Non-Zero Mean Gaussian Process Prior Wind Field Models

Dan Cornford

Abstract

This report outlines the derivation and application of a non-zero mean, polynomial-exponential covariance function based Gaussian process which forms the prior wind field model used in 'autonomous' disambiguation. It is principally used since the non-zero mean permits the computation of realistic local wind vector prior probabilities, $P(u_i, v_i)$, which are required when applying the *scaled-likelihood trick*, as the marginals of the full wind field prior $P(U, V)$. As $P(U, V)$ is multivariate normal, these marginals are very simple to compute.

1   Introduction

Prior models for wind fields, $P(U, V)$, are required for the disambiguation methodology proposed at the NCRG at Aston (Cornford and Nabney, 1998); either 'forward' disambiguation:

$$P(U, V \mid \mathbf{\Sigma}^o) = \left(\prod_i P(\boldsymbol{\sigma}_i^o \mid u_i, v_i) \right) P(U, V) \tag{1}$$

or 'inverse' disambiguation:

$$P(U, V \mid \mathbf{\Sigma}^o) = \left(\prod_i \frac{P(u_i, v_i \mid \boldsymbol{\sigma}_i^o)}{P(u_i, v_i)} \right) P(U, V). \tag{2}$$

A non-zero mean Gaussian process is chosen to capture useful information such as, typical length scales, magnitudes, degrees of smoothness and noise variance[1]. A non-zero mean is particularly important when implementing the 'inverse' disambiguation where the $P(u_i, v_i)$ term, which is an unconditional prior probability of the local wind vector, will in general be far from zero mean. The reason for this is that almost all regions of the Earth have dominant wind directions, related to the global climate system (Cornford, 1997b; Wallace and Hobbs, 1977).
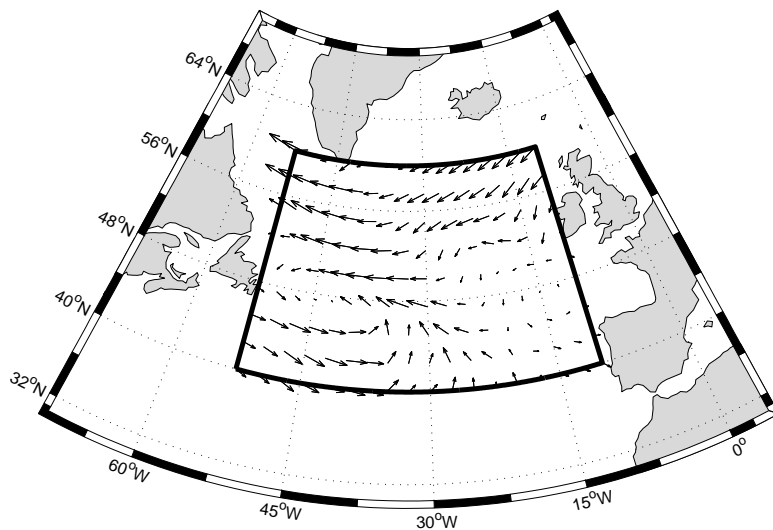
2   Data

Figure 1: The area of the North Atlantic over which the prior model is defined together with an example of the ECMWF model winds for midnight 25/04/95

This work considers a sector of the North Atlantic from $52.5°W, 40°N$ to $10°W, 60°N$ (Figure 1) and fits a non-zero mean Gaussian process over this region. This prior model can only be applied (for disambiguation) in the North Atlantic. In principle there is no reason why it could not be applied elsewhere, the only problem is one of data collection and determination of a suitable form for the mean function. Additionally it would become necessary to allow the parameters in the covariance function to adapt, since these are unlikely to remain stationary over the whole globe (as they are assumed to be over the region considered).

In order to set the parameters in the wind field prior a large amount of European Centre for Medium range Weather Forecasting (ECMWF) data is used, which consists of gridded analysis

---
[1] For a spatial Gaussian process this is often referred to as the nugget variance and also encapsulates sub-sampling scale fluctuations.

(`AS`) data on a regular 2.5 degree latitude-longitude grid. The surface wind field[2] for a small region of the North Atlantic is extracted from the global, gridded data set. The British Atmospheric Data Centre (BADC) performed the interpolation from the model grid to the regular grid using software provided by ECMWF. An example wind field is shown in Figure 1. In order to obtain reliable climatological estimates of the parameters in the wind field model (in particular the mean parameter) three complete years of data from 1995, 1996 and 1997 were used.

A smaller data set was chosen to take advantage of the revisions made to the ECMWF model surface wind parameterisations in 1994, which increased model 10 m wind speeds. Thus the more reliable, recent data (which also includes scatterometer derived winds which were assimilated during the period for which the data is extracted) are considered.

yieldThe wind patterns in the North Atlantic show strong seasonal periodicity. Generally, winds are stronger during the winter season than during the summer season. It is also possible that the characteristic length scales of features change through the seasons. In order to account for this a separate prior wind field model is developed for each month, using the three years data. Since ECMWF analyse the incoming synoptic data (including wind observations) every six hours, there are roughly 360 sample wind fields for each month, which are used to estimate (or sample) the parameters of the prior wind field models.

Since only a small region of the North Atlantic is considered, local Cartesian co-ordinates (x, y) are used rather than latitude and longitude, although this is purely a modelling decision and can be changed. Due to errors in data transmission or translation some of the wind fields are found to be highly un-meteorological, in that they consisted of purely divergent flow. All the data used was visually examined and the very few cases where bad wind fields were identified were removed.

3   Models

To implement the Bayesian methodology a flexible, realistic and computationally efficient prior model for wind fields is required. Starting with a very general model, data is used to determine whether parameters could be fixed to a simplified model.

3.1   Modified Bessel function based wind fields

A modified Bessel covariance function based zero mean Gaussian process is the starting model (which actually places prior models over the stream function and velocity potential – see Cornford (1997a)). This covariance function has the general form:

$$C\left(r\right) = E^2 \left(\frac{1}{2^{\nu-1}\Gamma\left(\nu\right)} \left(\frac{r}{L}\right)^\nu K_\nu\left(\frac{r}{L}\right) \right) + \eta^2 \qquad (3)$$

where r is the separation distance of two points, L is a characteristic length scale parameter, E^2 gives the energy (variance) in the wind, ν gives the order of the Bessel function which controls the differentiability (smoothness) of the wind fields and η^2 is the so called 'nugget' variance which represents the noise and the sub-sampling scale variability. For an applied discussion of Gaussian process see Abrahamsen (1997) or Cornford (1997a). Further details of the use of modified Bessel covariance function based Gaussian processes can be found in Cornford (1998).

