Inference for variograms

Adrian W. Bowman[a,*], Rosa M. Crujeiras[b]

[a] *School of Mathematics & Statistics, The University of Glasgow, U.K.*

[b] *Department of Statistics and OR, University of Santiago de Compostela, Spain*

Abstract

The empirical variogram is a standard tool in the investigation and modelling of spatial covariance. However, its properties can be difficult to identify and exploit in the context of exploring the characteristics of individual datasets. This is particularly true when seeking to move beyond description towards inferential statements about the structure of the spatial covariance which may be present. A robust form of empirical variogram based on a fourth-root transformation is used. This takes advantage of the normal approximation which gives an excellent description of the variation exhibited on this scale. Calculations of mean, variance and covariance of the binned empirical variogram then allow useful computations such as confidence intervals to be added to the underlying estimator. The comparison of variograms for different datasets provides an illustration of this. The suitability of simplifying assumptions such as isotropy and stationarity can then also be investigated through the construction of appropriate test statistics and the distributional calculations required in the associated p-values can be performed through quadratic form methods. Examples of the use of these methods in assessing the form of spatial covariance present in datasets are shown, both through hypothesis tests and in graphical form. A simulation study explores the properties of the tests while pollution data on mosses in Galicia (North-West Spain) are used to provide a real data illustration.

Keywords: isotropy, nonparametric smoothing, standard error, stationarity, variogram.

1. Introduction

The variogram is a very well established tool in the analysis of spatial data as it describes the nature of spatial covariance in a very convenient manner. Cressie (1993)

*corresponding author

Email addresses: adrian.bowman@glasgow.ac.uk (Adrian W. Bowman), rosa.crujeiras@usc.es (Rosa M. Crujeiras)

Supplementary material is available on-line

gives a broad overview of spatial methods in general, including a detailed treatment of the variogram, while Diggle and Ribeiro (2007) approach the topic in the context of regression and other forms of spatial modelling. From this latter approach, spatial covariance can be viewed as a nuisance parameter which is not of primary interest but which needs to be accommodated appropriately in a model in order to avoid compromising the estimation and interpretation of regression and other parameters. This is a very appealing perspective. However, it remains clear that there is considerable motivation for assessing the characteristics of spatial covariance in a dataset or model and the variogram offers a natural route to doing so. The abundance of papers in both the methodological and the applied literature, with Marchant and Lark (2007) and Emery and Ortiz (2007) as indicative examples, testifies to the popularity and value of the variogram.

Suppose that the random variable $Z(\mathbf{s})$ denotes the value of a process at spatial location $\mathbf{s}$ within a region $\mathcal{D}$. A simple model of spatial variation assumes that the process is Gaussian and intrinsically stationary with mean zero. Spatial covariance can then be characterised through the variogram $2\gamma(\mathbf{h}, \mathbf{s})$ as

$$2\,\gamma(\mathbf{h}, \mathbf{s}) = \mathsf{var}\{Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})\}\,.$$

The function γ is referred to as the semivariogram. When the process is weakly stationary, with mean and covariance independent of location $\mathbf{s}$, this reduces to the simpler form

$$2\,\gamma(\mathbf{h}) = \mathsf{var}\{Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})\}\,.$$

An assumption of isotropy, where covariance depends on the spatial separation vector $\mathbf{h}$ only through its size $h = ||\mathbf{h}||$, allows the further simplification

$$2\,\gamma(h) = \mathsf{var}\{Z(\mathbf{s}) - Z(\mathbf{s} + h\mathbf{e})\}\,,$$

for any unit vector $\mathbf{e}$. When data are available in the form of observations $Z(\mathbf{s}_i) = z_i$ at spatial locations $\{\mathbf{s}_i,\ i = 1, \ldots, n\}$ with inter-point distances denoted by $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, the variogram is usually estimated by its empirical version based on binning the data, as

$$2\,\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} (z_i - z_j)^2, \tag{1}$$

where $N(h) = \{(i, j) : h_{ij} \in b(h)\}$, $b(h)$ denotes an interval or 'bin' containing h and $|N(h)|$ denotes the cardinality of $N(h)$.

The large variability which the empirical variogram can exhibit is not always fully appreciated. The left hand panel of Figure 1 shows ten simulations of 100 observations from a stationary and isotropic Gaussian spatial process in $[0, 1]^2$ with exponential semivariogram $\gamma(h) = \sigma^2\{1 - \exp(-h/\phi)\}$, with the sill ($\sigma^2$) and range ($\phi$) parameters set to 1 and 0.15 respectively. We expect large variability at large distances, because of the sparsity of data; however, considerable variability is also displayed at smaller distances. There is also a strong degree of correlation across the values in the estimate, giving in some cases a false impression of a well estimated function from a single realisation.

