



Neural Computing Research Group
Dept of Computer Science & Applied Mathematics
Aston University
Birmingham B4 7ET
United Kingdom
Tel: +44 (0)121 333 4631
Fax: +44 (0)121 333 4586
<http://www.ncrg.aston.ac.uk/>

Flexible Gaussian Process Wind Field Models

Dan Cornford
d.cornford@aston.ac.uk

Technical Report NCRG/98/017

August 3, 1998

Abstract

This technical report builds on previous reports to derive the likelihood and its derivatives for a Gaussian process with a modified Bessel function based covariance function. The full derivation is shown. The likelihood (with gradient information) can be used in maximum likelihood procedures (i.e. gradient based optimisation) and in Hybrid Monte Carlo sampling (i.e. within a Bayesian framework).

Keywords: *Gaussian process, wind field, modified Bessel function, Helmholtz' theorem, maximum likelihood.*

1 Introduction

This section briefly outlines the the reasons for investigating the modified Bessel covariance function based Gaussian processes. More details on the overall aims of the NEUROSAT project can be found in Cornford and Nabney (1998) and details of the reasons for using Gaussian process priors together with some background on Gaussian process models can be found in Cornford (1997) or Abrahamsen (1997). We shall provide a brief re-cap here.

1.1 Bayes theorem and wind fields

Recall Bayes theorem, which states:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{normalising factor}} \quad (1)$$

If we observed a wind field D_{uv} , then given that we have a probabilistic model for wind fields we can assess the data likelihood under that model. Let us assume that we have formulated a random field model for U, V (that is a (spatial) wind field made up of several local wind measurements (u, v)) which depends on hyper-parameters ξ ¹. We can use Bayes theorem in the form:

$$P(\xi | D_{uv}) \propto P(D_{uv} | \xi)P(\xi) \quad (2)$$

to examine and sample from the posterior distribution of ξ given a wind field D_{uv} . In the first instance this will be useful to assess suitable hyper-priors on ξ .

In the general problem of obtaining wind vectors from the scatterometer measurements we must be able to develop realistic wind fields from our model for U, V . In the general problem of wind retrieval from scatterometer measurements we obtain local so called sigma nought measurements (denoted $\sigma^\circ = (\sigma_f^\circ, \sigma_m^\circ, \sigma_a^\circ)$) for a series of cells covering a small region of the ocean's surface, which we together denote by Σ° . We wish to infer the wind field, U, V , from these measurements and we do so using Bayes theorem again:

$$P(U, V | \Sigma^\circ) \propto \left(\prod_i P(\sigma_i^\circ | u_i, v_i) \right) P(U, V) \quad (3)$$

where $P(\sigma_i^\circ | u_i, v_i)$ is a local 'forward' model (that is a probabilistic sensor model), $P(U, V)$ is our prior (Gaussian process) model for wind fields and $P(U, V | \Sigma^\circ)$ is the posterior distribution of the wind field given the scatterometer measurements. Thus the prior model $P(U, V)$, which depends on hyper-parameters ξ , will be critically important in the NEUROSAT project.

Thus we use historical data to set the hyper-priors on the hyper-parameters ξ which will produce the most realistic prior wind field models, $P(U, V)$. However, care must be taken since these hyper-priors will vary in space and time. In the final formulation we can use Monte Carlo integration to remove the dependence on the hyper-parameters.

1.2 Posterior over ξ

Although in earlier sections we call ξ the hyper-parameters (as they are in the context of Equation (3)) from now on we will refer to them simply as parameters since this report deals exclusively with the prior wind field model in Equation (3). The aim initially will be to sample from the posterior distribution of ξ (Equation (2)) in order to produce simulated wind fields (using a generative version of the Gaussian process model) and visually assess their adequacy (and hence that of the underlying model). It may also be necessary to determine a unique (most probable) parameter set so that we remove the need to integrate these parameters out from the final model (the 'correct' thing to do) and simply use the maximum *a posteriori* probability values for ξ in order to make the implementation sufficiently fast for operational use.