Essentially a modified Bessel covariance function based Gaussian process is applied to both the stream function (Ψ) and velocity potential (Φ) rather than directly to the wind vector components. This allows control over the ratio of divergence to vorticity in the resulting wind field, and automatically produces valid, positive definite, joint covariance matrices. By controlling the length scales, variance, smoothness and nugget variance a very flexible model for wind fields is

---

[2]By this is meant the standard 10 m vector component wind field

produced with the additional benefit of being able to control the ratio of divergence to vorticity through the relative magnitudes of the variances of the stream function and velocity potential. Using Helmholtz' theorem:

$$u = -\frac{\partial \Psi}{\partial y} + \frac{\partial \Phi}{\partial x}$$

$$v = \frac{\partial \Psi}{\partial x} + \frac{\partial \Phi}{\partial y}$$

Helmholtz' theorem produces a separation of the vector wind field into two scalar components, the stream function which represents purely rotational flow and the velocity potential which defines the divergent flow (Daley, 1991).
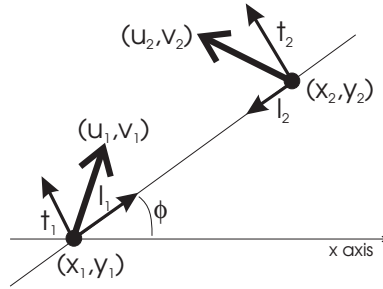

Figure 2: A graphical description of the the conversion of velocity components from (u, v) to (l, t). Note that this is with reference to vector pairs.

Considering two wind vectors (u_1, v_1) and (u_2, v_2) at locations separated by $\boldsymbol{r} = (x_1 - x_2, y_1 - y_2)$ where the Ψ and Φ's are continuous and differentiable, the definitions of expectations and derivatives permit the computation of the wind component covariances in terms of the velocity potential and stream function covariances. This may not appear to have brought very much to the problem, however the simple isotropic covariances on the stream function and velocity potential produce anisotropic, flow dependent covariances on the wind components. Furthermore assuming that the covariances of the stream function and velocity potential are isotropic (that is depend only on r) and that the cross covariance $C_{\Psi\Phi}$ is zero, that is the velocity potential and stream function are uncorrelated, a simple form for the (u, v) covariances is obtained. Defining longitudinal (along flow) and transverse (across flow) velocity components (Figure 2):

$$
\begin{aligned}
l &= u \cos(\phi) + v \sin(\phi) \\
t &= -u \sin(\phi) + v \cos(\phi)
\end{aligned}
\tag{4}
$$

where ϕ is the angle between the x-axis and l. The covariances for l and t are given by:

$$
\begin{pmatrix} C_{ll} & C_{lt} \\ C_{tl} & C_{tt} \end{pmatrix} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{pmatrix} \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}
\tag{5}
$$

Working in radial coordinates gives:

$$
\begin{aligned}
C_{ll}(r) &= -\frac{1}{r}\frac{\partial}{\partial r}C_{\Psi\Psi} - \frac{\partial^2}{\partial r^2}C_{\Phi\Phi} \\
C_{tt}(r) &= -\frac{\partial^2}{\partial r^2}C_{\Psi\Psi} - \frac{1}{r}\frac{\partial}{\partial r}C_{\Phi\Phi} \\
C_{lt} &= C_{tl} = 0.
\end{aligned}
\tag{6}
$$

Thus given $C_{\Psi\Psi}$ and $C_{\Phi\Phi}$ the wind covariances - which are not isotropic in general can be computed from simple scalar isotropic covariance models for the stream function and velocity potential. Working with correlation functions rather than covariances[3] gives:

$$
C_{ll}(r) = -E_{\Psi}^2 L_{e\Psi}^2 \frac{1}{r}\frac{\partial}{\partial r}\rho_{\Psi\Psi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{\partial^2}{\partial r^2}\rho_{\Phi\Phi}
\tag{7}
$$

$$
C_{tt}(r) = -E_{\Psi}^2 L_{e\Psi}^2 \frac{\partial^2}{\partial r^2}\rho_{\Psi\Psi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{1}{r}\frac{\partial}{\partial r}\rho_{\Phi\Phi}
$$

---

[3] The correlation function is simply the covariance divided by the variance (i.e. covariance at $r = 0$).

where E_Ψ^2 and E_Φ^2 are the variances of the rotational and divergent components of the wind respectively. This is a sensible parameterisation because it maintains the independence of the parameters (as much as possible) in the model. This can be expressed using:

$$v^2 = \frac{E_\Phi^2}{E_\Psi^2 + E_\Phi^2}, \qquad E_u^2 = E_\Psi^2 + E_\Phi^2$$

where v^2 gives the ratio of the kinetic energy in the divergent flow (as given by Φ) to that in the total wind flow given by E_u^2 allowing control over the ratio of divergence and vorticity in the resulting flow fields. $L_{e\Phi}^2$ is the squared (effective)[4] length scale associated with the correlation function $\rho_{\Phi\Phi}$ and is defined as:

$$L_{e\Phi}^2 = -\frac{2\rho_{\Phi\Phi}}{\nabla^2 \rho_{\Phi\Phi}}\bigg|_{r=0}$$

where:

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}$$

Finally, the (u, v) covariances are given by:

$$\begin{aligned} C_{uu}\left(r, \phi\right) &= \cos^2\left(\phi\right)C_{ll}\left(r\right) + \sin^2\left(\phi\right)C_{tt}\left(r\right) \\ C_{uv}\left(r, \phi\right) = C_{vu}\left(r, \phi\right) &= \cos\left(\phi\right)\sin\left(\phi\right)\left(C_{ll}\left(r\right) - C_{tt}\left(r\right)\right) \\ C_{vv}\left(r, \phi\right) &= \sin^2\left(\phi\right)C_{ll}\left(r\right) + \cos^2\left(\phi\right)C_{tt}\left(r\right) \end{aligned} \qquad (8)$$

with the full joint covariance matrix being:

$$K_{uv} = \begin{pmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{pmatrix}. \qquad (9)$$

This gives a very general method for constructing covariance functions for wind fields, although care must be taken to ensure that the correlation functions for the velocity potential and stream function satisfy several conditions - largely on the continuity of their derivatives at the origin, see Cornford (1997a).