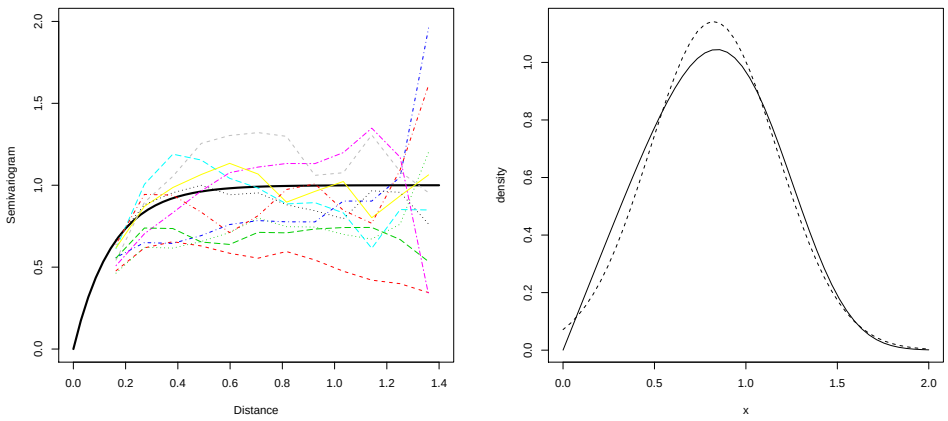

Figure 1: The left hand panel shows ten empirical semivariograms (dashed lines), each constructed from data simulated from a Gaussian spatial process with exponential semivariogram (full line). One hundred data points were arranged in a regular grid and the sill and range parameters were set to 1 and 0.15 respectively. The right hand panel shows the density function of the fourth-root transformation of a χ^2_1 random variable (full line) and the density function of a normal random variable with the same mean and standard deviation (dashed line).

The properties of the sample variogram are not easy to evaluate. The finite sampling distribution of $2\hat{\gamma}(h)$ has been tabulated by Davis and Borgman (1979) for univariate data using Fourier inversion methods, although the procedure could in principle be applied in higher dimensions. Asymptotic analysis has been carried out by Davis and Borgman (1982) and Cressie (1985), who established the sampling distribution for the empirical variogram and a robust version based on the fourth-root transform, introduced by Cressie and Hawkins (1980). Pardo-Igúzquiza and Dowd (2001) and Marchant and Lark (2004) discuss the computation of the covariance matrix of the empirical variogram, in the original scale, as a potential route to confidence intervals based on the asymptotically normal distribution derived by Davis and Borgman (1979). However, even in the Gaussian scenario, skewness is an issue as demonstrated by Baczkowski and Mardia (1987) who propose a log-normal approximation for the distribution of the sample variogram. Cressie (1993, Section 2.4.2) shows that the distribution can be written as a weighted sum of independent χ^2_1 random variables.

These issues have hampered the use of the sample variogram as an inferential tool. In the case of independent spatial data, the distributional properties of the sample variogram can be evaluated and exploited to form the basis of a test for the presence of spatial correlation, as described by Diblasi and Bowman (2001). Similarly, a test of goodness-of-fit assuming known parameters can be constructed, as discussed in Maglione and Diblasi (2004). However, beyond these rather special cases, diagnostic and inferential procedures based on the sample variogram have proved difficult to develop.

In the following sections, the sampling properties of the empirical variogram will be explored and used to construct diagnostic and inferential procedures. This is ap-

proached not on the scale defined by (1) but on the scale used by the robust approach proposed by Cressie and Hawkins (1980). Section 2 considers the issue of estimating the standard error of the binned variogram, allowing pointwise confidence bands to be constructed. This has an immediately useful application in allowing the informative comparison of variograms from different datasets, as described in Section 3. The provision of information on variance also allows procedures to be developed to assess the assumption of isotropy and this is discussed in Section 4. Another simplifying assumption is stationarity and assessment procedures for this are proposed in Section 5. Some further discussion is given in Section 6. Real data illustrations are provided throughout the paper for the methods developed.

2. Standard errors and confidence intervals

The raw materials for the sample variogram (1) are the squared differences $(z_i - z_j)^2$, which are scaled χ^2_1 random variables under the assumption of a Gaussian process with constant mean; see Cressie and Hawkins (1980). The considerable skewness associated with this distribution creates difficulties for further analysis which can be addressed by a suitable transformation. A very convenient candidate is the square-root absolute value transformation $\sqrt{|z_i - z_j|}$ or, equivalently, the fourth-root transformation of the squared differences. This has an effect which is very close to that of the function $\Phi^{-1}(F(\cdot))$ which transforms a χ^2_1 random variable with cdf F to a normal random variable with cdf Φ (see Figure 1, right hand panel) and it forms the basis of the robust estimate of the variogram proposed by Cressie and Hawkins (1980). This is based on

$$\hat{\gamma}^*(h) = \frac{1}{|N(h)|} \sum_{(i,j)\in N(h)} \sqrt{|z_i - z_j|} \tag{2}$$

which, as the analysis below shows, is an estimate of $\gamma^*(h) = 0.977741\{\gamma(h)\}^{\frac{1}{4}}$. An estimate of the semivariogram in the original scale can then be produced by the back-transformation $\tilde{\gamma}(h) = \{\hat{\gamma}^*(h)/0.977741\}^4$. Cressie and Hawkins (1980) show that the modified transformation $\{\hat{\gamma}^*(h)\}^4/(0.457 + 0.494/|N(h)|)$ adjusts for bias in the mean; however, the simpler transformation remains appropriate when transforming the endpoints of confidence intervals.

