NEURAL COMPUTING
NCRG
RESEARCH GROUP

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/

Random Field Models and Priors on Wind

Dan Cornford
d.cornford@aston.ac.uk

Technical Report NCRG/97/023 March 13, 1998

1 Introduction

This report seeks to make concrete some of the ideas we have been discussing about sensible priors for winds over the ocean. In particular, random field models are reviewed, as are permissible covariance functions. The criteria which these covariance functions must satisfy in order that vorticity and divergence exist and are continuous are defined. The impact of Helmholtz theorem is discussed, and sensible choices for the covariances are suggested.

2 Random Field Models

2.1 Random variables

If we recall basic probability theory, a probability space (Ω, F, P) where Ω is the sample space (i.e. all possible elementary events), F is a σ-field of subsets of Ω (i.e. a collection of subsets of Ω that also contains the empty set and Ω itself, which is closed under finite intersection, countable union

and set complement) and P is a probability measure on (Ω, F) must satisfy Kolmogorov's axioms:

$$\begin{aligned} P(\Omega) &= 1 \\ 0 \leq P(A_i) \leq 1 \qquad & \forall A_i \in F \\ \text{if} A_i \cap A_j = \emptyset \qquad (\forall i \neq j) \Longrightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

where the A_i's are called events. This probability space is the underlying model for all of the forthcoming discussion.

We can use this definition of probability to define a random variable (Christakos, 1992). Let $z(u)$ be a random variable where $u \in \Omega$ are the elementary events. Thus z is a (measurable) mapping from (Ω, F) into $(\mathbb{R}^1, F^1)$ such that:

$$\forall B \in F^1, \qquad z^{-1}(B) = \{u \in \Omega : z(u) \in B\} \in F$$

This measureability induces a probability measure μ_z on $(\mathbb{R}^1, F^1)$, such as:

$$\forall B \in F^1, \qquad \mu_z(B) = P(z^{-1}(B)) = P(z = z(u) \in B)$$

We usually study the distribution of this probability measure through the (cumulative) distribution function:

$$F_z(\chi) = P(z \leq \chi)$$

where $\chi \in \mathbb{R}^1$ or the probability density function:

$$f_z(\chi) = P(z = \chi)$$

when F_z is absolutely continuous. Using these definitions we can go on to define a random field.

2.2 Random fields

A random field is a collection of random variables defined by their *joint* distribution functions. Now define $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^2$ to be a (spatial) random field[1], then we have:

$$F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(\chi_1, \ldots, \chi_n) = P(z_1 \leq \chi_1, \ldots, z_n \leq \chi_n) \tag{1}$$

where $z_1 = Z(\boldsymbol{x}_1)$ and n can be any integer. We can also define the probability density function:

$$\begin{aligned} f_Z(\chi_1, \ldots, \chi_n) &= f_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(\chi_1, \ldots, \chi_n) \\ &= \frac{\partial^n}{\partial\chi_1 \cdots \partial\chi_n} F_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n}(\chi_1, \ldots, \chi_n) \end{aligned}$$

A random field can also be characterised in terms of its characteristic function, which is the n-dimensional Fourier transform of the probability density (Christakos, 1992).

In a very straightforward manner we can extend the concept of a scalar random field (as outlined above) to a vector random field $\boldsymbol{Z}(\boldsymbol{x})$. This will be necessary when we consider winds.

Classically, random fields are characterised by their first and second order moments. These are the mean:

$$m_z(\boldsymbol{x}) = E[Z(\boldsymbol{x})] = \int \chi f_{\boldsymbol{x}}(\chi) d\chi \tag{2}$$

[1] In this report we will consider 2 dimensional random fields since these are most relevant to our problem.

and covariance:

$$\begin{aligned} c_z(\boldsymbol{x_i}, \boldsymbol{x_j}) &= E[(Z(\boldsymbol{x_i}) - m_z(\boldsymbol{x_i}))(Z(\boldsymbol{x_j}) - m_z(\boldsymbol{x_j}))] \\ &= E[Z(\boldsymbol{x_i})Z(\boldsymbol{x_j})] - m_z(\boldsymbol{x_i})m_z(\boldsymbol{x_j}) \end{aligned} \tag{3}$$

It is also useful to define the (spatial) correlation function:

$$\rho_z(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{c_z(\boldsymbol{x_i}, \boldsymbol{x_j})}{\sqrt{\sigma_z^2(\boldsymbol{x_i})}\sqrt{\sigma_z^2(\boldsymbol{x_j})}} \tag{4}$$

where $\sigma_z^2(\boldsymbol{x_i})$ is the variance of the random field at the point $\boldsymbol{x_i}$. When the random field is Gaussian, then these first two moments completely characterise its behaviour (i.e. Gaussian processes[2]). Making the 'Gaussian assumption' is very common in practice, since inference and estimation are then straightforward.

To be a valid covariance, the function $c_z(\boldsymbol{x_i}, \boldsymbol{x_j})$ must be non-negative definite[3], that is:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} q_i q_j c_z(\boldsymbol{x_i}, \boldsymbol{x_j}) \geq 0 \tag{5}$$

$\forall n \in \mathbb{N}$, points $\boldsymbol{x_i}, \boldsymbol{x_j}$ and $q_1, \ldots, q_n \in \mathbb{R}$. Every permissible covariance is a symmetric function and if the random field is ergodic:

$$\lim_{|\boldsymbol{x_i} - \boldsymbol{x_j}| \to \infty} c_z(\boldsymbol{x_i}, \boldsymbol{x_j}) = 0$$

If we move to vector random fields then if we consider a two component (e.g. u, v) random field, with components $Z_1(\boldsymbol{x})$ and $Z_2(\boldsymbol{x})$ then in addition to the standard means and covariances we can define the cross-covariance:

$$c_{z_1,z_2}(\boldsymbol{x_i}, \boldsymbol{x_j}) = E[(Z_1(\boldsymbol{x_i}) - m_{z_1}(\boldsymbol{x_i}))(Z_2(\boldsymbol{x_j}) - m_{z_2}(\boldsymbol{x_j}))]$$

and a cross-correlation as before. We now can define the joint covariance matrix:

$$\boldsymbol{C_Z} = \begin{bmatrix} c_{z_1}(\boldsymbol{x_i}, \boldsymbol{x_j}) & c_{z_1,z_2}(\boldsymbol{x_i}, \boldsymbol{x_j}) \\ c_{z_2,z_1}(\boldsymbol{x_j}, \boldsymbol{x_i}) & c_{z_2}(\boldsymbol{x_i}, \boldsymbol{x_j}) \end{bmatrix}$$

The condition corresponding to Equation 5 for this vector case is:

$$\boldsymbol{q'C_Z q} \geq 0$$

where $q' = [q_1, \ldots, q_{2n}]$, the augmented vector, and the other conditions are as before.

2.3 Convergence properties

To study the properties of random fields we need to consider the concept of stochastic convergence[4]. There are several types of stochastic convergence. If $\{z_n\}$ is a sequence of random variables (Christakos, 1992) $\{z_n\}$ is said to converge to z (also a random variable) with

[2] In the geophysical sciences a process is generally regarded as being one dimensional and usually refers to time-series modelling.

[3] This ensures that the variance is non-negative (as it must be).

[4] This is rather like the deterministic case where we study the continuity, differentiability etc. of functions through the convergence of their series.