1.3 Likelihood of the data given ξ

In order to compute the posterior distribution of ξ we need to be able to compute the likelihood of the random field model given the data. Consider the log likelihood (Neal, 1997):

$$\log(P(D_{uv} | \xi)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det(K_{uv})) - \frac{1}{2} D'_{uv} K_{uv}^{-1} D_{uv} \quad (4)$$

¹We assume that given ξ the model form is also completely specified, and later we drop the dependence on ξ totally to clarify the notation.

where $D_{uv} = (u_1, u_2, \dots, u_n, v_1, v_2, \dots, v_n)'$ with n the number of observations and:

$$K_{uv} = \begin{pmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{pmatrix} \quad (5)$$

where the C_{uu} 's are the (cross)-covariance matrices of the wind components (u, v) (see Cornford (1997)). In this report we assume a zero mean Gaussian process since we are more interested in priors which mainly embody the local *structure* in the wind fields, rather than the mean structure. We are also concerned to ensure that we do not bias our satellite derived winds preferentially in the prevailing direction, since it is often those winds not in the prevailing direction which are most important to identify. Later work will include non-zero means.

1.4 Prior over ξ

Initially the prior over the parameters ξ will be uninformative since we want to examine the information in the data, rather than confirm our prejudices. The prior will, however restrict the parameters in ξ to sensible values which could be physically justified². In the full analysis the priors used are not uninformative, but introduce very 'weak' information. These are discussed later.

1.5 Application

In order to implement the above methodology we have to select a suitable model for the wind fields. We have already stated that it will be a random field model, thus our remaining choice will be on the functional form of the covariance function we use. We assume a stationary random field, and following developments outlined in Cornford (1997) we apply Helmholtz theorem to split the wind into divergent and rotational components, thus allowing us to control the ratio of vorticity to divergence. We shall try to encompass as broad a range of processes as possible by using a very flexible covariance function based on Bessel functions. The details of the computations are given below.

2 The Bessel Based Covariance

To start we shall re-cap the implications of using Helmholtz' theorem to model the vector winds as the combination of two scalar fields. Then we introduce the modified Bessel covariance function and show how it can be used.

2.1 Helmholtz' theorem

The Helmholtz theorem allows us to separate a vector flow field into two scalar components - a non-divergent (rotational) component and an irrotational (divergent) component (Daley, 1991). Then if we define Ψ (stream function) and Φ (velocity potential) using:

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

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

²This is a benefit of the Gaussian process model, since the parameters all have a physical interpretation.

Helmholtz theorem is useful because it allows us to manipulate the covariance function for the vector winds into two scalar covariances for Ψ and Φ . Here we follow the derivation of Daley (1991). If we consider two wind vectors (u_1, v_1) and (u_2, v_2) at locations separated by $\mathbf{r} = (x_1 - x_2, y_1 - y_2)$ then if the Ψ and Φ 's are continuous and differentiable, thus using the definitions of expectations and derivatives we can express the wind covariances C_{uu} etc. in terms of the velocity potential and stream function covariances $C_{\Phi\Phi}$ etc.. This may not appear to have brought very much to the problem, however we now find that we need not assume isotropic covariances on the wind components, but we can maintain the simplicity of isotropic covariances on the velocity potential and stream function covariances. 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 - we obtain a simple form for the (u, v) covariances. We define longitudinal (along flow) and transverse (across flow) velocity components:

$$\begin{aligned} l &= u \cos(\phi) + v \sin(\phi) \\ t &= -u \sin(\phi) + v \cos(\phi) \end{aligned} \quad (6)$$

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

$$\begin{pmatrix} C_{ll} & C_{lt} \\ C_{lt} & C_{tt} \end{pmatrix} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{pmatrix} \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \quad (7)$$

If we work in radial coordinates we can now write:

$$\begin{aligned} C_{ll}(r) &= -\frac{1}{r} \frac{\partial}{\partial r} C_{\Psi\Psi} - \frac{\partial^2}{\partial r^2} C_{\Phi\Phi} \\ C_{tt}(r) &= -\frac{\partial^2}{\partial r^2} C_{\Psi\Psi} - \frac{1}{r} \frac{\partial}{\partial r} C_{\Phi\Phi} \\ C_{lt} &= C_{tl} = 0 \end{aligned} \quad (8)$$