In terms of exploring the properties of the empirical variogram and constructing estimates of its variability, the square-root absolute value scale has the further strong advantage of using as building blocks quantities which are very well approximated by normal random variables. This gives a very helpful route to theoretical properties and a secure basis for the construction of standard errors and other characterisations of variability. In order to explore this, it is convenient to denote the covariance function by $C(h) = \sigma^2 - \gamma(h)$, where $\sigma^2 = \mathsf{var}\{Z(\mathbf{s})\}$, and to evaluate the simple properties of the differences $(Z_i - Z_j)$ as

$$\begin{aligned} \mathbb{E}\{Z_i - Z_j\} &= 0, \\ \mathsf{var}\{Z_i - Z_j\} &= 2\gamma(h_{ij}), \end{aligned}$$

$$\begin{aligned}
\mathsf{corr}\{Z_i - Z_j, Z_k - Z_l\} &= \frac{C(h_{ik}) + C(h_{jl}) - C(h_{jk}) - C(h_{il})}{\sqrt{\mathsf{var}\{Z_i - Z_j\}\,\mathsf{var}\{Z_k - Z_l\}}} \\
&= \frac{\gamma(h_{jk}) + \gamma(h_{il}) - \gamma(h_{ik}) - \gamma(h_{jl})}{2\sqrt{\gamma(h_{ij})\gamma(h_{kl})}}, \qquad (3)
\end{aligned}$$

where, for future reference, this last correlation is denoted by ρ_{ijkl}. The corresponding properties of $d_{ij} = \sqrt{|Z_i - Z_j|}$ can then be derived from mean and variance results in Diblasi and Bowman (2001) and a correlation result in Cressie (1985):

$$\begin{aligned}
\mathbb{E}\{d_{ij}\} &= \frac{2^{1/4}\Gamma(3/4)}{\sqrt{\pi}}\left(2\gamma(h_{ij})\right)^{1/4} = 0.977741\gamma(h_{ij})^{1/4}, \\
\mathsf{var}\{d_{ij}\} &= \frac{\sqrt{2}\left(\sqrt{\pi} - \Gamma^2(3/4)\right)}{\pi}\sqrt{2\gamma(h_{ij})} = 0.172402\sqrt{\gamma(h_{ij})}, \\
\mathsf{corr}\{d_{ij}, d_{kl}\} &= \Gamma^2(3/4)\left(\sqrt{\pi} - \Gamma^2(3/4)\right)^{-1}\left(\left(1 - \rho_{ijkl}^2\right)H(3/4; 3/4; 1/2; \rho_{ijkl}^2) - 1\right) \\
&= 5.545063\left((1 - \rho_{ijkl}^2)H(3/4; 3/4; 1/2; \rho_{ijkl}^2) - 1\right),
\end{aligned}$$

where H and Γ denote the hypergeometric and the Gamma functions, respectively.

It is not practical to work with the full set of square-root differences d_{ij} because of the size of the resulting covariance matrix which has $\{n(n-1)/2\}^2$ entries and so the means of the d_{ij}'s within the bins $b(h)$ as shown in (2) and originally defined in (1) are considered. When constructing the bins for this procedure, it is common practice to use a regularly spaced set of boundary points on the distance axis. However, as the variogram is likely to change more rapidly at smaller distances than at larger ones it is worthwhile considering other strategies. Possibilities include boundary points which, as far as possible, equalise the number of pairs in each bin and the use of the list of unique distances where there is a high degree of replication for data lying on a spatial grid. After some experimentation, it was decided to adopt boundary points which are regularly spaced on a log scale which has the advantage of focussing attention on smaller distances and stabilising behaviour at larger distances by including more observations than a simple regular grid on the distance scale. As a further useful adjustment, the bin is represented not by its mid-point but by the mean of the distances lying inside the bin, to track the observed data as closely as possible. For the number of bins B, analogy with the suggestion by Freedman and Diaconis (1981) for the bin size of histograms leads to the proposal of $B = \lceil\{n(n-1)/2\}^{1/3}\rceil$ for the variogram, where $\lceil\cdot\rceil$ denotes the integer ceiling.

If the binned means are denoted by $\{d_1, \ldots, d_B\}$, then

$$\begin{aligned}
\mathbb{E}\{d_b\} &= 0.977741\frac{1}{n_b}\sum_{(i,j)\in\mathcal{I}_b}\gamma(h_{ij})^{1/4}, \\
\mathsf{var}\{d_b\} &= \frac{1}{n_b^2}\sum_{(i,j),(k,l)\in\mathcal{I}_b (k,l)\neq(i,j)}\mathsf{cov}\{d_{ij}, d_{kl}\} + \frac{1}{n_b^2}\sum_{(i,j)\in\mathcal{I}_b}\mathsf{var}\{d_{ij}\}, \\
\mathsf{cov}\{d_{b_1}, d_{b_2}\} &= \frac{1}{n_{b_1}n_{b_2}}\sum_{(i,j)\in\mathcal{I}_{b_1},(k,l)\in\mathcal{I}_{b_2}}\mathsf{cov}\{d_{ij}, d_{kl}\}
\end{aligned}$$