- mean square convergence:

$$z_n \xrightarrow{m.s.} z \iff \lim_{n\to\infty} E[|z_n - z|^2] = 0$$

- almost sure convergence:

$$z_n \xrightarrow{a.s.} z \iff P[\lim_{n\to\infty} z_n = z] = 1$$

- convergence in probability:

$$z_n \xrightarrow{p} z \iff \forall \epsilon > 0, \qquad \lim_{n\to\infty} P[|z_n - z| > \epsilon] = 0$$

Mean square and almost sure convergence both imply convergence in probability. If $z_n \xrightarrow{m.s.} z$ then $E[z_n] \to E[z]$ and the operators lim and $E[\cdot]$ commute. Thus a random field $Z(\boldsymbol{x})$ is said to converge to $Z(\boldsymbol{x_0})$ when $\boldsymbol{x} \to \boldsymbol{x_0}$ in one of the above categories if the sequence of random variables $\{z_n = Z(\boldsymbol{x})\}$ at $\boldsymbol{x} = \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}, \ldots$ tends to $z_0 = Z(\boldsymbol{x_0})$ as $n \to \infty$. We can now state the mean square convergence criterion for a random field:

Let $Z(\boldsymbol{x})$ be a random field with $\boldsymbol{x_0}$ a fixed point in two dimensional space. Then:

$$Z(\boldsymbol{x}) \xrightarrow{m.s.} Z(\boldsymbol{x_0}) \iff E[Z(\boldsymbol{x_i})Z(\boldsymbol{x_j})] \to E[Z(\boldsymbol{x_0})]^2 \qquad \boldsymbol{x_i}, \boldsymbol{x_j} \to \boldsymbol{x_0} \tag{6}$$

where the convergence of the expectations is in the usual sense (i.e. not mean square). We can now define continuity, differentiability (and integrability if needed) in terms of mean square (m.s.) or almost sure (a.s.) convergence of the random field. Note that in general m.s. convergence does not imply a.s. convergence nor vice-versa.

2.4 Continuity

A (second-order)[5] random field $Z(\boldsymbol{x})$ is m.s. continuous at $\boldsymbol{x} \in \mathbb{R}^2$ if:

$$Z(\boldsymbol{x_i}) \xrightarrow{m.s.} Z(\boldsymbol{x}), \qquad \boldsymbol{x_i} \to \boldsymbol{x}$$

or

$$Z(\boldsymbol{x} + \boldsymbol{h}) \xrightarrow{m.s.} Z(\boldsymbol{x}), \qquad \boldsymbol{h} \to 0 \tag{7}$$

This definition can be extended to the whole domain to yield an everywhere m.s. continuous random field. Now $Z(\boldsymbol{x})$ is m.s. continuous at $\boldsymbol{x} \in \mathbb{R}^2$ if and only if $c_z(\boldsymbol{x}, \boldsymbol{x_i})$ is continuous at $(\boldsymbol{x}, \boldsymbol{x_i} = \boldsymbol{x})$ (assuming that the mean is continuous - indeed here without loss of generality we will assume a zero mean).

A random field $Z(\boldsymbol{x})$ is a.s. continuous at $\boldsymbol{x} \in \mathbb{R}^2$ if:

$$Z(\boldsymbol{x} + \boldsymbol{h}) \xrightarrow{a.s.} Z(\boldsymbol{x}), \qquad \boldsymbol{h} \to 0 \tag{8}$$

If $Z(\boldsymbol{x})$ is everywhere a.s. continuous it is called sample function continuous. Sample function continuity can be expressed in terms of the covariance function of a random field. If $\boldsymbol{x}, \boldsymbol{h} \in C \subset \mathbb{R}^2$ then:

$$c_z(\boldsymbol{x} + \boldsymbol{h}, \boldsymbol{x} + \boldsymbol{h}) - c_z(\boldsymbol{x} + \boldsymbol{h}, \boldsymbol{x}) - c_z(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{h}) + c_z(\boldsymbol{x}, \boldsymbol{x}) \leq \frac{\alpha \|\boldsymbol{h}\|^4}{|\log \|\boldsymbol{h}\||^{1+\beta}} \tag{9}$$

[5] this refers to the order of the norm of the probability space - which is almost always 2 - i.e. Euclidean

for $\alpha > 0, \beta > 2$ implies that $Z(\boldsymbol{x})$ is a.s. continuous (Christakos, 1992; Adler, 1981). It should be noted that this is a very tight bound which is very difficult to attain (none of the practically used covariances satisfy this bound). If we make a 'Gaussian' assumption on the random field, and assume it to be stationary (see later), then we can significantly relax the bound to give conditions for sample function continuity (Abrahamsen, 1997; Adler, 1981):

$$c_z(\mathbf{0}) - c_z(\boldsymbol{h}) \leq \frac{\alpha}{|\log \|\boldsymbol{h}\||^{1+\epsilon}}, \qquad \alpha > 0, \epsilon > 0 \tag{10}$$

This much weaker bound is met by almost all used covariance functions, thus if a *Gaussian* random field possesses continuous expectations and a continuous covariance function then it has continuous sample paths with probability one (Abrahamsen, 1997).

Sample function continuity tells us about the behaviour of the realisations (simulations) of a random field. M.s. continuity does not necessarily imply a.s. continuity although in the Gaussian case it is an 'almost' sufficient condition (Abrahamsen, 1997).

2.5 Differentiability

A random field $Z(\boldsymbol{x})$ is m.s. differentiable at $\boldsymbol{x} \in \mathbb{R}^2$ with respect to the x component of $\boldsymbol{x} = (x, y)$ if $\exists Z_{(x)}(\boldsymbol{x})$ such that:

$$\frac{Z(\boldsymbol{x} + (h_x, 0)) - Z(\boldsymbol{x})}{h_x} \xrightarrow{m.s.} Z_{(x)}(\boldsymbol{x}), \qquad h_x \to 0$$

where $Z_{(x)}(\boldsymbol{x})$ is the partial derivative with respect to x:

$$\frac{\partial Z(\boldsymbol{x})}{\partial x} = \lim_{h_x \to 0} \frac{Z(\boldsymbol{x} + h_x) - Z(\boldsymbol{x})}{h_x}$$

This can be written in terms of the covariance function if the mean value $E[Z(\boldsymbol{x})]$ is differentiable and the covariance:

$$\begin{aligned} \operatorname{cov}\left(\frac{\partial Z(\boldsymbol{x_i})}{\partial x_i}, \frac{\partial Z(\boldsymbol{x_j})}{\partial x_j}\right) &= \frac{\partial^2 c_z(\boldsymbol{x_i}, \boldsymbol{x_j})}{\partial x_i \partial x_j} \\ &= \lim_{h_{x_i}, h_{x_j} \to 0} \frac{1}{h_{x_i} h_{x_j}} [c_z(\boldsymbol{x} + h_{x_i}, \boldsymbol{x} + h_{x_j}) \\ &\quad - c_z(\boldsymbol{x} + h_{x_i}, \boldsymbol{x}) - c_z(\boldsymbol{x}, \boldsymbol{x} + h_{x_j}) + c_z(\boldsymbol{x}, \boldsymbol{x})] \end{aligned} \tag{11}$$

exists and is finite at all diagonal points $\boldsymbol{x_i} = \boldsymbol{x_j}$. We can extend this definition of differentiability to higher order derivatives viz:

$$\operatorname{cov}\left(\frac{\partial^v Z(\boldsymbol{x_i})}{\partial x_i^v}, \frac{\partial^v Z(\boldsymbol{x_j})}{\partial x_j^v}\right) = \frac{\partial^{2v} c_z(\boldsymbol{x_i}, \boldsymbol{x_j})}{\partial x_i^v \partial x_j^v}$$

which must exist and be finite at $\boldsymbol{x_i} = \boldsymbol{x_j}$.