In order to estimate the parameters of the covariance function a maximum *a posteriori* probability estimate was found to determine the parameters which are most likely given the data. Writing the parameter vector as $\boldsymbol{\xi}$ and the dataset as D_{uv} then the likelihood of the data is:

$$P(D_{uv} \,|\, \boldsymbol{\xi}) = \frac{1}{(2\pi)^{\frac{n}{2}}\det\left(K_{uv}\right)^{\frac{1}{2}}}\exp\left(-\frac{1}{2}D_{uv}'K_{uv}^{-1}D_{uv}\right) \qquad (10)$$

where K_{uv} is the joint covariance matrix of both wind components and n is the sample size[5]. The gradient of this likelihood with respect to the parameters can be computed, and the parameters are optimised using a scaled conjugate gradient algorithm (Bishop, 1995). More details can be found in Cornford (1998). A Bayesian approach (O'Hagan, 1994) is adopted throughout the project thus when implementing the procedure in its most general form we sample over (or optimise) the posterior distribution of the parameters $\boldsymbol{\xi}$ given the data D_{uv}:

$$P(\boldsymbol{\xi} \,|\, D_{uv}) = \frac{P(D_{uv} \,|\, \boldsymbol{\xi})P(\boldsymbol{\xi})}{P(D_{uv})} \propto P(D_{uv} \,|\, \boldsymbol{\xi})P(\boldsymbol{\xi}) \qquad (11)$$

where $P(\boldsymbol{\xi})$ is a prior[6] distribution over the Gaussian process parameters and $P(D_{uv})$ is the normalising constant of the posterior probability, which we can ignore since Markov Chain Monte Carlo (MCMC) (Neal, 1996) methods are used to draw samples from the posterior distribution, $P(\boldsymbol{\xi}|D_{uv})$ or $P(\boldsymbol{\xi}|D_{uv})$ is optimised in which case multiplication by a constant make no difference to the minimum determined.

---

[4] The length scale L is the parameter in the correlation function while the effective length scale L_e gives the gives the true correlation length scale, that is the distance at which the correlation function has fallen to 0.05.

[5] n will be twice the number of observations, since each observation comprises a pair of wind components.

[6] This raises a notational issue. Strictly these distributions are hyper-priors, since these are priors on hyper-parameters in our prior wind field model. The difference is not made explicit in the text since it is felt that the context will clearly show which prior model is being referenced.

The modified Bessel covariance function based Gaussian process is a very flexible model but is rather slow to compute due to the necessity of recomputing a non-integer order Bessel function (Press *et al.*, 1992). Figure 3 shows cross sections through the maximum *a posteriori* probability values of the individual parameters in the modified Bessel covariance function based Gaussian process. These show that the maximum (= minimum negative log probability) is well defined for all parameters, although this is less marked for the smoothness parameters. Note that to ensure that the vorticity is continuous, the smoothness parameter must be at least two. These profiles are similar in appearance to those obtained by using maximum likelihood methods using the likelihood in (11) without any priors (not shown).

The prior distributions are Weibull (since all parameters must be positive) and are shown in Figure 4. These prior distributions have the general form:

$$P(\xi) = g * d * \left(\xi^{g-1} \right) \exp \left(-d * \xi^g \right) \tag{12}$$

where the parameters g, d are determined using expert knowledge (Dan Cornford) since the parameters of the Gaussian process all have physical interpretations. Note identical priors on the length and smoothness scales of the stream function and velocity potential were imposed. The nugget variance, in particular, is strongly constrained since it is not well identified from the data, other than having small magnitude. Thus the constraint imposed by the priors produces a reasonable, well defined maximum *a posteriori* probability value, which is consistent with the data.

Figure 5 shows the maximum *a posteriori* probability values for the parameters of the modified Bessel covariance function based Gaussian process and how they vary over the course of the year. This brings out several features of the data. The same prior distributions are used as shown in Figure 4. The annual cycle is clear, notably in the variance of the wind field. This corresponds to mid-latitude winds being driven largely (and indirectly) by the pole-equator temperature gradient which is largest in winter, and thus the winds have most energy during this season. The energy in the rotational wind (top right, Figure 5) can be seen to be more than ten times greater than the energy in the divergent wind (top left, Figure 5). Note that characteristic length scales of approximately $500\ km$ which correspond to effective correlation scales of $\sim 850\ km$ are consistent with the scale of meteorological features such as cyclones, fronts and anticyclones.

The bottom two plots of Figure 5 show the smoothness scale maximum *a posteriori* probability values. These can be seen to be roughly 2.5. If these are set to 2.5 exactly then the modified Bessel covariance function simplifies to a polynomial-exponential covariance function and the time spent in the computation of the covariance matrix can be reduced by a factor of 4.81[7]. The bottom two graphs of Figure 3 indicate that the posterior distribution of the smoothness parameters is relatively flat around the maximum *a posteriori* probability values, so this simplification is near the maximum *a posteriori* probability values. Setting the smoothness parameters of the stream function and velocity potential to 2.5 is consistent with the data, the physics (that is it ensures that the vorticity and divergence are continuous) and greatly improves modelling speed.

3.2   Polynomial-exponential based wind fields

By fixing the smoothness parameters to 2.5 the covariance function is given by:

$$C\left(r \right) = E^2 \left(1 + \frac{r}{L} + \frac{r^2}{3L^2} \right) \exp \left(-\frac{r}{L} \right) + \eta^2 \tag{13}$$

where, as before, r is the separation distance of two points, L is a characteristic length scale parameter, E^2 is the energy (variance) in the wind and η^2 is the so called 'nugget' variance. Using this form of covariance in exactly the same way as the previous covariance was applied, gives a flexible yet fast wind field model. The maximum *a posteriori* probability values of the parameters of the zero mean polynomial-exponential covariance function based Gaussian process are shown in Figure 6.

---

[7] This figure is computed averaging ten runs of the `profile` command in MATLAB on a Silicon Graphics Indy R5000.

So far all the Gaussian processes considered have been zero mean. Since there are generally predominant wind directions at every location on the Earth's surface a zero mean model may be improved by the addition of a location dependent mean term.

3.3 Adding a non-zero mean

In order to improve the model and to produce meaningful marginal probabilities of the wind in each cell a non-zero mean component (on the wind components) is added to the Gaussian process model. Since only a small region of the North Atlantic is being considered a second order polynomial is used:

$$m_u(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \tag{14}$$

and:

$$m_v(x, y) = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy \tag{15}$$

where (x,y) are the (Cartesian) co-ordinates and the coefficients (a_i and b_i) are determined by maximising their *a posteriori* probability jointly with the parameters of the polynomial-exponential covariance function. Weakly informative, proper, Gaussian priors are put on the coefficients, corresponding to ridge regression (O'Hagan, 1994). In all cases the variance of the Gaussian is 100 while the mean is zero for all but a_0 for which it is 4 ms^{-1}. This corresponds to a very vague, realistic prior belief that the mean winds are light westerlies in the North Atlantic.

The likelihood of the non-zero mean Gaussian process wind field model is given by:

$$P(D_{uv} \,|\, \boldsymbol{\xi}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det (K_{uv})^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \left(D_{uv} - m_{uv} \right)' K_{uv}^{-1} \left(D_{uv} - m_{uv} \right) \right) \tag{16}$$

where:

$$m_{uv} = \begin{bmatrix} m_u(x, y) \\ m_v(x, y) \end{bmatrix} \tag{17}$$

and $\boldsymbol{\xi}$ now also includes the mean coefficients. Together with the priors over the parameters (as described above) samples from the posterior distribution of the parameters can be taken or their maximum *a posteriori* probability values can be found (given the derivatives which are straightforward to compute). The mean parameters are initialised to the least squares fit of the polynomials, (14) and (15), to the monthly mean wind vector for the region considered. The covariance parameters are initialised to the good guess values determined using geophysical knowledge.