where $\mathcal{I}_b$ denotes the set of index pairs identifying the d_{ij} values in bin b and n_b is the number of index pairs in this bin. The standard error at each bin is then given by an estimate of $\sqrt{\mathsf{var}\{d_b\}}$. A simple strategy for constructing this is to substitute $\tilde{\gamma}$ for γ in the expressions derived for $\mathsf{var}\{d_{ij}\}$ and $\mathsf{cov}\{d_{ij}, d_{kl}\}$ above and to substitute these in turn into the expression for $\mathsf{var}\{d_b\}$. A confidence interval on the fourth-root scale is then easily constructed by adding and subtracting two standard errors to d_b. To translate this back to the variogram scale, the same transform used for the semivariogram estimator can be applied to the end points to obtain an approximate confidence interval for γ.

Simulations were conducted to assess the performance of this approach. Data were simulated from a Gaussian process with exponential semivariogram for range parameter $\phi = 0.15, 0.25$ and unit variance. The spatial locations took the form of a regular grid for a variety of sample sizes, namely 49, 100 and 225 points in $[0, 1]^2$. The Freedman–Diaconis rule suggests 11, 17 and 30 bins for these sample sizes, although some of these are empty due to the regular structure, reducing to 10, 15 and 24 bins respectively. The number of replicates was 500 in each case. If the true coverage is 95% then the empirical coverage is likely to lie in the range $0.95 \pm 2\sqrt{0.95 \times 0.05/500}$ and this region is highlighted with grey shading in the panels of Figure 2 for reference. In order to check the suitability of the approach to the construction of confidence intervals described above, $\mathsf{var}\{d_b\}$ was calculated by substitution of the true values of γ rather than estimates. The results are shown in the right hand column where the empirical coverage lies comfortably within the shaded tolerance region. Following the recommendation of Cressie (1985) and many other authors, the range of spatial distances has been restricted because of small sample sizes and high variability at large distances. Here the upper limit has been set to 0.7 which is close to one half of the maximum observed distance.

The left hand column in Figure 2 shows the effect of using the estimate $\tilde{\gamma}$ in the construction of $\mathsf{var}\{d_b\}$. This shows marked departures from the target coverage. A potential explanation lies in the numerator of the correlations computed in (3). Even when the inter-pair distances h_{ij}, h_{kl} are small, the derived distances h_{jk}, h_{il}, h_{ik}, h_{jl} may still be large if $\mathbf{s}_i$, $\mathbf{s}_j$ are well separated from $\mathbf{s}_k$, $\mathbf{s}_l$. The poor estimation of γ at large h may therefore have a large and deleterious effect.

There are three features which it is reasonable to expect in a variogram, namely that it is a smooth function, that it is non-negative everywhere and that it is monotonically increasing. (The monotonicity requirement is applicable in most situations although hole-effect variograms for periodic processes are an exception.) The estimate of γ which is used in the construction of the standard errors may therefore usefully be stabilised by incorporating these features.

An alternative convenient method of achieving this is through p-spline smoothing, described by Eilers and Marx (1996) and developed by many subsequent authors. A smooth function can be represented as $\sum_{j=1}^{p} \alpha_j \varphi_j(x)$, using a large number (say $p = 100$) of b-spline basis functions φ_j. In vector-matrix form, this can be expressed as