For sample function differentiability we have:

$$\frac{Z(\boldsymbol{x} + (h_x, 0)) - Z(\boldsymbol{x})}{h_x} \xrightarrow{a.s.} Z_{(x)}(\boldsymbol{x}), \qquad h_x \to 0$$

so long as $Z_{(x)}(\boldsymbol{x})$ exists. If we assume that $Z_{(x)}(\boldsymbol{x})$ is the m.s. partial derivative derivative of $Z(\boldsymbol{x})$ which has covariance:

$$c_{z_{(x_k, x_l)}}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{\partial^2 c_z(\boldsymbol{x_i}, \boldsymbol{x_j})}{\partial x_k \partial x_l} \tag{12}$$

then using (9) we can see that $Z_{(x)}(\boldsymbol{x})$ is a.s. continuous if:

$$c_{z_{(x_k,x_l)}}(\boldsymbol{x}+\boldsymbol{h},\boldsymbol{x}+\boldsymbol{h}) - c_{z_{(x_k,x_l)}}(\boldsymbol{x}+\boldsymbol{h},\boldsymbol{x}) \\ -c_{z_{(x_k,x_l)}}(\boldsymbol{x},\boldsymbol{x}+\boldsymbol{h}) + c_{z_{(x_k,x_l)}}(\boldsymbol{x},\boldsymbol{x}) \leq \frac{\alpha\|\boldsymbol{h}\|^4}{|\log\|\boldsymbol{h}\||^{1+\beta}} \tag{13}$$

This rather odd notation is simply saying that the derivative of the original random field is a.s. continuous if the random field defined by the second derivative of its covariance function is a.s. continuous as defined in (9). If we make the 'Gaussian' assumption then we can relax the bound to be that in (10). For criteria concerning the integrability of random fields see Abrahamsen (1997) or Christakos (1992).

2.6 Spectral representation

All random fields can be expressed in a Fourier-Stieltjes representation:

$$Z(\boldsymbol{x}) = \int_{\mathbb{R}^2} \exp(i\boldsymbol{w}\cdot\boldsymbol{x})dN_z(\boldsymbol{w})$$

where $dN_z(\boldsymbol{w})$ is a random field over $\boldsymbol{w}\in\mathbb{R}^2$. $Z(\boldsymbol{x})$ has a covariance given by:

$$c_z(\boldsymbol{x_i},\boldsymbol{x_j}) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp(i(\boldsymbol{w_i}\cdot\boldsymbol{x_i} - \boldsymbol{w_j}\cdot\boldsymbol{x_j}))dQ_z(\boldsymbol{w_i},\boldsymbol{w_j}) \tag{14}$$

where $Q_z(\boldsymbol{w_i},\boldsymbol{w_j})$ is the spectral distribution function of $Z(\boldsymbol{x})$ - and is not necessarily differentiable.

If $dN_z(\boldsymbol{w})$ is differentiable then we obtain the Riemann integral representation:

$$Z(\boldsymbol{x}) = \int_{\mathbb{R}^2} \exp(i\boldsymbol{w}\cdot\boldsymbol{x})\tilde{Z}(\boldsymbol{w})d\boldsymbol{w} \tag{15}$$

Thus $Z(\boldsymbol{x})$ and $\tilde{Z}(\boldsymbol{w})$ are Fourier pairs with:

$$\tilde{Z}(\boldsymbol{w}) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \exp(-i\boldsymbol{w}\cdot\boldsymbol{x})Z(\boldsymbol{x})d\boldsymbol{x} \tag{16}$$

The covariance function of $Z(\boldsymbol{x})$ can be written as:

$$c_z(\boldsymbol{x_i},\boldsymbol{x_j}) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp(i(\boldsymbol{w_i}\cdot\boldsymbol{x_i} - \boldsymbol{w_j}\cdot\boldsymbol{x_j}))C_z(\boldsymbol{w_i},\boldsymbol{w_j})d\boldsymbol{w_i}d\boldsymbol{w_j} \tag{17}$$

where $C_z(\boldsymbol{w_i},\boldsymbol{w_j})$ is the spectral density function of $Z(\boldsymbol{x})$ - and corresponds to the derivative of the spectral distribution function. A covariance function that admits the representation (17) is called harmonizable.

The spectral representation of random fields is important because it offers another method of checking for and generating (conditionally) positive definite covariance functions. Note that in all the above integrals when c_z is real then the exponentials can be replaced by cosines.

3 Using Random Field Models

We have defined some of the mathematics for random field models in the preceding section, however without further assumptions they are not practically applicable. This is because the full knowledge of the first and second moments (if we are assuming Gaussian processes) implies a complete knowledge about the variable of interest in any case. In reality we only have a finite sample, often from only one realisation, and thus to make progress we must make further assumptions.

3.1 Homogeneity and stationarity

A random field[6] is called homogeneous *in the wide sense* or *second-order stationary*[7] if

$$m_z(\boldsymbol{x}) = E[Z(\boldsymbol{x})] = m \tag{18}$$

and

$$c_z(\boldsymbol{x_i}, \boldsymbol{x_j}) = c_z(\boldsymbol{x_i} - \boldsymbol{x_j}) = c_z(\boldsymbol{h}) \tag{19}$$

where m is the constant mean value, and the covariance depends only on the separation vector $\boldsymbol{h} = \boldsymbol{x_i} - \boldsymbol{x_j}$. In practice this assumption allows us to make inferences from only one realisation of a random field, but care must be taken since if the field is not actually second-order stationary then the model may produce unrealistic results.

In geostatistics another form of stationarity is often invoked. *Intrinsic stationarity* is defined by:

$$E[Z(\boldsymbol{x} + \boldsymbol{h}) - Z(\boldsymbol{x})] = m(\boldsymbol{h}) = 0 \tag{20}$$

and

$$var[Z(\boldsymbol{x} + \boldsymbol{h}) - Z(\boldsymbol{x})] = E[(Z(\boldsymbol{x} + \boldsymbol{h}) - Z(\boldsymbol{x}))^2] = 2\gamma_z(\boldsymbol{h}) \tag{21}$$

where $\gamma_z(\boldsymbol{h})$ is the so called *variogram*[8] of the random field $Z(\boldsymbol{x})$. Intrinsic stationarity is weaker than second-order stationarity - thus second-order stationarity implies intrinsic stationarity but not vice-versa. We can see that the assumption of intrinsic stationarity is on the first order increments of the random field $Z(\boldsymbol{x}+\boldsymbol{h})-Z(\boldsymbol{x})$ rather than the actual field. Practical experience in geostatistical modelling of physical processes suggests that intrinsic stationarity is more common than second-order stationarity, and thus the variogram is frequently used in preference to a covariance function. Where a random field is second-order stationary we have:

$$\gamma_z(\boldsymbol{h}) = c_z(0) - c_z(\boldsymbol{h}) \tag{22}$$

and we can use either a variogram or covariance function. In general when we use random field models we will specify some form of parametric function (which satisfies the properties required for it to be valid) to represent the variogram or covariance. There are several permissible models, some of which are listed later.

Related to stationarity is the concept of *ergodicity* which implies that we can compute the means and covariances of a random field $Z(\boldsymbol{x})$ from only one realisation.

3.2 Isotropy

It is often observed that the random field being studied is invariant to rotation, that is the covariance or variogram depends only on the separation distance $h = \|\boldsymbol{x_i} - \boldsymbol{x_j}\|$ where $\|\cdot\|$ denotes the standard Euclidean norm, rather than the separation vector $\boldsymbol{h}$. For wind fields this will be a rather unlikely situation since there is generally considerably more correlation along the flow than across it. Thus we will generally be dealing with anisotropic covariance functions, although the form of the anisotropy will remain simple.

It is possible to have $h = \|\boldsymbol{x_i} - \boldsymbol{x_j}\|_K$ where $\|\cdot\|_K$ is a more general norm given by:

$$h = \|\boldsymbol{x_i} - \boldsymbol{x_j}\|_K = \sqrt{(\boldsymbol{x_i} - \boldsymbol{x_j})' K (\boldsymbol{x_i} - \boldsymbol{x_j})}$$

[6] Or indeed the natural process.
[7] Also called strict stationarity if the random field is Gaussian.
[8] Strictly $\gamma_z(\boldsymbol{h})$ is the *semivariogram* but I will use the term variogram.

and K is some positive semi-definite transformation matrix. This form of norm allows for anisotropic covariance functions. It corresponds to a scaling of the axes and is know as *geometric anisotropy* in the geostatistics literature (Wackernagel, 1995). Note that the isotropic case is simply given by $K = I$ the identity matrix.