Figure 7 shows the maximum *a posteriori* probability values of the covariance parameters for the non-zero mean Gaussian process. This can be contrasted with Figure 6. The mean function included in the results displayed in Figure 7 has the expected effect on the covariance parameters. Some of the energy in the winds is included in the mean component and thus the variances of the divergent and rotational winds are reduced. The length scale is also somewhat reduced as the more persistent, larger scale features are present in the mean. The nugget variance remains essentially unchanged.

Figure 8 shows the maximum *a posteriori* probability mean wind fields computed for four months of the year. These use three years of data and thus still contain 'non-mean' features (that is long-lived, yet nevertheless transitory, features) but the wind fields are consistent with climatological expectations. The wind field for April is probably less representative of the true mean and reflects the dominance of blocking anticyclones over Western Europe in recent years during the spring months.[8]

---

[8]This is an interesting point, since there really is not such thing as a stable (temporally stationary) climate thus it could be argued that a three year average is appropriate in the current scenario of potentially rapid global climate change.

Figure 9 shows some example wind fields simulated from the non-zero mean polynomial-exponential covariance function based Gaussian process for four months. These simulated wind fields capture many of the features typical of real wind fields (although they lack fronts). Since these models will form the prior models for wind fields in the next stage of processing where wind fields are determined on the basis of scatterometer observations and prior wind field models it is important to have good prior models. Visual assessment indicates that the proposed models meet the requirement.

3.4   Determination of the marginal probabilities

When using the inverse model in disambiguation (2) the local (marginal) probabilities $P(u_i, v_i)$ of the wind vectors in each cell are required. These are obtained as the marginal probabilities of the full joint wind field model $P(U, V)$. Since $P(U, V)$ is multivariate normal, the marginal probabilities are simply given by:

$$P(u_i, v_i) = \frac{1}{(2\pi)^{\frac{1}{2}} \det (K_{uv,ii})^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \left((u_i, v_i) - m_{uv,i} \right)' K_{uv,ii}^{-1} \left((u_i, v_i) - m_{uv,i} \right) \right) \qquad (18)$$

where the subscript i indicates that only to the value of the i'th cell is being referred to. Since a stationary covariance function is being used the variance-covariance matrix $K_{uv,ii}$ will be the same for all cells, however the mean will change. This prior local model then gives a good estimate of the unconditional local distribution of wind vectors for any cell.

4   Sampling or Optimisation?

When the NEUROSAT project was proposed a fully Bayesian approach was envisaged where the integral over the parameters in the Gaussian process (whether it be zero or non-zero mean) would be approximated using Monte Carlo methods - that is averaging over the wind fields computed using samples from the posterior distribution of the Gaussian process parameters of the wind field prior model. This will be extremely time consuming since it would necessitate repeated computation (or storage) of the inverse covariance matrices for all samples (that is covariance parameters) at each step in the computation of the posterior probability (or log probability = energy in the Markov Chain Monte Carlo context) of the wind fields given the scatterometer data.

Thus, operationally, it may be more effective to fix the Gaussian process parameters at their maximum *a posteriori* probability values which means that the inverse covariance matrix need only be computed once and stored leaving each sampling step with a (large) matrix multiplication rather than inversion and multiplication over many duplicate chains. Storage of multiple inverse covariance matrices is impractical due to their size (722×722 for a full scene, even larger for a swathe).

5   Potential Model Improvements

Although the model fits the data well and produces realistic wind fields when used in a generative sense, there are several improvements that could be envisaged in future work. Fronts are not represented in the model, but are common features in the atmosphere. Other work addresses fronts and will not be discussed here.

The model has a parametric mean which has a fixed functional form. This implicitly defines the (unknowable and arbitrary) split between the mean component and the stochastic component. Since we cannot know whether such a split is justified by the data we must accept a level of arbitrariness here. The second order polynomial was chosen to represent the mean since it was felt that the climatological mean (which this is to some extent representing) has a simple and

smooth form in the North Atlantic. However the exact functional form is arbitrary and so could be changed.

Currently a small region of the North Atlantic is considered, however, operationally all of the Earth's oceans must be included. This is a highly complex task. The mean function could be defined piece-wise over regions of the Earth's surface, or a more flexible form based on spherical geometry (spherical harmonics) could be used.

The covariance function would be still more complicated since there are compelling reasons to believe that it would not be stationary. For example, the mid-latitudes experience significantly different weather patterns from the sub-tropics, and thus the covariance functions will change from region to region. There is no existing theory that could produce such non-stationary covariance functions in a principled framework.

6   Conclusions

A flexible Gaussian process wind field model based on a modified Bessel covariance function was introduced. This was shown to be a good model of assimilated ECMWF winds and the maximum *a posteriori* probability values for the parameters were shown. These physically interpretable parameters agree with *a priori* beliefs. Fixing the smoothness parameters in the modified Bessel covariance function to 2.5, close to their maximum *a posteriori* probability values, allows a simplification of the covariance function, resulting in the polynomial-exponential covariance function.

Introducing a non-zero mean to the polynomial-exponential covariance function based Gaussian process allows a more sensible application of the model when using the *scaled-likelihood trick* to compute the posterior distribution of the wind fields using the inverse scatterometer model. The non-zero mean allows the calculation of the local probability of the wind vector in each cell as the marginal probability of the Gaussian process for that cell alone. The maximum *a posteriori* probability values of the parameters have been computed for both zero and non-zero mean Gaussian processes using the polynomial-exponential covariance function.

The posterior distribution of the Gaussian process parameters has also been sampled from using both Metropolis and hybrid Monte Carlo samplers (Neal, 1996). However practical considerations suggest the use of the maximum *a posteriori* probability parameters rather than implementing a fully Bayesian method. Possible improvements to the model are also noted.

References

Abrahamsen, P. 1997. A Review of Gaussian Random Fields and Correlation Functions, Second Edition. Technical Report 917, Norwegian Computing Center.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Cornford, D. 1997a. Random Field Models and Priors on Wind. Technical Report NCRG/97/023, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK.

Cornford, D. 1997b. Surface Wind Fields (on Earth). Technical Report NCRG/97/022, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK.

Cornford, D. 1998. Flexible Gaussian Process Wind Field Models. Technical Report NCRG/98/017, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK.