Thus given $C_{\Psi\Psi}$ and $C_{\Phi\Phi}$ we can compute the wind covariances - which are not isotropic in general - based on simple scalar isotropic covariance models for the stream function and velocity potential. Computing the covariances in this way ensures that the joint covariance matrix of (u, v) is positive definite. In practice we tend to work with correlation functions rather than covariances³ giving:

$$\begin{aligned} C_{ll}(r) &= -E_{\Psi}^2 L_{e\Psi}^2 \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Psi\Psi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{\partial^2}{\partial r^2} \rho_{\Phi\Phi} \\ C_{tt}(r) &= -E_{\Psi}^2 L_{e\Psi}^2 \frac{\partial^2}{\partial r^2} \rho_{\Psi\Psi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Phi\Phi} \end{aligned} \quad (9)$$

where E_{Ψ}^2 and E_{Φ}^2 are the variances of the rotational and divergent components of the wind respectively. We choose this parameterisation because this maintains the independence of the parameters (where possible) in the model. This can be expressed using:

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

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 . This allows us to explicitly control the ratio of divergence and vorticity in the resulting flow fields. $L_{e\Phi}^2$ is the squared (effective) length scale associated with the correlation function $\rho_{\Phi\Phi}$ and is defined as:

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

where:

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

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

Finally we can compute the (u, v) covariances using:

$$\begin{aligned} C_{uu}(r, \phi) &= \cos^2(\phi) C_{U}(r) + \sin^2(\phi) C_{tt}(r) \\ C_{uv}(r, \phi) = C_{vu}(r, \phi) &= \cos(\phi) \sin(\phi) (C_U(r) - C_{tt}(r)) \\ C_{vv}(r, \phi) &= \sin^2(\phi) C_U(r) + \cos^2(\phi) C_{tt}(r) \end{aligned} \quad (10)$$

This gives us a very general method for constructing covariance functions for wind fields, although we must take care 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, 1997)). We shall now illustrate the above using the modified Bessel covariance function.

It is worth noting at this point that in general for observed winds in the atmosphere (as we are studying), rather than analysis corrections⁴, the stream function and velocity potential are often strongly correlated. This is due to vorticity and divergence generally attaining their maxima at fronts and near the centre of cyclones. Thus one improvement to this formulation would be to allow correlation between the velocity potential and stream function, although this would further complicate the implementation.

2.2 The modified Bessel covariance function

The correlation function which we shall use is based on a modified Bessel function of the second kind with a distance growth factor. This form of correlation function was chosen because it has a form which is more likely to give a good wind field model prior due to its flexibility. In the following we drop the distinction between $\rho_{\Psi\Psi}$ and $\rho_{\Phi\Phi}$ since these will have identical forms, with different parameters. The modified Bessel correlation function is given by:

$$\rho(r; L, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{r}{L}\right)^{\nu} K_{\nu}\left(\frac{r}{L}\right) \quad (11)$$

where $\nu \geq 0$ and $L > 0$ and an example can be seen in Figure (1). This form of correlation function does not permit negative correlations and produces realisations which are $\lceil \nu \rceil$ times mean square differentiable. In order to use the function in practice we require $\nu \geq 1$ to ensure that the derivatives of $\rho(r; L, \nu)$ exist (and thus the wind correlations exist).

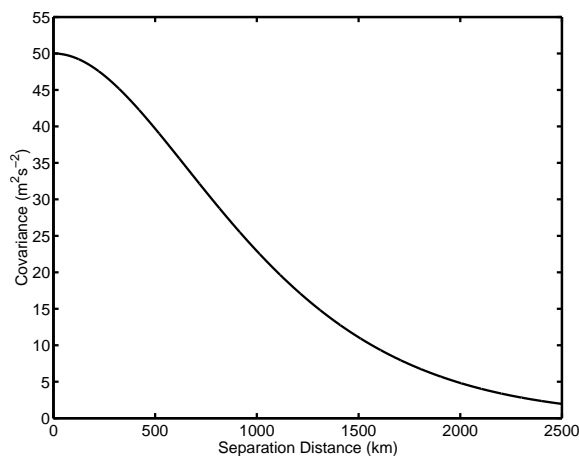


Figure 1: An example of the modified Bessel covariance function with $E^2 = 50 \text{ m}^2 \text{ s}^{-2}$, $L = 400 \text{ km}$, $\nu = 2.5$ and $\eta^2 = 0 \text{ m}^2 \text{ s}^{-2}$.