3.3 Valid covariance functions and the nugget effect

If $c_z(\boldsymbol{h})$ is a valid covariance function for a second-order stationary random field $Z(\boldsymbol{x})$ the following properties apply (Christakos, 1992)[9]:

$$c_z(\boldsymbol{h}) = c_z(-\boldsymbol{h})$$
$$c_z(\boldsymbol{h}) \leq c_z(\boldsymbol{0})$$

If $Z(\boldsymbol{x})$ is isotropic then also:

$$\lim_{|\boldsymbol{h}|\to\infty} \sqrt{|\boldsymbol{h}|}c_z(\boldsymbol{h}) = 0$$

These conditions do not imply the function is positive definite and this must be checked separately. In general a (covariance) function is positive definite if it can be represented in the form (14). If we go on to assume that the random field is real and the spectral distribution function is differentiable then:

$$c_z(\boldsymbol{h}) = \int_{\mathbb{R}^2} \cos(\boldsymbol{w}\cdot\boldsymbol{h}) C_z(\boldsymbol{w}) d\boldsymbol{w} \tag{23}$$

where $C_z(\boldsymbol{w})$ is the spectral density function (it is always assumed that these integrals exist). We can now state that $c_z(\boldsymbol{h})$ is a permissible covariance function if:

$$C_z(\boldsymbol{w}) \geq 0 \qquad \forall \boldsymbol{w} \in \mathbb{R}^2 \tag{24}$$

In general we will require some form of continuity on our random field. For wind vectors there are strong grounds for believing the large scale wind[10] will be continuous and at least once differentiable - indeed given that vorticity and divergence are also continuous and differentiable it could be argued that at a minimum the random field should be twice differentiable. This has implications for the covariance functions we can use, but we shall come back to this latter. However at smaller scales (both in space and time) there may be strong grounds for treating the wind (random) field as containing discontinuities.

This is a philosophical point but is important to modelling wind fields. How do we treat turbulence? The wind field over any non-smooth[11] can be decomposed into two components - the steady state flow (i.e. the average flow over a sufficiently long time interval) and the turbulent flow (i.e. the difference between the steady state flow and the observed flow). Typically the steady state flow can be modelled using the equations of motion, or the geostrophic approximation - that is by models based on the integration of the differential equations governing fluid motion. The same is not true for the turbulent terms.

In practice these are regarded as random and are modelled statistically (although some models now attempt to resolve the largest turbulent eddies over very small domains in a research setting

[9] assuming again that it is ergodic
[10] By which I mean wind averaged over large areas or times.
[11] That is effectively any real surface for other than trivially slow wind speeds.

- often with aim of improving the parametric models used to statistically represent turbulence). Thus we have a dilemma borne of scale[12] which is very difficult to resolve. At the very smallest observable level (time-scales of milliseconds) the wind appears continuous - if a little complicated. If we consider a larger averaging scale (say one second) then we might observe the wind field to be discontinuous - in both space and time - despite the fact that the underlying process is continuous. This is one source of the *nugget* effect often referred to in geostatistics (Wackernagel, 1995).

The nugget effect is used to model the variance associated with two sources of uncertainty. Firstly there is the error due to the representativity of the observations (in space and possibly time) - that is the error that comes from the observations being made at inconsistent scales compared to the process as discussed above. The nugget effect can also be used to model uncorrelated observing errors. The covariance is then given by:

$$c_z(\boldsymbol{h}) = \alpha\delta_z(\boldsymbol{h}) + \beta c_b(\boldsymbol{h}) \tag{25}$$

where $\delta_z(\boldsymbol{h})$ is the Kronecker delta ($\delta_z(\boldsymbol{h}) = 1 \iff \boldsymbol{h} = 0$), α is the amount of 'uncertainty' variance and $c_b(\boldsymbol{h})$ is a continuous valid covariance function. Given this model we could separate $Z(\boldsymbol{x})$ into:

$$Z(\boldsymbol{x}) = Z_\alpha(\boldsymbol{x}) + Z_b(\boldsymbol{x}) \tag{26}$$

where $Z_\alpha(\boldsymbol{x})$ is a white noise random field and $Z_b(\boldsymbol{x})$ is the 'smooth' component of $Z(\boldsymbol{x})$.

In order to obtain a good model we must try to accommodate the process scales (with reference to the observation scales). We must also recall the objectives of the modelling. In many cases we will not want to represent the turbulent component of the wind (in simple numerical models of the large scale atmospheric flow these turbulent components can amplify and propagate in alarming fashion and are thus filtered in various ways) rather we desire to predict the large scale flow. Thus we can view our problem as a smoothing or filtering one.

The data we are dealing with is scatterometer estimated wind fields which have a spatial resolution of $50 \times 50km$. Thus we can reasonably assume that at this resolution we are observing the large scale wind flow. An interesting question (which I have not seen addressed) is what happens when the wind field is inhomogeneous. There are two types - firstly and most simply - that due to the turbulent nature of wind over the ocean (just watch the ripples - which is what the satellite is 'seeing' - over a lake). How does this 'patchy' ripple pattern aggregate over the $50 \times 50km$ that the satellite observes? It is not clear that this should be linear - and is a useful justification for the use of neural networks to model the (inverse) transfer function. The second question involves what happens across fronts. Here there is a large scale (but still less than the observation scale) change in wind direction which may occur under a front. It is not known what the scatterometer will return.

So what does this mean for wind field modelling? Well I think we must insist that we are modelling at the observation scale but ensuring that we incorporate our knowledge of the process scales. This means that we will want to use a covariance function that has a nugget component (that is the noise component will cause the field to be sample function discontinuous) and a 'smooth' component that represents a random field which is mean square differentiable at least twice.

3.4 Practical differentiability

If we are considering a homogeneous (i.e. second-order stationary) random field $Z(\boldsymbol{x})$ then this is m.s. continuous if its covariance function $c_z(\boldsymbol{h})$ is continuous at $\boldsymbol{h} = 0$. It is m.s. differentiable

[12] Which is often referred to in the literature on fractals and chaos.

to the v'th order if the $2v$'th partial derivative of $c_z(\boldsymbol{h})$ exists and is finite at $\boldsymbol{h} = 0$ (Christakos, 1992). This means that

$$\frac{\partial^v Z(\boldsymbol{x_i})}{\partial x_i^v} \tag{27}$$

exists in the m.s. sense iff:

$$\left.\frac{\partial^{2v-1} c_z(\boldsymbol{h})}{\partial h_x^{2v-1}}\right|_{\boldsymbol{h}=0} = 0 \tag{28}$$

Thus we have the necessary and sufficient conditions for our covariance to be at least twice differentiable. This may be clarified by an example.

If we consider the two dimensional 'squared exponential' covariance function given by:

$$c_z(\boldsymbol{h}) = \exp\left(-\frac{\|\boldsymbol{h}\|^2}{a^2}\right)$$

with the width parameter $a > 0, \boldsymbol{h} = (h_x, h_y)$ then we can show that the $2v-1$'th derivative (with respect to x) contains only terms in odd powers (greater than 0) of h_x times the exponential term. At $\boldsymbol{h} = 0$ this expression is equal to zero for all integer $v > 0$, thus a random field with this covariance function is infinitely many times m.s. differentiable - something which we rarely observe. Indeed this case can be thought of as pathological in that it implies that the knowledge of the field over a very small area (which includes knowledge of all the derivatives up to a given order) would make it possible to predict the value of the field anywhere over the domain (c.f. Taylor expansion).

Now consider the one dimensional example:

$$c_z(h) = \cos(bh)\exp\left(-\frac{h}{a}\right)$$

with both $a, b > 0$, which is a suitable covariance function for a process with a periodic component[13]. This covariance produces a m.s. continuous process, however it is not m.s. differentiable:

$$\frac{\partial c_z(h)}{\partial h} = -\frac{1}{a}\cos(bh)\exp\left(-\frac{h}{a}\right) + b\sin(bh)\exp\left(-\frac{h}{a}\right)$$

and at $h = 0$ this is equal to $1/a$ which is non-zero.