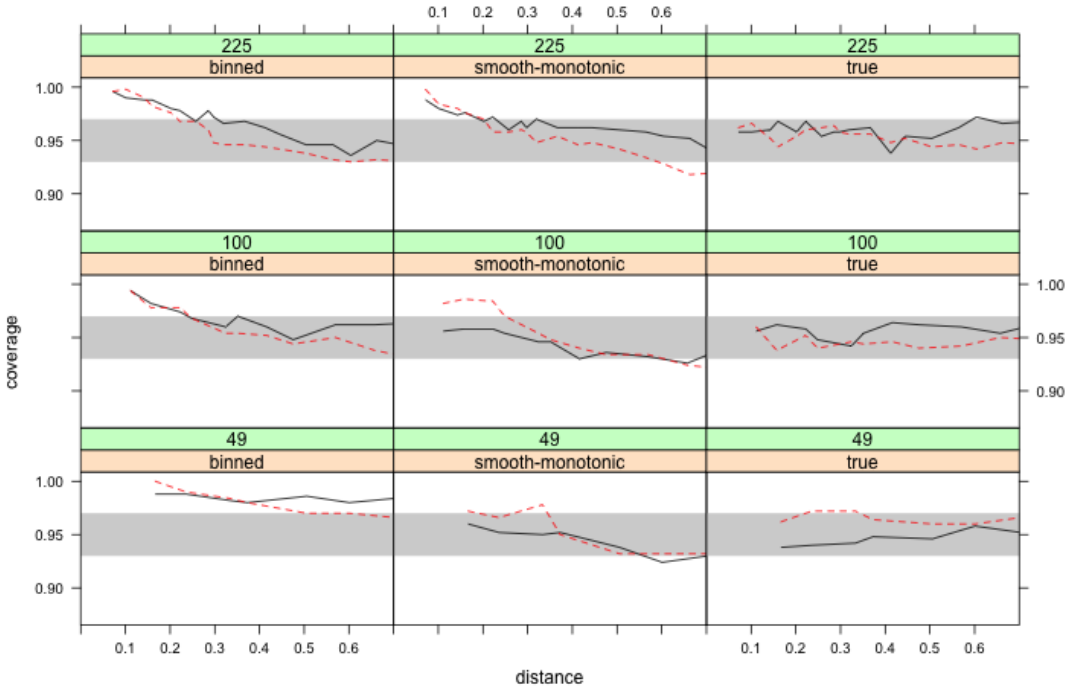

Figure 2: Empirical coverages of confidence intervals for the semivariogram. The columns refer to the estimate of γ used in the calculation of standard error. The rows refer to different sample sizes. The full and dashed lines show the results for data simulated from an exponential model with variance 1 and range parameter ϕ set to 0.15 and 0.25 respectively. The shaded region indicates a tolerance band for the empirical coverage when the true coverage is the target value of 0.95.

$\mathbf{v} = A\boldsymbol{\alpha}$, where $\mathbf{v}$ denotes the vector of bin means of the squared differences $(z_i - z_j)^2$, A is a design matrix whose columns evaluate the b-spline basis functions at the bin distances and $\boldsymbol{\alpha}$ represents the vector of b-spline coefficients. The p-spline approach is to use least squares to estimate $\boldsymbol{\alpha}$ but to incorporate a penalty term to ensure smoothness. This is achieved by minimising $(\mathbf{v}-A\boldsymbol{\alpha})^T(\mathbf{v}-A\boldsymbol{\alpha})+\lambda\boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha}$ where D is a suitable differencing matrix, for example $\boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha} = \sum_{i=2}^{p-1}(\alpha_{i-1} - 2\alpha_i + \alpha_{i+1})^2$. This leads to the solution $\hat{\boldsymbol{\alpha}} = (A^T A + \lambda D^T D)^{-1} A^T \mathbf{v}$. The smooth estimate can be written as $S\mathbf{v}$, with the smoothing matrix $S = A(A^T A + \lambda D^T D)^{-1} A^T$. The penalty parameter λ can be chosen to give the smoothing procedure a specified approximate degrees of freedom or 'effective dimension' expressed in $\mathrm{tr}\{S\}$, where tr denotes the trace; see Eilers and Marx (1996) for the details. The degrees of freedom in our study are set to $0.8B$, where B is the number of bins in the empirical variogram. This produces light smoothing. The choice of the degree of smoothing was explored by simulation and the results are available in the supplementary material which is available on the journal website. The results provide some evidence of stability over the choice of degrees of freedom and that the monotonicity is the more important feature of the estimator. However, in the more demanding cases such as $\phi = 0.25$ with larger sample size, the 0.8 setting proves to be the most effective.

A possible strategy to achieve monotonicity is isotonic regression, as proposed by Kim and Boos (2004). However, an alternative approach, which fits well with a p-spline formulation, is simply to add a further penalty term to penalise non-monotonic behaviour, as proposed by Bollaerts et al. (2006). This uses the function $\delta(x)$, which is 1 if $x > 0$ and 0 otherwise, to create the penalty $\sum_{j=2}^{p} \delta(\alpha_{j-1} - \alpha_j)\{\alpha_{j-1} - \alpha_j\}^2 = \boldsymbol{\alpha}^T M^T M\boldsymbol{\alpha}$ for a suitable matrix M. We then minimise $(\mathbf{v}-A\boldsymbol{\alpha})^T(\mathbf{v}-A\boldsymbol{\alpha})+\lambda\boldsymbol{\alpha}^T D^T D\boldsymbol{\alpha}+\kappa\boldsymbol{\alpha}^T M^T M\boldsymbol{\alpha}$. The penalty form of construction does not strictly guarantee monotonicity but ensures that any non-monotonicity is minor if a suitably large value of κ, for example $\kappa = 10\lambda$, is employed. Non-negativity can also be promoted in a very similar manner, by including the further penalty $\kappa\boldsymbol{\alpha}^T N^T N\boldsymbol{\alpha}$, where N is a diagonal matrix with 1's corresponding to negative values of $\boldsymbol{\alpha}$. Note that, because M and N involve $\boldsymbol{\alpha}$, iteration is required to reach a solution.

The left hand panel of Figure 3 shows the empirical variogram (1) constructed from 100 observations simulated from a Gaussian process with exponential covariance function, with variance 1 and range parameter 0.15. The true variogram is displayed as a dashed line. Substantial underestimation at high distances is apparent. The use of p-splines with very light smoothing, corresponding to 10 degrees of freedom, stabilises the estimate a little by reducing some of the fluctuations. However, in this case it is the additional use of the monotonicity penalty which produces a very beneficial effect. The right hand panel of Figure 3 illustrates the construction of confidence intervals for the semivariogram from the same set of simulated data, with the square-root scale as a starting point and using a smooth-monotonic approach in the calculation of the standard errors.