⁴Analysis corrections are the observed minus the forecast winds in the context of numerical weather prediction data assimilation.

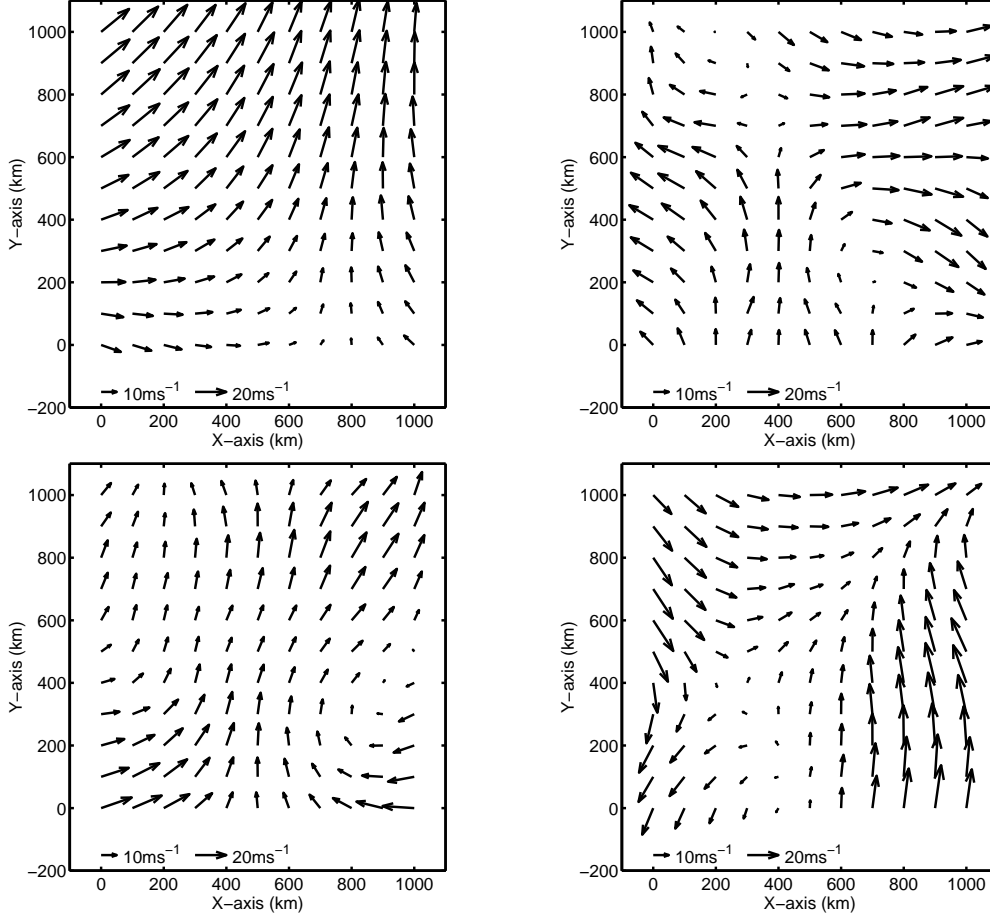


Figure 2: Examples of realisations simulated from the modified Bessel covariance function based Gaussian process with various physically motivated parameter settings. The top left figure is based on physically realistic parameters, the top right has very short length scales (150 km), the bottom left is purely rotational flow and the bottom right is purely divergent flow.

In order to compute ρ_{ui} and ρ_{tt} we need:

$$\frac{1}{r} \frac{\partial \rho(r)}{\partial r} = -\frac{1}{\Gamma(\nu) 2^{\nu-1} L^2} \left(\frac{r}{L}\right)^{(\nu-1)} K_{\nu-1}\left(\frac{r}{L}\right) \quad (12)$$

$$\frac{\partial^2 \rho(r)}{\partial r^2} = -\frac{1}{\Gamma(\nu) 2^{\nu-1} L^2} \left(\left(\frac{r}{L}\right)^{\nu-1} K_{\nu-1}\left(\frac{r}{L}\right) - \left(\frac{r}{L}\right)^{\nu} K_{\nu-2}\left(\frac{r}{L}\right) \right) \quad (13)$$

