figure 17, right plot, estimates of  $E[N_w]$  based on reanalysis (dashed dotted line) and the observed  $N_w$  (the solid irregular line) are compared with the theoretical  $E[N_w]$  computed using (9) for routes shown in Figure 16 and Figure 15 top left plot. One can see that the lines are close except for the high wind speed. The observed crossings of high wind speed (solid irregular line) are fewer than theoretically predicted. This we attribute to use of routing programs that successfully choose calmer route than the average one. This claim is also supported by studies of the estimate of  $E[N_w]$  derived from 10 years of reanalysis, shown as the dashed line. One can see that these estimates are higher than on-board observed  $N_w$  for wind speed above 12 m/s. A more extensive statistical description of the encountered speed is presented in [18].

## 9. Conclusions

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.

## Appendix A:

In this appendix (13) is proved. Let assume that  $W(s)$  is a smooth process. Using Fubinni's theorem

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

Since  $\int_0^{+\infty} N(T^{st} > t) dt = \int_s^{s+S} \mathbf{1}_{\{W(t) \geq u\}} dt$ , 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

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

and hence

$$E [T^{st}] = S \frac{P(W > u)}{E [N]} = S \frac{P(W > u)}{E [N_u] / 2}.$$

## Appendix B:

In this appendix methods used to estimate parameters of the proposed spatio-temporal model are sketched. The parameters of the model has been fitted for the North Atlantic. Here the ERA Interim data has been used. A moments method and regression fit were employed to estimate the parameters. In the following the measured wind speed 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 (3) fitted using LS regression. Empirical cumulative distribution function (CDF) and Gaussian (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 Step1) 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  $\sigma_i^2(t_j)$  are found. Then  $\sigma_i^2(t)$  are estimated by fitting seasonal components, similar to (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. 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.

## 10. Acknowledgments

Support of Chalmers Energy Area of Advance is acknowledge. Research was also supported by Swedish Research Council Grant 340-2012-6004 and by Knut and Alice Wallenberg

stiftelse. The author also would like to thank Wallenius Lines AB for providing on-board wind measurement data and two anonymous referees for valuable comments.

## References

- [1] ADLER R.J. (1981) *The geometry of random fields*. Wiley
- [2] ADLER R.J. AND TAYLOR J.E. (2007) *Random fields and Geometry*. Springer Monographs in Mathematics.
- [3] BANACH, S. (1925) Sur les lignes rectifiables et les surfaces dont l'aire est finie *Fund. Math.*, **7** 225-237.
- [4] BAXEVANI, A., PODGÓRSKI, K. and RYCHLIK, I. (2003) Velocities for moving random surfaces. *Prob. Eng. Mechanics*, **18** 251-271.
- [5] BAXEVANI, A. and RYCHLIK, I. (2006) Maxima for Gaussian seas. *Ocean Engineering*, **33** 895-911.
- [6] BAXEVANI, A., CAIRES, S. and RYCHLIK, I. (2008) Spatio-temporal statistical modelling of significant wave height. *Environmetrics*, **20** 14-31.
- [7] 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.
- [8] 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 Energ.* **13** 339-355.
- [9] 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, L., KÅLLBERG, P., KÖHLER, M., MATRICARDI, M., MCNALLY, A.P., MONGE-SANZ, B.M., MORCRETTE, B.M., PARK, B.M., 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.
- [10] KISS, P. and JÁNOSI, I.M. (2008) Limitations of wind power availability over Europe: a conceptual study. *Nonlin. Processes Geophys.* **15** 803-813.
- [11] LONGUET-HIGGINS, M. S.. (1957) The statistical analysis of a random, moving surface. *Phil. Trans. Roy. Soc. A*, **249** 321-387.
- [12] 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.
- [13] MONBET, V., AILLIOT, P., and PREVOSTO, M. (2007) Survey of stochastic models for wind and sea state time series. *Prob. Eng. Mechanics*, **22** 113-126.
- [14] PODGÓRSKI, K., RYCHLIK, I. and MACHADO, U. E. B. (2000) Exact Distributions for Apparent Waves in Irregular Seas, *Ocean. Engng.*, **27** 979-1016.

- [15] RICE, S. O. (1944, 1945) The mathematical analysis of random noise part I and II. *Bell Syst Tech J.* **23** 282–332, **24** 46–156.
- [16] 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.
- [17] RYCHLIK, I. (2000) On some reliability applications of Rice formula for intensity of level crossings. *Extremes*, **3** 331-348.
- [18] RYCHLIK, I. and MAO, W. (2014) Probabilistic Model for Wind Speed Variability Encountered by a Vessel. *Natural Resources*, **5** 837-855 . DOI: 10.4236/nr.2014.513072

Igor Rychlik  
 Mathematical Sciences,  
 Chalmers University of Technology,  
 SE-412 96 Göteborg, Sweden. e-mail: rychlik@chalmers.se