In the above sections we have examined two particular cases. This analysis can be generalised for any covariance function. We often assume isotropy of the homogeneous random fields - this may be appropriate for the stream function and velocity potential fields. In this case we have:

$$c_z(\boldsymbol{h}) = c_z(r) \tag{29}$$

where $r = \|\boldsymbol{h}\|$. In two dimensions we are now working in polar coordinates, with only r being relevant. Thus we can express continuity conditions on the candidate covariance functions in terms of r. Julian and Thiebaux (1975) suggest that the necessary criteria for differentiability (and thus existence of a solution) are:

$$\lim_{r\to 0}\left(\frac{1}{r}\frac{\partial c_z(r)}{\partial r}\right) \tag{30}$$

is finite and:

$$\lim_{r\to 0}\left(\frac{1}{r}\frac{\partial c_z(r)}{\partial r} - \frac{\partial^2 c_z(r)}{\partial r^2}\right) = 0 \tag{31}$$

[13] which is typical of the transient waves in the atmosphere - that is the cycle of highs and lows

This is different from the criterion for m.s. differentiability of an isotropic random field given in Christakos (1992) which is:

$$\left.\frac{\partial c_z(r)}{\partial r}\right|_{r=0} = 0 \tag{32}$$

It can be seen that (31) is a zero condition on the Laplacian of the covariance function (in plane polar coordinates) and corresponds to the condition:

$$\lim_{x,y\to 0} -\left(\frac{\partial^2 c_z(x,y)}{\partial x^2} + \frac{\partial^2 c_z(x,y)}{\partial y^2}\right) = \lim_{\cdot\to 0} -\nabla^2 c_z(\cdot) = 0 \tag{33}$$

where $r = \sqrt{x^2+y^2}$. Thus we are saying that the second derivative of the covariance function (in polar coordinates) tends to zero as $r \to 0$. Adler (1981) gives the definition of mean square differentiability as being that:

$$\frac{\partial^2 c_z(\boldsymbol{s},\boldsymbol{t})}{\partial s_i \partial t_i} \tag{34}$$

exists and is finite at the point $(\boldsymbol{t},\boldsymbol{t})$ for all $\boldsymbol{t}$. This corresponds to the earlier definition in Christakos (1992) given in (11). If we assume the field is homogeneous and that the covariance is therefore a symmetric function then the requirement (32) can be seen to be a sufficient and necessary condition for (34). Thus it would seem that the conditions given in Julian and Thiebaux (1975) are a little difficult to reconcile with those above. [The different conditions expressed in Julian and Thiebaux (1975) to those elsewhere can be reconciled if we swap the conditions of the derivatives of the covariance functions in (30) and (31). I believe there may have been a typo in the original paper (Julian and Thiebaux, 1975).]

3.5 Prediction with random fields

In the next sections we deal with the practical application of random field models to prediction and simulation of two dimensional Gaussian random fields. In many instances the Gaussian assumption is not needed for the results to be true, however under this assumption many of the results are optimal. In cases where the Gaussian assumption is inappropriate other methods should be used (Diggle *et al.*, 1998; de Oliveira *et al.*, 1997). We shall generally consider scalar output variables, but the extension to vector variables is trivial, with no major modifications necessary.

Consider a random variable $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^2$ with $n \times 1$ observation vector $\boldsymbol{z_m} = z(\boldsymbol{x_m})$. If these observations are assumed to come from a (zero mean) random field model with covariance function c_z the predicted value $z_o = z(\boldsymbol{x_o})$ at some unsampled location $\boldsymbol{x_o}$ is given by:

$$\hat{z}_o = E[z_o|\boldsymbol{z_m}] = \boldsymbol{c_o}'C^{-1}\boldsymbol{z_m} \tag{35}$$

To see the meaning of the terms, consider the joint $n+1 \times n+1$ covariance matrix of $[\boldsymbol{z_m}, z_o]'$ given by:

$$\begin{bmatrix} C & \boldsymbol{c_o} \\ \boldsymbol{c_o}' & c_{oo} \end{bmatrix}$$

where C is the $n \times n$ covariance matrix of $\boldsymbol{z_m}$, $\boldsymbol{c_o}$ is the $n \times 1$ covariance vector between $\boldsymbol{z_m}$ and z_o and c_{oo} is the (co)variance of z_o. This can all be derived from the properties of the multivariate Normal distribution (in the Gaussian case) or from the minimisation of the (expected) squared error of prediction under unbiased, linear conditions. Since the distribution $P(z_o|\boldsymbol{z_m})$ is Gaussian the estimation variance is given by:

$$\hat{\sigma}^2_{z_o} = var[z_o] = c_{oo} - \boldsymbol{c_o}'C^{-1}\boldsymbol{c_o} \tag{36}$$

which does not depend on $\boldsymbol{z_m}$ but only on c_z, $\boldsymbol{x_m}$ and $\boldsymbol{x_o}$.

The vector case is similar although the matrix dimension changes and the notation becomes less clear. In the above results we have assumed that the covariance function is completely known, and that the random field is zero mean. In practice we attempt to estimate the covariance function from the data, which is the reason for all the stationarity (ergodicy) assumptions.

The next section examines how. Strictly speaking we should adopt a Bayesian approach and evaluate the prediction through the posterior distribution using (hyper)priors over the parameters in the covariance function (Neal, 1997; Brown *et al.*, 1994; Handcock and Stein, 1993).

3.6 Estimation of the covariance

There are two commonly used methods for estimating the parameters in the covariance function; (restricted) maximum likelihood and method of moments (Cressie, 1993). Maximum likelihood estimation is the more principled approach. Under the Gaussian assumption, with covariance function c_z parameterised by parameter vector $\boldsymbol{\theta}$ the log likelihood of the the data $\boldsymbol{z_m}$ is:

$$l = P(\boldsymbol{z_m}|\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|C|) - \frac{1}{2}\boldsymbol{z_m}'C^{-1}\boldsymbol{z_m} \tag{37}$$

where $|C|$ means the determinant of C. Thus we can compute the likelihood of the data for any given model, and also the derivatives (Williams, 1998; Neal, 1997) which will allow us to determine the most likely parameters to have generated the observations using gradient based minimisation methods (conditional on the model being sensible as ever).

In the case where we cannot assume zero mean we will also have to estimate the mean function. In this case we might use the (restricted) maximum likelihood estimator (Cressie, 1993) and some parametric mean (often given by a combination of 'regression' basis functions). Note that in the finite data case there is no unique partition into mean and covariance component - this depends on the investigators beliefs and the aim of the model.

Ripley (1988) notes that in practice maximum likelihood estimates of parameter values do not agree with visual inspection of the empirical covariogram given by:

$$\bar{C}(h) = \frac{1}{N(h)} \sum_{N(h)} \left((Z(\boldsymbol{x_i}) - m_z)\,(Z(\boldsymbol{x_j}) - m_z) \right) \tag{38}$$

where $N(h)$ is the number of pairs with separation distance $h_{lower} \leq h < h_{upper}$ and m_z is some estimate of the mean function - which we usually assume zero. Effectively one divides the range of h into a number of classes, usually equally spaced, and using the replications within each class one computes the mean covariance in of each class at the mean separation vector.

Plotting this allows for visual inspection of the covariance function estimator. One can take the analysis a step further and estimate the parameters of the covariance function by fitting the function to the empirical estimator. This is know as method of moments estimation, and is recommended by Cressie (1993). There are various estimators that can be used and methods for determining the parameters of the fitted model, although most use a weighted least squares estimator with gradient based optimisation - although since the parameter space is small and of low dimension exhaustive searches are quick and easy.