These derivatives are obtained using the identity:

$$\frac{\partial}{\partial z} (z^{\nu} K_{\nu}(z)) = -z^{(\nu)} K_{\nu-1}(z)$$

(see (Abramowitz and Stegun, 1972)) and the chain rule:

$$\frac{\partial \rho}{\partial r} = \frac{\partial \rho}{\partial z} \frac{\partial z}{\partial r}, \quad z = \frac{r}{L}.$$

When using the formulae for the derivatives care must be taken when $r \rightarrow 0$ and a modified form must be used:

$$\lim_{r \rightarrow 0} \left(\frac{1}{r} \frac{\partial \rho(r)}{\partial r} \right) = -\frac{1}{2(\nu-1)L^2}$$

$$\lim_{r \rightarrow 0} \left(\frac{\partial^2 \rho(r)}{\partial r^2} \right) = -\frac{1}{2(\nu-1)L^2}$$

using:

$$K_\nu(z) \rightarrow \frac{1}{2}(z)^{-\nu} 2^{\nu-1} \Gamma(\nu), \quad z \rightarrow 0$$

and:

$$\Gamma(\nu) = (\nu - 1) \Gamma(\nu - 1).$$

Thus we can now compute ρ_{ll} and ρ_{tt} using (9) with:

$$L_e^2 = -\left. \frac{2\rho(r)}{\nabla^2 \rho(r)} \right|_{r=0} = 2(\nu - 1)L^2$$

which can in turn be used to compute ρ_{uu} , ρ_{uv} and ρ_{vv} using (10). We then finally combine these correlations with the wind variance to produce the desired covariance matrix using (14). In practice it is also useful to add a certain amount of ‘noise’ or ‘jitter’ (Neal, 1997) to the diagonal of the covariance matrix to stabilise the computations and to represent the real noise in the data such that:

$$K_{uv} = \begin{pmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{pmatrix} + \eta^2 \mathbf{I} \quad (14)$$

where \mathbf{I} is the $2n \times 2n$ identity matrix and η^2 is the noise variance⁵.

Figure (2) show several realisations (i.e. wind fields) produced using the modified Bessel covariance function based Gaussian process as a generative model. All plots have $\eta^2 = 0.1 \text{ m}^2 \text{ s}^{-2}$ and $\nu = 2.5$. The top left plot has $L = 500 \text{ km}$ and $E_\Psi^2 = 70 \text{ m}^2 \text{ s}^{-2}$ and $E_\Phi^2 = 7 \text{ m}^2 \text{ s}^{-2}$, while the top right has $L = 150 \text{ km}$ and $E_\Psi^2 = 40 \text{ m}^2 \text{ s}^{-2}$ and $E_\Phi^2 = 4 \text{ m}^2 \text{ s}^{-2}$. The bottom plots both have $L = 200 \text{ km}$, with the left plot having $E_\Psi^2 = 50 \text{ m}^2 \text{ s}^{-2}$ and $E_\Phi^2 = 0 \text{ m}^2 \text{ s}^{-2}$ and the right $E_\Psi^2 = 0 \text{ m}^2 \text{ s}^{-2}$ and $E_\Phi^2 = 50 \text{ m}^2 \text{ s}^{-2}$ to illustrate the difference between purely rotational flow and purely divergent flow respectively.

2.3 Likelihood

As we have seen earlier, once we have computed K_{uv} , then the likelihood is straightforward to compute using (4). If we wish to sample from the posterior with a general prior, then we will have to use some form of Markov Chain Monte Carlo (MCMC) sampling procedure (Gilks *et al.*, 1996) since the posterior will generally not have a tractable analytical form. The most sensible method to sample from the posterior seems to be Hybrid Monte Carlo (HMC) sampling (Duane *et al.*, 1987), which is already implemented in NETLAB⁶ and samples preferentially from regions where the posterior has greater mass and includes a momentum term that should minimise random walks and allow us to efficiently explore the posterior. Unfortunately to use HMC we need the derivatives of the likelihood with respect to all the parameters. Note that:

$$P(D_{uv} | \xi) = P(D_{uv} | E_\Psi^2, E_\Phi^2, L_\Psi, L_\Phi, \nu_\Psi, \nu_\Phi, \eta^2) \quad (15)$$

There is no general reason why the length scales and smoothness for the stream function and velocity potential should be different, however initially both will be allowed to vary independently and thus the data can ‘speak for itself’. In addition to sampling from the posterior, we may well wish to compute the maximum likelihood estimates or maximum *a posteriori* probability estimates of the parameters, ξ , and to do this we shall also require derivative information.