The central column of Figure 2 shows the improvement in coverage which is gained by applying p-spline smoothing to the empirical variogram (1) and substituting the

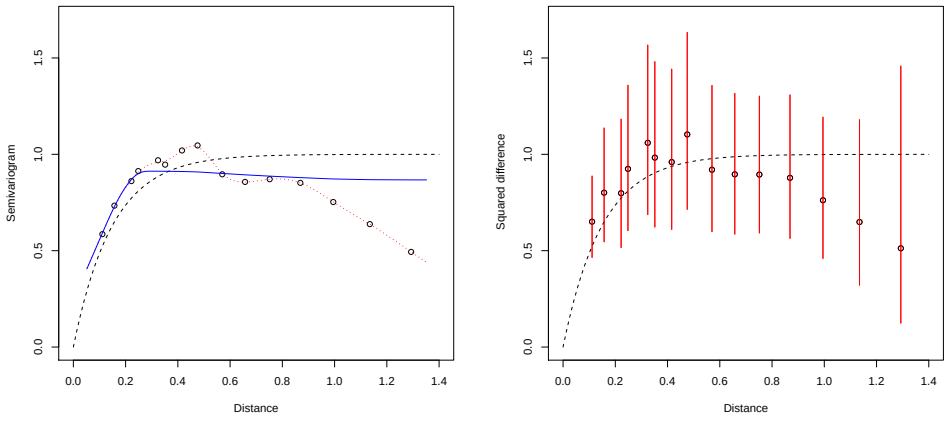

Figure 3: The left hand panel shows an empirical semivariogram (points) with smooth (dotted line) and smooth monotonic (full line) estimators superimposed, constructed from 100 observations simulated from a Gaussian process with exponential semivariogram. The right hand panel shows confidence intervals attached to an empirical semivariogram constructed from an estimate on the fourth-root scale. In both plots, the true semivariogram is shown as a dashed line.

result into the expression for the standard error of $\hat{\gamma}^*$, the semivariogram on the fourth-root scale. Note that the estimator $\hat{\gamma}^*$ itself remains in simple binned form, while smoothing is employed in the construction of the standard errors. There is a marked improvement in coverage, which is now more successful in hitting the target value of 95%.

More extensive simulations were carried out to confirm that this effect is consistent across other variogram shapes. The Matérn family allows easy movement from the exponential (shape parameter $\kappa = 0.5$) to smoother shapes defined by $\kappa = 1, 5$. As the κ increases, the variogram approaches the Gaussian model and the value $\kappa = 5$ is already close to this. A nugget effect can also be included. Although this simply introduces a shift in the vertical scale of a variogram, examining the performance through simulation allows any effects of the non-linear transformations involved in the square-root absolute value to be monitored. The details of these simulations are provided in the supplementary material. A summary is provided in Table 1 which reports the proportion of cases where the coverage falls within the tolerance range for the simulation. These proportions are higher, and sometimes substantially so, for the smooth-monotonic approach in the vast majority of cases, indicating improved coverage properties over the use of the simple binned estimator in the calculation of standard errors.

3. Comparing variograms from different datasets

The availability of an estimated covariance matrix for an empirical variogram has an immediate application in comparing variograms from different datasets. There

				Binned			Smooth–monotonic		
ϕ	κ	τ	n:	49	100	225	49	100	225
	0.50	0.00		0.00	0.52	0.50	0.95	0.82	0.71
		0.25		0.00	0.10	0.71	0.83	0.60	0.65
0.15	1	0.00		0.33	0.50	0.41	0.33	0.70	0.76
		0.25		0.17	0.70	0.65	1.00	0.90	0.59
	5	0.00		0.67	0.40	0.76	0.67	0.80	0.88
		0.25		0.00	0.70	0.47	0.83	0.80	0.82
	0.50	0.00		0.00	0.48	0.64	0.77	0.58	0.59
		0.25		0.33	0.40	0.35	0.83	0.40	0.53
0.25	1	0.00		0.33	0.40	0.59	0.83	0.60	0.53
		0.25		0.00	0.40	0.53	0.67	0.50	0.47
	5	0.00		0.33	0.60	0.65	0.83	0.80	0.82
		0.25		0.33	0.70	0.47	0.83	0.30	0.47

Table 1: The proportion of cases where the coverage of confidence intervals for the variogram falls within the simulation tolerance range, across all spatial distances less than 0.7. Confidence intervals constructed from standard errors based on the binned empirical variogram and from a smooth monotonic estimate are considered. The results are documented for data simulated on a regular grid across the unit square from Matérn models with shape parameter κ, range parameter ϕ, nugget effect τ and sample size n.

are relatively few proposals for methods of tackling this problem. Crujeiras et al. (2007) and Crujeiras et al. (2008) consider spectral methods in the rather special case where the data are regularly spaced. From a very different perspective, Zhu et al. (2002) provide asymptotic inference for the spatial cumulative distribution function, developing hypothesis testing to detect a difference in the spatial random processes at two time points. However, the availability of estimates of variograms and their standard errors, as discussed in Section 2, provides a very direct route to the assessment of evidence for differences.