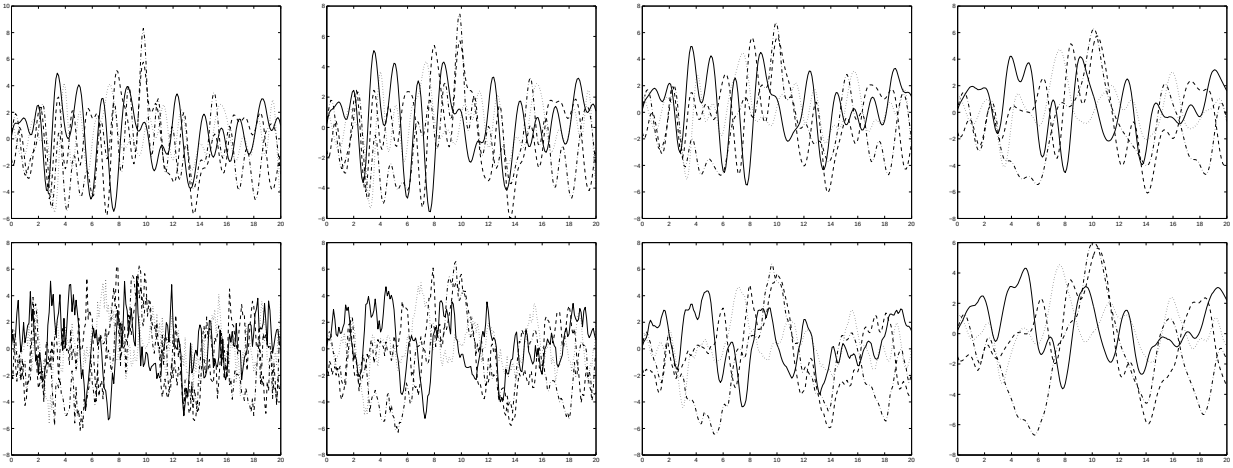

Figure 1: Four realisations from a 1D Gaussian process based on Bessel function covariance (Equation (53) - top) and modified Bessel function covariance (Equation (52) - bottom) with varying smoothness ($\nu = 0.5, 1, 2, 5$), using the same seeds. In all realisations the variance is 5, the length scale 1 and the nugget 0.0001.

3.7 Simulation with random fields

In addition to (optimal) prediction (which can be seen as smoothing) we may be interested in realisations of the process defined by the covariance function we are using. To look at these sample paths we need to know how to simulate realisations from a given random field model. Note that the aim of simulation is different from prediction. The simulation is achieved by a Cholesky decomposition LL' of C_{ss} the $n_s \times n_s$ covariance matrix of the points $\boldsymbol{x_s}$ at which the simulation is desired. One then takes a $n_s \times 1$ vector of iid Gaussian random variables $\boldsymbol{\varepsilon}$ and forms the simulated values:

$$\boldsymbol{z_s} = L\boldsymbol{\varepsilon} + \boldsymbol{m_z} \tag{39}$$

where $\boldsymbol{m_z}$ is the mean vector as usual. The main problem with this technique is that the Cholesky decomposition scales as n_s^3 with storage of n_s^2 which makes it slow to compute for $n > 1000$. Other methods such as turning-bands simulation are available and less computationally demanding (Cressie, 1993).

In addition to the above unconditional simulation, one may want to simulate conditionally on some observed data in the $n \times 1$ data vector $\boldsymbol{z_m} = z_m(\boldsymbol{x_m})$ at simulation points $\boldsymbol{x_s}$. If we denote the conditional simulation by the $n_s \times 1$ vector $\boldsymbol{z_{cs}}$ then:

$$\boldsymbol{z_{cs}}(\boldsymbol{x_s}) = \boldsymbol{z_s}(\boldsymbol{x_s}) + C'_{ms} C_{mm}^{-1} (\boldsymbol{z_m}(\boldsymbol{x_m}) - \boldsymbol{z_s}(\boldsymbol{x_m})) \tag{40}$$

where the full covariance matrix between the data and simulation points is given by:

$$\begin{bmatrix} C_{mm} & C_{ms} \\ C'_{ms} & C_{ss} \end{bmatrix}$$

with C_{mm} the $n \times n$ covariance matrix of the observed data, C_{ms} the $n \times n_s$ (cross) covariance between the observation and simulation points and C_{ss} is the $n_s \times n_s$ covariance between the simulation points. The conditional simulation honours the data, yet retains the variability associated with the random field model fitted, and has an estimation variance twice that of prediction (Cressie, 1993).

4 Implications of Helmholtz Theorem for Wind Field Modelling

We now look at the application of random field models to vector data, in particular wind fields. Helmholtz Theorem (Cornford, 1997b) is useful because it allows us to manipulate the covariance function for the vector winds into two scalar covariances for Ψ and Φ. If we consider two wind vectors (u_1, v_1) and (u_2, v_2) at locations (x_1, y_1), (x_2, y_2) respectively then the cross-covariance:

$$\begin{aligned} C_{uv}((x_1,y_1)(x_2,y_2)) &= E[u_1.v_2] \\ &= E\left[\left(\frac{\partial\Phi_1}{\partial x}-\frac{\partial\Psi_1}{\partial y}\right).\left(\frac{\partial\Phi_2}{\partial y}+\frac{\partial\Psi_2}{\partial x}\right)\right] \\ &= E\left[\left(\frac{\partial\Phi_1}{\partial x_1}\frac{\partial\Phi_2}{\partial y_2}\right)\right]+E\left[\left(\frac{\partial\Phi_1}{\partial x_1}\frac{\partial\Psi_2}{\partial x_2}\right)\right] \\ &\quad -E\left[\left(\frac{\partial\Psi_1}{\partial y_1}\frac{\partial\Phi_2}{\partial y_2}\right)\right]-E\left[\left(\frac{\partial\Psi_1}{\partial y_1}\frac{\partial\Psi_2}{\partial x_2}\right)\right] \end{aligned} \tag{41}$$

The Ψ and Φ's are continuous and differentiable, thus using the definitions of expectations and derivatives we get:

$$E\left[\left(\frac{\partial\Phi_1}{\partial x_1}\frac{\partial\Phi_2}{\partial y_2}\right)\right]=\frac{\partial^2}{\partial x_1\partial y_2}E[\Phi_1.\Phi_2]=\frac{\partial^2}{\partial x_1\partial y_2}C_{\Phi\Phi}((x_1,y_1)(x_2,y_2)) \tag{42}$$

If we assume the flow is homogeneous then the covariances no longer depend on absolute locations, merely the relative locations of the two points:

$$C_{\Phi\Phi}((x_1,y_1)(x_2,y_2))=C_{\Phi\Phi}(\boldsymbol{r}) \tag{43}$$

where $\boldsymbol{r}=(x_1-x_2, y_1-y_2)$ is the displacement vector, which is usually considered in plane polar coordinates (r,ϕ). Thus we find that:

$$C_{uv}=\frac{\partial^2}{\partial x^2}C_{\Phi\Phi}-\frac{\partial^2}{\partial y^2}C_{\Psi\Psi}+\frac{\partial^2}{\partial x\partial y}C_{\Phi\Psi}-\frac{\partial^2}{\partial y\partial x}C_{\Psi\Phi} \tag{44}$$

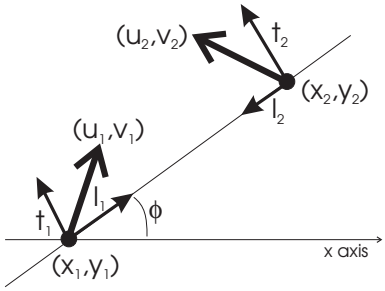

Figure 2: A graphical description of the the conversion of velocity components from (u, v) to (l, t). Note that this is with reference to vector pairs.