⁵Also referred to as the nugget variance or jitter.

⁶The NETLAB tool-box for MATLAB is available from <http://www.ncrg.aston.ac.uk/netlab/index.html>.

2.4 Derivative of the Likelihood

In the general setting, for any one parameter ξ the derivative of the log likelihood can be written (Neal, 1997):

$$\frac{\partial \log (P(D_{uv} | \xi))}{\partial \xi} = -\frac{1}{2} \text{tr} \left(K_{uv}^{-1} \frac{\partial K_{uv}}{\partial \xi} \right) + \frac{1}{2} D'_{uv} K_{uv}^{-1} \frac{\partial K_{uv}}{\partial \xi} K_{uv}^{-1} D_{uv} \quad (16)$$

where everything is defined as before and tr means trace. Unfortunately this derivative is not simple to compute since each parameter has a different derivative and thus must be computed separately.

$$\begin{array}{lll} \frac{\partial K_{uv}}{\partial E_{\Psi}^2} & \frac{\partial K_{uv}}{\partial E_{\Phi}^2} & \text{see later} \\ \frac{\partial K_{uv}}{\partial L_{\Psi}} & \frac{\partial K_{uv}}{\partial L_{\Phi}} & \text{see later} \\ \frac{\partial K_{uv}}{\partial \nu_{\Psi}} & \frac{\partial K_{uv}}{\partial \nu_{\Phi}} & \text{no closed form} \\ & \frac{\partial K_{uv}}{\partial \eta^2} & = \text{I} \end{array}$$

Worse still for the smoothness parameters there is no closed form for the derivatives. The next few sections deal with computation of the derivatives not specified above.

2.4.1 Variance component derivatives

The variance parameters appear in (9) and the derivatives are easy to compute:

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

and the derivative of the full covariance matrix is formed using the equations as before using (10) to give:

$$\frac{\partial K_{uv}}{\partial E_{\Psi}^2} = \begin{pmatrix} \frac{\partial C_{uu}}{\partial E_{\Psi}^2} & \frac{\partial C_{uv}}{\partial E_{\Psi}^2} \\ \frac{\partial C_{vu}}{\partial E_{\Psi}^2} & \frac{\partial C_{vv}}{\partial E_{\Psi}^2} \end{pmatrix} \quad \frac{\partial K_{uv}}{\partial E_{\Phi}^2} = \begin{pmatrix} \frac{\partial C_{uu}}{\partial E_{\Phi}^2} & \frac{\partial C_{uv}}{\partial E_{\Phi}^2} \\ \frac{\partial C_{vu}}{\partial E_{\Phi}^2} & \frac{\partial C_{vv}}{\partial E_{\Phi}^2} \end{pmatrix}$$

2.4.2 Length scale derivatives

The length scale L appears twice in the computations of the wind covariances, in equation (9) indirectly and explicitly in (11). In (9):

$$\begin{aligned} \frac{\partial C_{ll}(r)}{\partial L_{\Phi}} &= -E_{\Phi}^2 \frac{\partial L_{e\Phi}^2}{\partial L_{\Phi}} \frac{\partial^2}{\partial r^2} \rho_{\Phi\Phi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{\partial}{\partial L_{\Phi}} \frac{\partial^2}{\partial r^2} \rho_{\Phi\Phi} \\ \frac{\partial C_{tt}(r)}{\partial L_{\Phi}} &= -E_{\Phi}^2 \frac{\partial L_{e\Phi}^2}{\partial L_{\Phi}} \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Phi\Phi} - E_{\Phi}^2 L_{e\Phi}^2 \frac{\partial}{\partial L_{\Phi}} \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Phi\Phi} \\ \frac{\partial C_{ll}(r)}{\partial L_{\Psi}} &= -E_{\Psi}^2 \frac{\partial L_{e\Psi}^2}{\partial L_{\Psi}} \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Psi\Psi} - E_{\Psi}^2 L_{e\Psi}^2 \frac{\partial}{\partial L_{\Psi}} \frac{1}{r} \frac{\partial}{\partial r} \rho_{\Psi\Psi} \\ \frac{\partial C_{tt}(r)}{\partial L_{\Psi}} &= -E_{\Psi}^2 \frac{\partial L_{e\Psi}^2}{\partial L_{\Psi}} \frac{\partial^2}{\partial r^2} \rho_{\Psi\Psi} - E_{\Psi}^2 L_{e\Psi}^2 \frac{\partial}{\partial L_{\Psi}} \frac{\partial^2}{\partial r^2} \rho_{\Psi\Psi}. \end{aligned}$$