Cornford, D. and I. T. Nabney 1998. NEUROSAT: An Overview. Technical Report NCRG/98/011, Neural Computing Research Group, Aston University, Aston Triangle, Birmingham, UK.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge: Cambridge University Press.

Neal, R. M. 1996. *Bayesian Learning for Neural Networks*. Springer. Lecture Notes in Statistics 118.

O'Hagan, A. 1994. *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*. London: Edward Arnold.

Press, W. H., S. A. Teukolsky, W. T. Vettering, and B. P. Flannery 1992. *Numerical Recipes in C* (2nd Edition ed.). Cambridge, UK: Cambridge University Press.

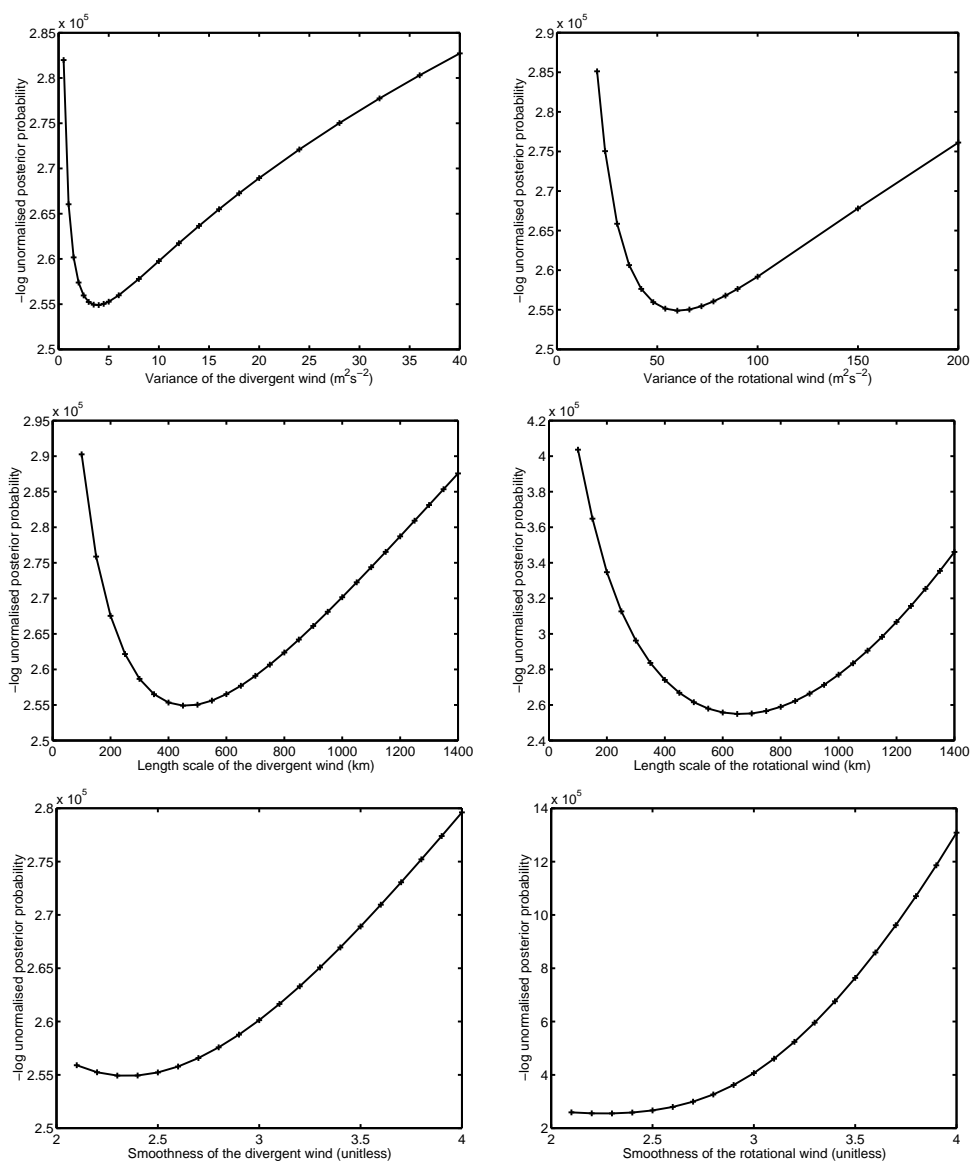Wallace, J. M. and P. V. Hobbs 1977. *Atmospheric Science - An Introductory Survey*. London: Academic Press.