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 and we define longitudinal and transverse velocity components (Figure 2):

$$\begin{aligned} l &= u\cos(\phi) + v\sin(\phi) \\ t &= -u\sin(\phi) + v\cos(\phi) \end{aligned} \tag{45}$$

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

$$\begin{bmatrix} C_{ll} & C_{lt} \\ C_{tl} & C_{tt} \end{bmatrix} = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} C_{uu} & C_{uv} \\ C_{vu} & C_{vv} \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \tag{46}$$

where the $\cos(\phi)$'s etc. are square matrices of the cosine of the angle (relative to the x-axis) between each observation pair. We can now write:

$$\begin{aligned} C_{ll}(r) &= -\frac{1}{r}\frac{d}{dr}C_{\Psi\Psi} - \frac{d^2}{dr^2}C_{\Phi\Phi} \\ C_{tt}(r) &= -\frac{d^2}{dr^2}C_{\Psi\Psi} - \frac{1}{r}\frac{d}{dr}C_{\Phi\Phi} \\ C_{lt} &= C_{tl} = 0 \end{aligned} \tag{47}$$

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. This is necessary if we are to simulate realisations (sample) from this prior (covariance).

4.1 Common covariance models for wind data

This is a bit of an oxymoron since there are very few published results which use covariances of the wind fields directly. Almost all studies are concerned with the analysis increments - that is the difference between the observations and the forecast background. These increments will have very different properties from the true winds.

4.1.1 Suitable covariance functions

If we now replace the covariance functions by correlations $C_{uv}(r, \phi) = E_u^2 \rho_{uu}(r, \phi)$ for all covariances where $E_u^2 = C_{uu}(0, 0)$, then we can rewrite the equations in terms of 'standard' correlation functions. We shall start by considering correlation functions of the form:

$$\rho_{\Phi\Phi}(r; L_\Phi) = \left(1 + \frac{r}{L_\Phi}\right) \exp\left(-\frac{r}{L_\Phi}\right) \tag{48}$$

$$\rho_{\Psi\Psi}(r; L_\Psi) = \left(1 + \frac{r}{L_\Psi}\right) \exp\left(-\frac{r}{L_\Psi}\right) \tag{49}$$

where L_Φ and L_Ψ are characteristic length scales for the stream function and velocity potential. These will typically be of the order of $1000km$ with there being little evidence to suggest different length scales for the different fields. If we now define:

$$v^2 = \frac{E_\Phi^2}{L_\Phi^2 E_u^2}$$

which gives the ratio of the kinetic energy in the divergent flow (as given by Φ) to that in the total wind flow. This allows us to explicitly control the ratio of divergence and vorticity in the resulting flow fields.

It can be noted that the random fields defined by the correlation (i.e. covariance) functions given above are once differentiable since:

$$\left.\frac{\partial \rho_{\Phi\Phi}}{\partial r}\right|_{r=0} = \left(-\frac{r}{L_\Phi^2}\right) \exp\left(-\frac{r}{L_\Phi}\right)\Bigg|_{r=0} = 0 \tag{50}$$

however the random field is not twice differentiable (as it should be so that the divergence and vorticity exist and are continuous) because:

$$\left.\frac{\partial^3 \rho_{\Phi\Phi}}{\partial r^3}\right|_{r=0} = \frac{1}{L_\Phi^3}\left(2 - \frac{r}{L_\Phi}\right) \exp\left(-\frac{r}{L_\Phi}\right)\Bigg|_{r=0} \neq 0 \tag{51}$$

Thus the correlation functions of the form (49) are not very good candidates for modelling the wind field since they do not produce a continuous vorticity or divergence field.

One of the most general forms for isotropic correlation functions is:

$$\rho_{\Phi\Phi}(r; L, \nu) = \frac{1}{2^{\nu-1}?\,(\nu)} \left(\frac{r}{L'}\right)^{\nu} K_\nu \left(\frac{r}{L'}\right) \tag{52}$$

where $L' = L/(2\sqrt{\nu})$, ? is the gamma function and K_ν is a modified Bessel function of the second kind of order ν. In this model $L > 0$ is a scale parameter and $\nu > 0$ is a smoothness parameter. With $\nu = 0.5$ the function corresponds to the exponential class of covariance functions, and as $\nu \to \infty$ we obtain the 'squared exponential' covariance. A random field with this covariance function is $\lceil \nu \rceil - 1$ times m.s. differentiable and the realisations will be a similar number of times differentiable if $\nu > \lceil \nu \rceil - 0.5$ (Handcock and Stein, 1993). Thus we desire $\nu > 2$ for the vorticity and divergence to exist and be continuous. One of the problems with using covariance functions of this form is that their computation is rather time consuming, although in relation to the other computational burdens this is probably of marginal importance. Also this correlation function cannot have negative values, which is not very physically realistic - due to the wave-like nature of the atmosphere.

We can obtain negative correlations if we use:

$$\rho_{\Phi\Phi}(r; \omega, \nu) = ?\,(\nu + 1) 2^{\nu} (\omega r)^{-\nu} J_\nu(\omega r) \tag{53}$$

where $\nu \geq 0$ and $\omega > 0$. This form of correlation function permits negative correlations and is infinitely many times m.s. differentiable. This may be the most sensible choice of correlation function to use in practice, although like the squared exponential covariance function, the assumption is that the underlying random field is analytic. This means that given a large number of points in a small area (or equivalently a large number of derivatives at a point) we might know the value of the function everywhere as the number of points (derivatives) tends to infinity.

Actually, it may be that it will prove impossible to use a covariance function with the desirable differentiability properties (and which is still a valid covariance function). The periodic form outlined in (53) is likely to give good priors due to the inherent 'periodicity' of the atmosphere, but will produce very smooth wind fields. The more flexible modified Bessel based covariance function (52) may be the most suitable candidate since it allows a great deal of control over the smoothness of the generating random fields. Although it cannot produce strictly periodic behaviour is will produce quasi-periodic behaviour so long as the length scale is of the same order of magnitude as the dimensions of the sampling area (which will be true for scatterometer swathes) - see Figure 1.