Now:

$$\frac{\partial \cdot}{\partial L_\Phi} = \frac{\partial \cdot}{\partial z} \frac{\partial z}{\partial L_\Phi}, \quad z = \frac{r}{L_\Phi}, \quad \frac{\partial z}{\partial L_\Phi} = -\frac{r}{L_\Phi^2}$$

so that:

$$\begin{aligned} \frac{\partial}{\partial L} \frac{1}{r} \frac{\partial \rho}{\partial r} &= \frac{1}{\Gamma(\nu) 2^{\nu-1} L^3} \left(2 \left(\frac{r}{L} \right)^{\nu-1} K_{\nu-1} \left(\frac{r}{L} \right) - \left(\frac{r}{L} \right)^\nu K_{\nu-2} \left(\frac{r}{L} \right) \right) \\ \frac{\partial}{\partial L} \frac{\partial^2 \rho}{\partial r^2} &= \frac{1}{\Gamma(\nu) 2^{\nu-1} L^3} \left(\left(2 + \left(\frac{r}{L} \right)^2 \right) \left(\frac{r}{L} \right)^{\nu-1} K_{\nu-1} \left(\frac{r}{L} \right) - (2\nu + 1) \left(\frac{r}{L} \right)^\nu K_{\nu-2} \left(\frac{r}{L} \right) \right) \end{aligned}$$

where we used the identity (Abramowitz and Stegun, 1972):

$$K_{\nu-1}(z) - K_{\nu+1}(z) = -\frac{2\nu}{z} K_\nu(z)$$

Also:

$$\frac{dL_{e\Phi}^2}{dL_\Phi} = 4(\nu_\Phi - 1) L_\Phi$$

Care must be again taken as $r \rightarrow 0$ when:

$$\begin{aligned} \lim_{r \rightarrow 0} \left(\frac{\partial}{\partial L} \left(\frac{1}{r} \frac{\partial \rho}{\partial r} \right) \right) &= \frac{1}{(\nu - 1) L^3} \\ \lim_{r \rightarrow 0} \left(\frac{\partial}{\partial L} \left(\frac{\partial^2 \rho}{\partial r^2} \right) \right) &= \frac{1}{(\nu - 1) L^3} \end{aligned}$$

Thus we can explicitly compute the derivatives with respect to the length parameters computing only two modified Bessel functions (which is a good thing because (in MATLAB) the evaluation of the modified Bessel functions takes a large proportion of the time to compute the likelihood).

2.4.3 Smoothness parameter derivatives

There is no closed form derivative of the likelihood with respect to ν . Thus we use a first order centred finite difference approximation to the derivative which requires Σ_{uv} to be computed for two very similar values of ν : $\nu - \Delta\nu$ and $\nu + \Delta\nu$. Thus:

$$\frac{\partial K_{uv}}{\partial \nu} \approx \frac{K_{uv}(\nu + \Delta\nu) - K_{uv}(\nu - \Delta\nu)}{2\Delta\nu}$$

which has approximation error of the order $(\Delta\nu)^2$, thus so long as $\Delta\nu$ is small enough (but not so small as to cause round off errors) the approximation will be reasonable. We take $\Delta\nu = 10^{-4}$. These derivatives can be efficiently evaluated simultaneously with those for the length scale which reduces the number of modified Bessel function evaluations.