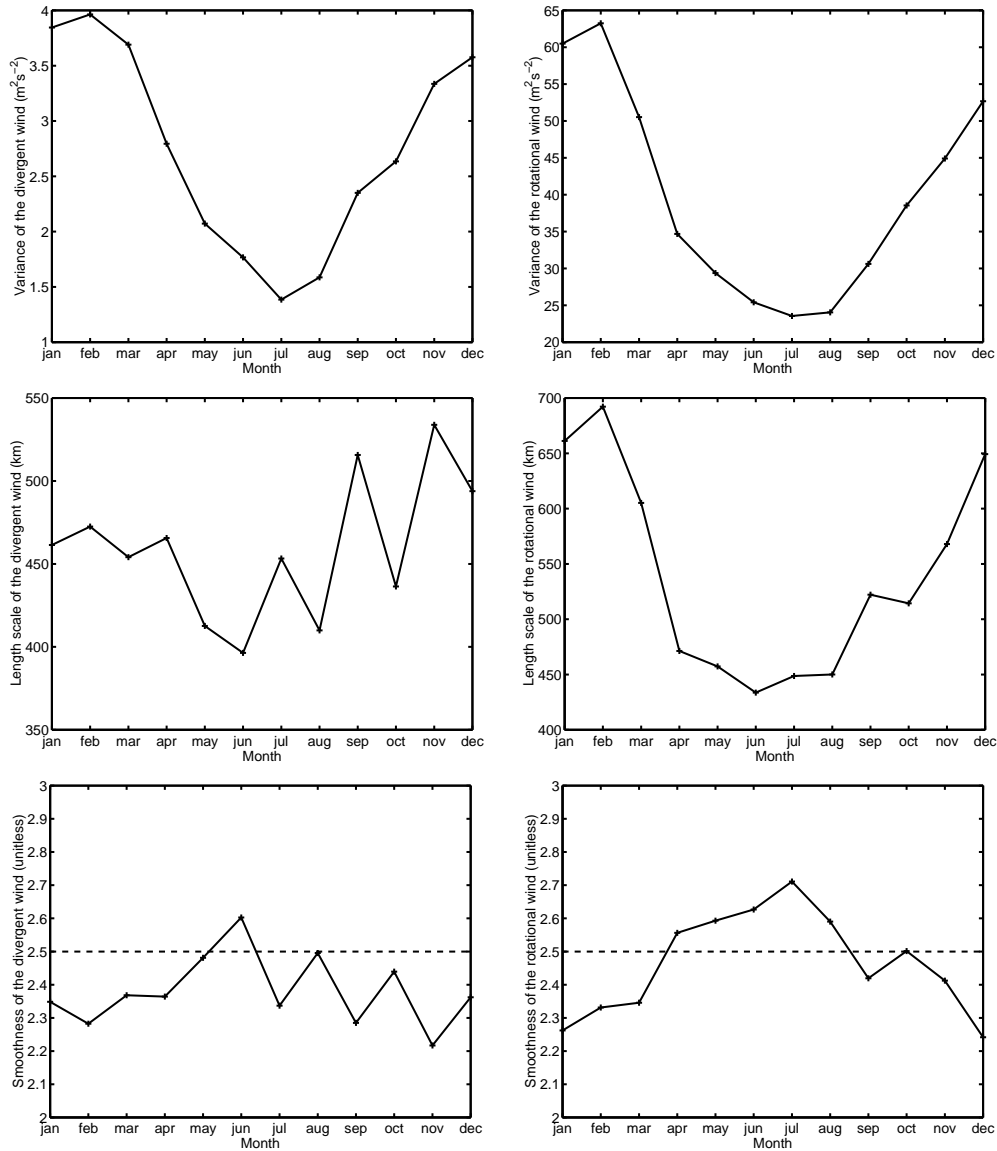
Figure 3: Cross sections through the posterior probability of the variance, length scale and smoothness parameters of the modified Bessel covariance function based Gaussian process respectively from top to bottom at the maximum *a posteriori* probability values of the other parameters for the month of January. Parameters for the velocity potential (divergent flow) are on the left and stream function (rotational flow) on the right.
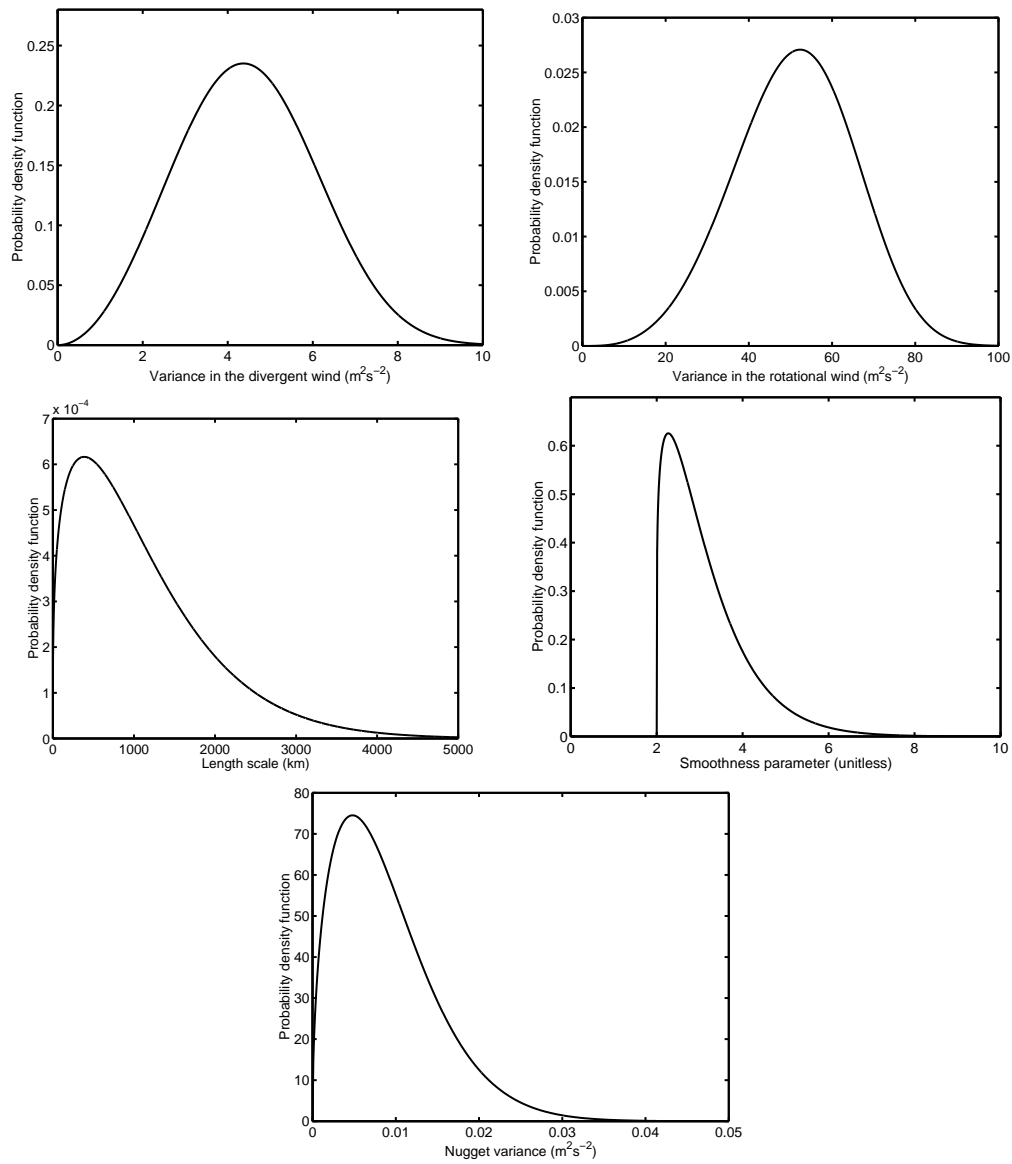
Figure 4: The prior distributions of the variance in the divergent and rotational winds (top), the length scale and smoothness scale (middle) and the nugget variance (bottom).