If $\hat{\gamma}_1^*$ and $\hat{\gamma}_2^*$ are constructed using the same binning scheme and are statistically independent, for example arising from different locations or the same location but well separated time points, then the covariance matrix of the difference between these two estimates is simply the sum of the individual covariance matrices, (V_1+V_2). Variograms which are identical on the usual scale will also lead to equality on the fourth-root scale, so evidence for differences can be evaluated through the test statistic

$$(\hat{\gamma}_1^* - \hat{\gamma}_2^*)^T (\hat{V}_1 + \hat{V}_2)^{-1} (\hat{\gamma}_1^* - \hat{\gamma}_2^*), \tag{4}$$

where the estimates $\hat{V}_1$ and $\hat{V}_2$ are constructed in the manner described in Section 2. Under the null hypothesis that γ_1 and γ_2 are equal, $(\hat{\gamma}_1^* - \hat{\gamma}_2^*)$ is approximately normally distributed with mean 0, and the test statistic (4) is approximately χ^2_B, where B is the number of bins in the common binning scheme.

The effectiveness of this approach is assessed through simulations in Table 2. Data

			$\phi = 0.15$		$\phi = 0.25$	
c_σ	c_ϕ	n:	49	100	49	100
1.0	1.0		0.052	0.050	0.026	0.036
1.5	1.0		0.086	0.140	0.062	0.126
2.0	1.0		0.198	0.374	0.178	0.262
4.0	1.0		0.786	0.980	0.716	0.934
1.0	2.0		0.090	0.174	0.090	0.146
1.0	5.0		0.508	0.842	0.576	0.688

Table 2: Simulated power for a test of equality of two variograms, with nominal significance level 0.05. The data are simulated from a Gaussian process over a regular square grid of sample size n, using an exponential semivariogram. One group has $\sigma = 1$ and the value of ϕ shown in the top row of the table, while the other group has the parameters $c_\sigma\sigma$ and $c_\phi\phi$, using the scalings shown in the left hand columns.

were simulated from Gaussian processes with mean 0 and an exponential semivariogram over a square regular grid in $[0, 1]^2$ with a variety of sample sizes. One group was generated with scale and range parameters $\sigma = 1$ and $\phi = 0.15, 0.25$, while the parameters used to generate the data for the other group were adjusted by scaling as $c_\sigma\sigma$ and $c_\phi\phi$ respectively. Table 2 shows the empirical power of the test, for a nominal significance level of 0.05. The apparent small deflation of empirical size for large ϕ may be a result of the estimation of the covariance matrices in (4). However, the evidence for this behaviour is weak because the quoted values are only around two standard errors from 0.05 using 500 simulations. More generally, the test is shown to be very effective.

A real application arises from a biomonitoring study in Galicia, in North-West Spain. Mosses are used as a means of measuring levels of heavy metal concentrations in the atmosphere, since most of the nutrient uptake of the mosses is from the air. This technique for large-scale monitoring of long-range transport processes has been used in Galicia over the last decade, as described by Fernández et al. (2005). In 2006, in both March and September, measurements of different metals were collected at 148 points lying almost in a regular grid over the region with 15 km spacing in north-south and east-west directions, as shown in the top left panel of Figure 4. According to the ecologists expertise, the period between the two samples, passing from a humid to a dry season, is enough time to guarantee the independence of the observed processes.

The top right panel of Figure 4 shows the empirical variograms ($\tilde{\gamma}$ as defined near the beginning of Section 2) for log(Co) data collected in March and September. A visual comparison raises the question of whether these two processes have the same covariance structure, and in particular whether the sills (variances) may be different. The addition of pointwise confidence intervals in the bottom left hand panel of Figure 4 immediately casts doubt on this suggestion, through an indication of the large variability present in these estimates. A neater display is based on the standard error of the difference between the two variograms, which is easily computed on the fourth-root scale as $\sqrt{\text{se}_1^2 + \text{se}_2^2}$, where se_1 and se_2 denote the standard errors of $\hat{\gamma}_1^*$ and $\hat{\gamma}_2^*$. Constructing a