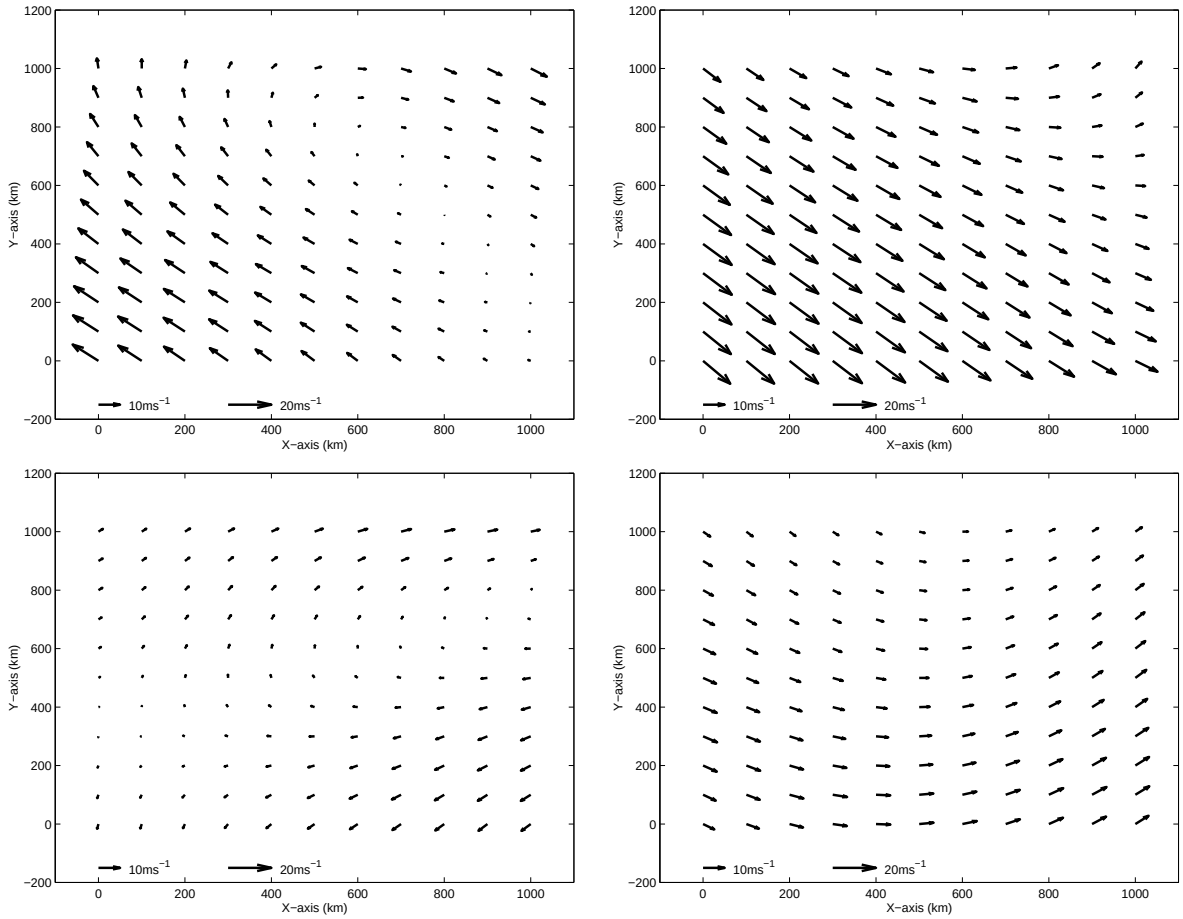

Figure 3: Four realisations from a vector random field model based on a modified Bessel function covariance (Equations (52) and (55)) with varying parameters (top - maximum likelihood parameters for January 1996, bottom - July 1996, North Atlantic). Note the lighter winds in the summer in the North Atlantic.

4.1.2 Wind correlations from velocity potential and stream function correlations

Whatever form of correlation function we choose we can write:

$$\rho_{ll}(r) = -L_\Psi^2\left(1-v^2\right)\frac{1}{r}\frac{d}{dr}\rho_{\Psi\Psi} - L_\Phi^2 v^2\frac{d^2}{dr^2}\rho_{\Phi\Phi} \tag{54}$$

$$\rho_{tt}(r) = -L_\Psi^2\left(1-v^2\right)\frac{d^2}{dr^2}\rho_{\Psi\Psi} - L_\Phi^2 v^2\frac{1}{r}\frac{d}{dr}\rho_{\Phi\Phi} \tag{55}$$

Thus we now have a general method for producing (u, v) covariances under the assumptions given above (homogeneity, isotropy - of the scalar covariances, and no correlation between Φ and Ψ) when we have a suitable form of correlation function for Φ and Ψ with given length scale (that is the scale at which the correlation first approaches zero for the field in question - which relates to the typical size of features (depressions, anticyclones) in the atmosphere). We also need to specify the kinetic energy of the wind ($\|\boldsymbol{u}\|^2/2$) which is assumed to be about $32m^2s^{-2}$. This will allow us to simulate realisations of the given covariance functions with given parameter values. We could hypothesise sensible priors for these hyper-parameters.

5 Practicalities

Recall, in practice we need to obtain some *parametric* functional form for the covariances. In wind vector assimilation into numerical weather prediction models one usually assumes some form of background field - which is typically a forecast field (but was sometimes assumed to be a climatology in the very early days of meteorological data assimilation). Thus one would actually analyse the error between the observed and background values (Daley, 1991). This is not practicable for us to do - and anyway it is probably better left to the Meteorological Office, who have readily available background fields.

We can take two choices. Effectively we have to assume a non-informative background field (i.e. no background information). While we could use a climatological background this will probably not be very useful since we are primarily interested in those events dissimilar to the climatology. Thus we can try to estimate sensible covariances for (u, v) - quite possibly through (Φ, Ψ) **or** we can use a regularisation approach (Cornford, 1997a) and let the choice of penalty dictate the covariance (strictly the basis). These approaches are similar but not necessarily parallel.

The success of both methods will depend on the choice of a sensible covariance (constraint). This is a very difficult area, because there have been no publications (which I know of!) that have discussed choices of priors appropriate to every atmospheric situation. In particular we have data at $25km$ resolution, which is much denser than other observed wind data - and thus has unknown properties. It is quite likely that errors in the data will be strongly correlated (since we are using the same instrument to observe all observations) as is true of many space based observation systems (Daley, 1991).

In the barotropic atmosphere (a strongly constrained atmosphere where the density depends only on pressure - which implies there are no thermal gradients on surfaces of constant pressure) there should be relatively few problems. The barotropic model is a good approximation for many atmospheric situations - but not all. Fronts (Cornford, 1997b) present a particular problem since in these regions one must assume the atmosphere is baroclinic - that is density depends on both pressure and temperature. Thus in the vicinity of fronts priors based on a barotropic model of the atmosphere (which most currently used 'priors' are) will be inappropriate.

5.1 A Hierarchical Approach

I think the way that we will have to tackle this problem is hierarchical. Firstly we will have priors on the probability of each type of feature (high, low, ridge, trough, **front**). These will vary in space and time. Then for each feature we have a fair prior on the distribution of vorticity and divergence, and thus we could simulate these and then conditionally simulate (u, v) on the vorticity and divergence fields. The trouble is that this is a lot of work (computationally and human wise).

There is an easier way, which is to use priors on (u, v) which already contain the features. This is true for most priors we have discussed for the larger scale features (lows, highs, troughs and ridges - see Figure 3) but is not true for **fronts**. Thus these will need special attention.

6 Conclusion

This short document forms the basis for the Bayesian approach to disambiguation. It has clarified many of the concepts surrounding random field theory. Much time has been spent examining the

implications of the choice of the covariance function on the properties of the resulting field. It has been shown that covariances on the stream function and velocity potential can give rise to sensible and realistic covariances on (u, v). Some potentially useful models for wind fields have been suggested and their merits examined. A short discussion about methods to include features such as fronts, which will be expanded elsewhere as progress is made, has highlighted their potential importance.

7 Acknowledgements

I would like to thank Chris Williams and Ian Nabney for constructive comments on the form and content of this report. The work has been carried out as part of the European Union funded NEUROSAT program.

References

Abrahamsen, P. 1997. A review of gaussian random fields and correlation functions, second edition. Technical Report 917, Norwegian Computing Center.

Adler, R. J. 1981. *The Geometry of Random Fields*. Chichester: John Wiley and Sons.

Brown, P. J., N. D. Le, and J. V. Zidek 1994. Multivariate spatial interpolation and exposure to air polutants. *The Canadian Journal of Statistics* **22**, 489–509.

Christakos, G. 1992. *Random Field Models in Earth Sciences*. London: Academic Press.

Cornford, D. 1997a. Splines and vector splines. Technical Report NCRG/97/024, 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.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*. New York: John Wiley and Sons.

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

de Oliveira, V., B. Kedem, and D. A. Short 1997. Bayesian prediction of transformed gaussian random fields. *Journal of the American Statistical Society* **92**, 1422–1433.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed 1998. Model-based geostatistics. *Applied Statistics* **47** (2).

Handcock, M. S. and M. L. Stein 1993. A bayesian analysis of kriging. *Technometrics* **35**, 403–410.

Julian, P. R. and H. J. Thiebaux 1975. On some properties of correlation functions used in optimum interpolation schemes. *Monthly Weather Report* **103**, 605–616.

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.

Ripley, B. D. 1988. *Statistical Inference for Spatial Processes*, Chapter Likelihood Analysis for Spatial Gaussian Processes, pp. 9–21. Cambridge, UK: Cambridge University Press.

Wackernagel, H. 1995. *Multivariate Geostatistics*. Berlin: Springer-Verlag.

Williams, C. K. I. 1998. Prediction with gaussian processes: From linear regression to linear prediction and beyond. In M. I. Jordan (Ed.), *Learning and Inference in Graphical Models*. Kluwer.