Figure 5: The maximum *a posteriori* probability values of the variance, length scale and smoothness parameters of the modified Bessel covariance function based Gaussian process respectively from top to bottom. Parameters for the velocity potential (divergent flow) are on the left and stream function (rotational flow) on the right. Each figure shows the maximum *a posteriori* probability value of the respective parameter for the 12 months of the year. The dotted line in the bottom two figures gives the value of the smoothness parameters for which the modified Bessel covariance function can be simplified to a polynomial-exponential covariance function.
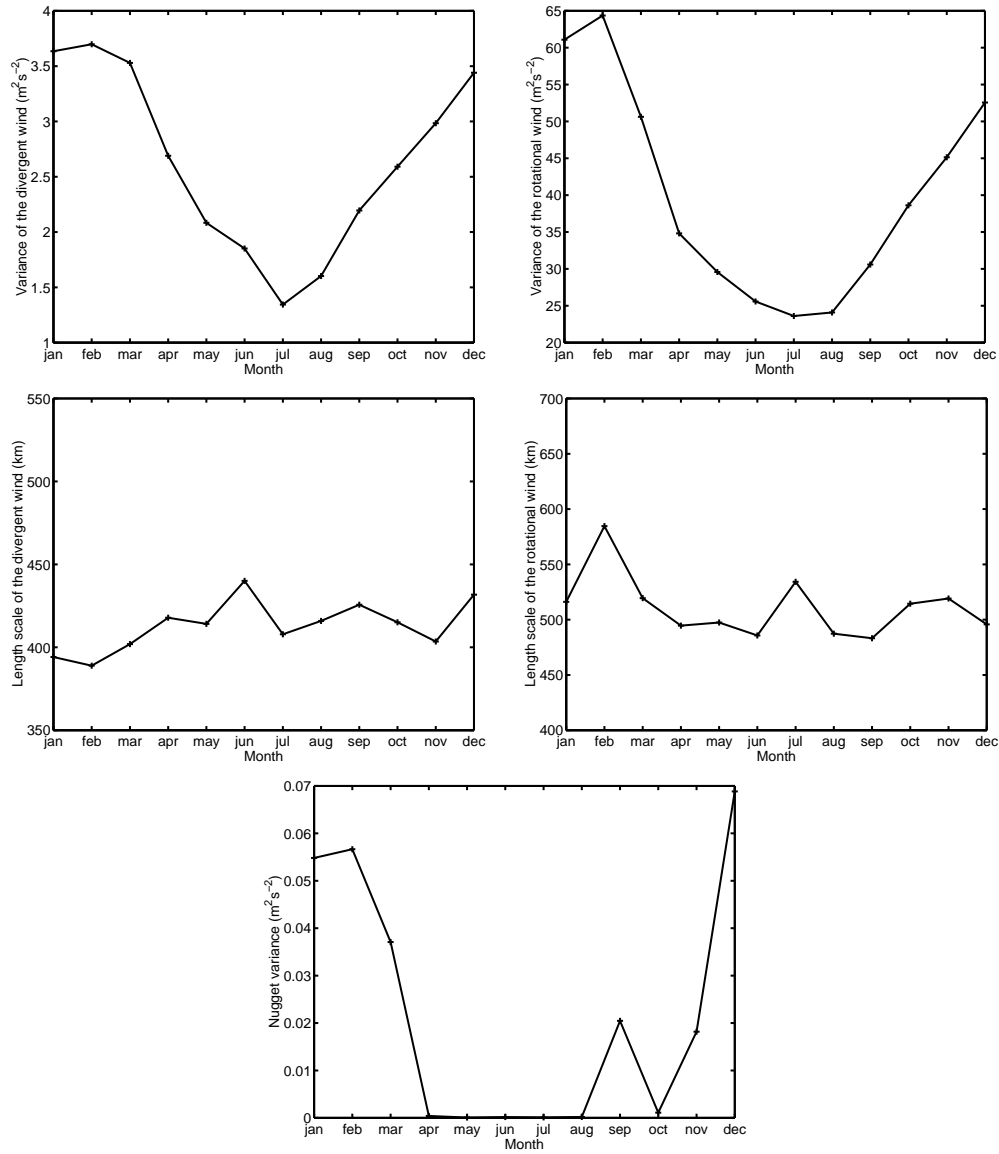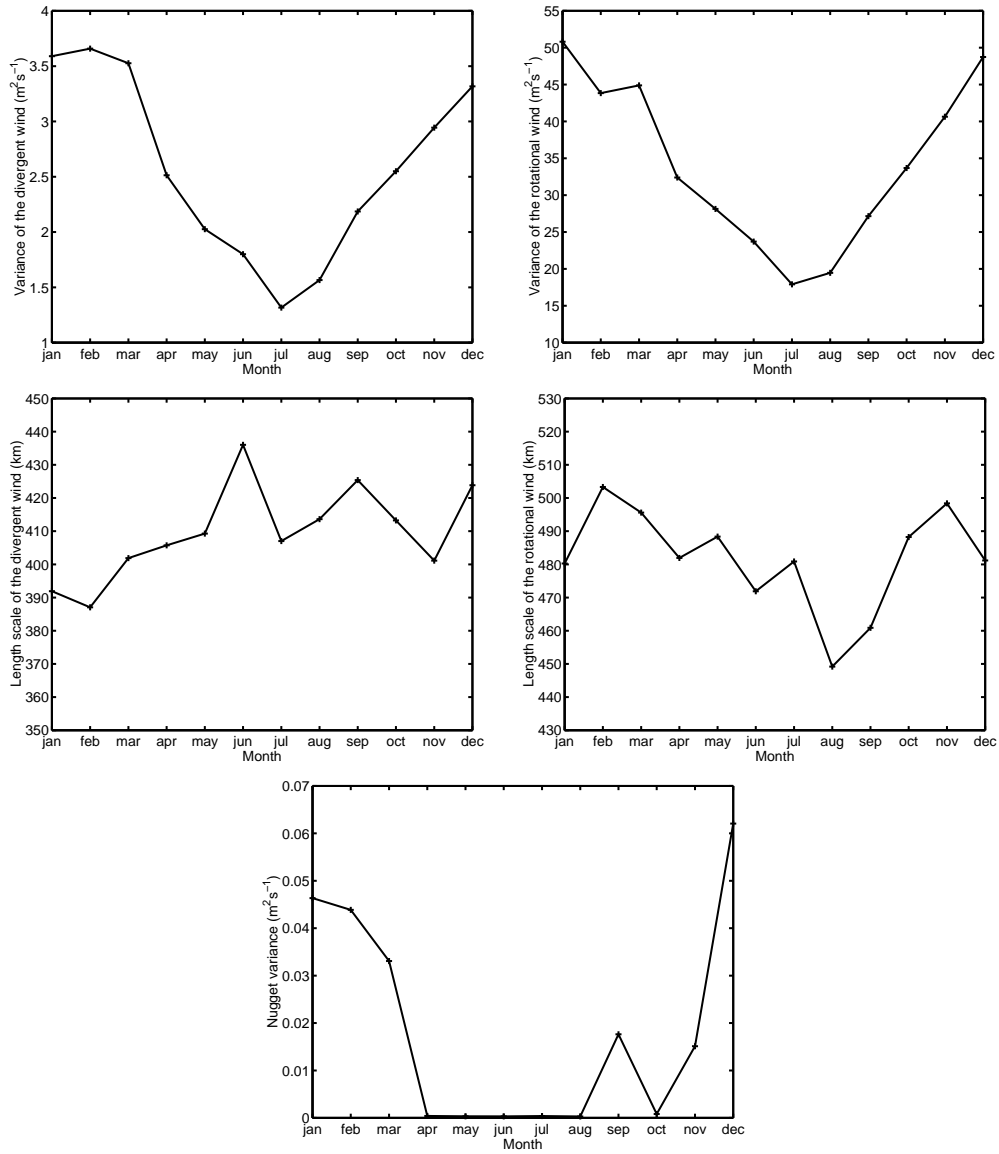
Figure 6: The maximum *a posteriori* probability values of the variance, length scale and the nugget variance of the polynomial-exponential covariance function based Gaussian process respectively from top to bottom. Parameters for the velocity potential (divergent flow) are on the left and stream function (rotational flow) on the right. Each figure shows the maximum *a posteriori* probability value of the respective parameter for the 12 months of the year.