The derivative of the likelihood can then be evaluated using (16) for each parameter in turn to give a vector of the derivatives with respect to each of the parameters. Note that efficiency could be further improved by calculating K_{uv} at the same time as its derivatives.

3 Likelihood for Several Wind fields

In the first instance we are trying to establish sensible values for the parameters in the covariance function by sampling from their posterior distribution with uninformative priors. Thus we will be interested in sampling from a posterior that is representative of the broad range of weather conditions that can generate wind fields. In this work we will focus on the north Atlantic region

between 40° and 60° N. Since there is likely to be strong temporal variability in the covariance parameters, we will derive the posterior for each month⁷. Initially we will use only one year of European Centre for Medium range Weather Forecasting 10 *m* wind analyses. These have a grid spacing of approximately 250 *km* with 18×9 observations every 6 hours. The evolution of the atmosphere in the north Atlantic is rather fast and thus we regard each times wind field as being independent of all other wind fields.⁸

The negative log likelihood for several datasets D_1, D_2, \dots, D_N can be written:

$$-\log \left(\prod_{i=1}^N p(D_i | \boldsymbol{\xi}) \right) = -\sum_{i=1}^N \log p(D_i | \boldsymbol{\xi}) \quad (17)$$

if the datasets are independent (as they are here). If we consider each data set to have n observations and we have N datasets, then if the observations are made at the same locations in every dataset we have:

$$\begin{aligned} \sum_{i=1}^N \log p(D_i | \boldsymbol{\xi}) &= \sum_{i=1}^N \left(-\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det(K_{uv})) - \frac{1}{2} D_i' K_{uv}^{-1} D_i \right) \\ &= -\frac{nN}{2} \log(2\pi) - \frac{N}{2} \log(\det(K_{uv})) - \frac{1}{2} \text{tr}(D' K_{uv}^{-1} D) \end{aligned}$$

where D is the data matrix made up of the columns D_i and tr stands for the trace as usual. For the derivatives of the likelihood we obtain:

$$\begin{aligned} \frac{\partial P(D | \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} &= \sum_{i=1}^N \left(-\frac{1}{2} \text{tr} \left(K_{uv}^{-1} \frac{\partial K_{uv}}{\partial \boldsymbol{\xi}} \right) + \frac{1}{2} D_i' K_{uv}^{-1} \frac{\partial K_{uv}}{\partial \boldsymbol{\xi}} K_{uv}^{-1} D_i \right) \\ &= -\frac{N}{2} \text{tr} \left(K_{uv} \frac{\partial K_{uv}}{\partial \boldsymbol{\xi}} \right) + \frac{1}{2} \text{tr} \left(D' K_{uv}^{-1} \frac{\partial K_{uv}}{\partial \boldsymbol{\xi}} K_{uv}^{-1} D \right) \end{aligned}$$

Thus we can quickly compute the likelihood over a series of wind fields, and this can then be used to find the maximum likelihood solution or sample from the posterior (for a given month).

4 Conclusions

This technical report has shown exactly how derive the necessary components to construct, optimise and sample from a Gaussian process model using a modified Bessel covariance function. This has been implemented in MATLAB (software is available from the author on request). It is also shown how the modified Bessel covariance function fits in with the larger NEUROSAT project goals.

5 Acknowledgements

I would like to thank Ian Nabney and Chris Williams for their help and advice during the writing of this report. The work was carried out as part of the NEUROSAT project.

References

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

⁷ It may be sufficient to have seasonal parameters, indeed future work could consider a space-time model.

⁸ This approximation would be considerably improved taking wind fields every 24 hours; however this could cause problems because these wind-fields would be for a fixed time (e.g. 12 Greenwich Mean Time) and we wish to capture the full range of behaviour including the weak diurnal signal.

- Abramowitz, M. and I. A. Stegun (Eds.) 1972. *Handbook of Mathematical Functions*. New York: Dover Publications Inc.
- Cornford, D. 1997. Random Field Models and Priors on Wind. Technical Report NCRG/97/023, 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.
- Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth 1987. Hybrid Monte Carlo. *Physics Letters B* **195**, 216–222.
- Gilks, W., S. Richardson, and D. J. Spiegelhalter 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Neal, R. M. 1997. Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification. Technical Report 9702, Department of Statistics, University of Toronto.