Figure 7: The maximum *a posteriori* probability values of the variance, length scale and the nugget variance of the non-zero mean polynomial-exponential covariance function based Gaussian process respectively from top to bottom. Parameters for the velocity potential (divergent flow) are on the left and stream function (rotational flow) on the right. Each figure shows the maximum *a posteriori* probability value of the respective parameter for the 12 months of the year.
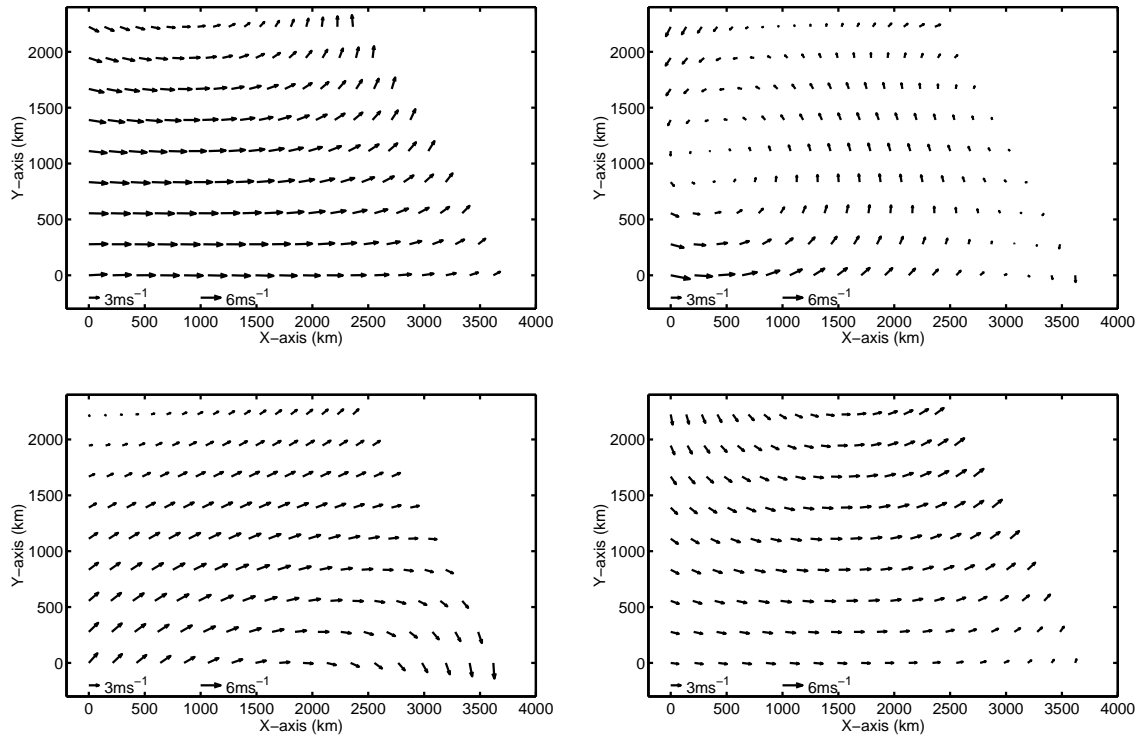
Figure 8: The maximum *a posteriori* probability values of the mean for January (top left), April (top right), July (bottom left) and October (bottom right) for the North Atlantic.
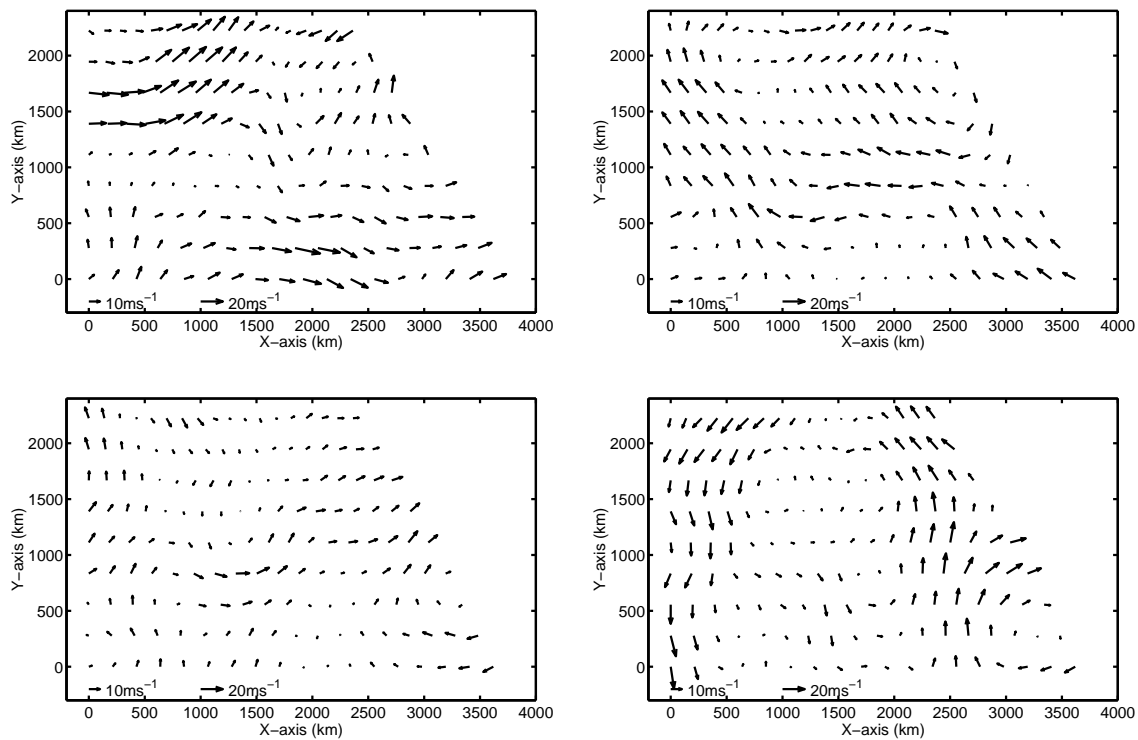

Figure 9: Simulated wind fields using maximum *a posteriori* probability values for all the parameters of the non-zero mean polynomial-exponential covariance function based Gaussian process for January (top left), April (top right), July (bottom left) and October (bottom right) for the North Atlantic.