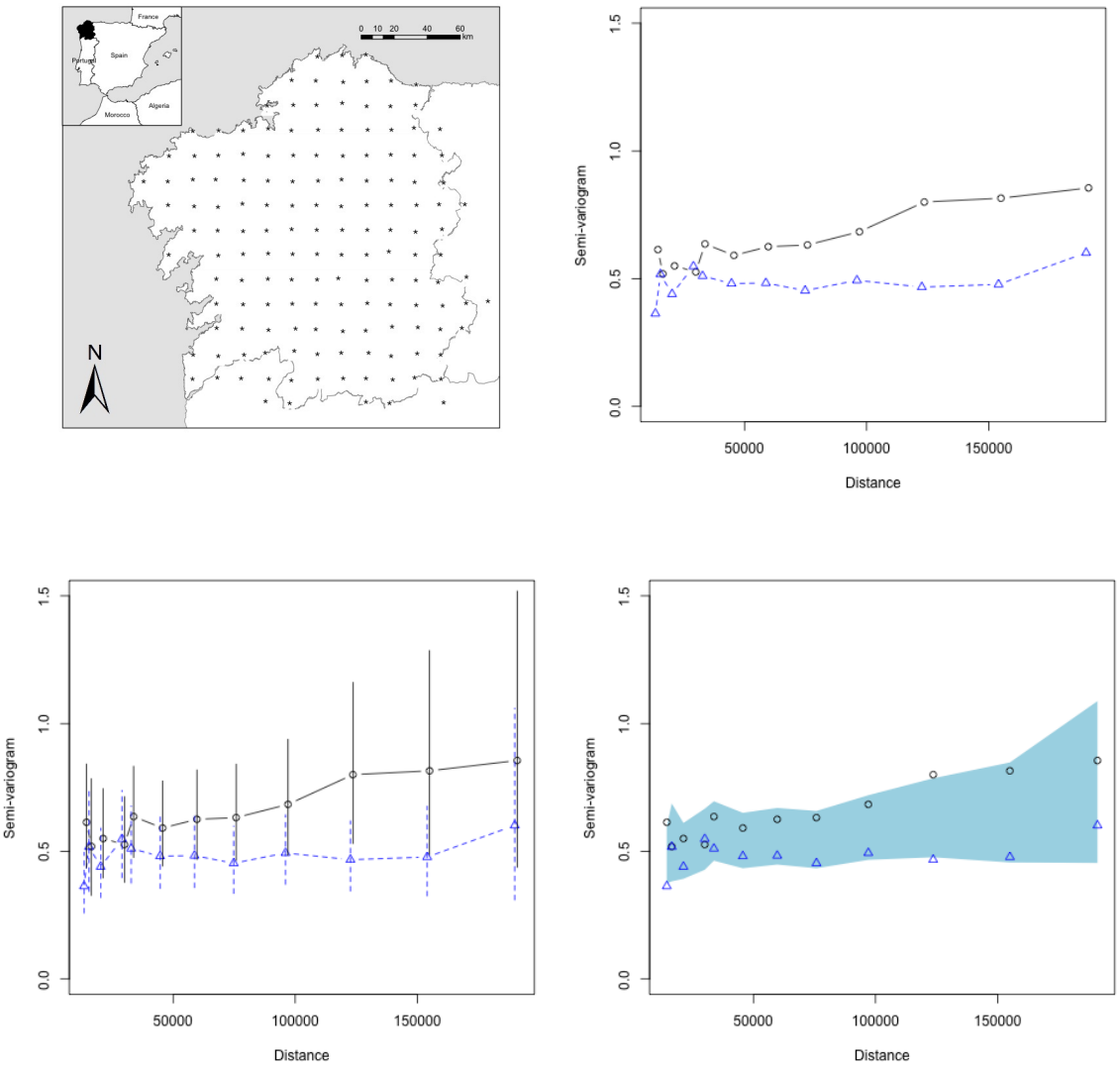

Figure 4: The top left panel shows sampling points in Galicia, in North-West Spain in 2006. The top right panel compares the semivariograms between March (circles and full lines) and September (triangles and dashed lines) for log(Co) from the mosses data. The lower panels show semivariograms with pointwise confidence intervals (left) and a reference band for equality (right).

band which is centered at the mean value of the two estimates of γ^*, and whose width is two standard errors of the difference, provides a very convenient graphical reference against which evidence of differences can be assessed. The bottom right hand panel of Figure 4 illustrates this on the original semivariogram scale, simply by transforming the end points of the reference band. Bowman and Young (1996) illustrate the use of reference bands of this type in a variety of other settings. On this occasion the reference band prevents over-interpretation of the differences between the two empirical semivariograms. This is confirmed by the application of the χ^2 test described above, which produces a p-value of 0.116, indicating lack of convincing evidence against the hypothesis that the two processes have the same spatial covariance structure.

4. Assessing isotropy

One of the simplifying assumptions discussed in Section 1 was isotropy, where the semivariogram $\gamma(\mathbf{h})$ of a stationary process reduces to the simpler form $\gamma(h)$ which depends only on the size $h = ||\mathbf{h}||$ of the separation distance and not its direction vector. If $\gamma(\mathbf{h})$ is reparameterised as $\gamma(h, \theta)$ where $\mathbf{h} = (h\cos\theta, h\sin\theta)$, then the suitability of an isotropic model can be examined by plotting separate empirical variograms for different ranges of direction θ or by plotting contours of the empirical variogram surface over the two-dimensional space defined by $\mathbf{h}$, as described by Banerjee et al. (2004, Section 2.3.2). Some formal tests for isotropy have been proposed, based on asymptotic results for the empirical variogram (1). Lu and Zimmerman (2005) use the spectral density, which requires regularly spaced data. Guan et al. (2004) use sub-sampling to construct an estimate of the covariance matrix and this requires selection of the size of the sub-samples as well as the number of lags. These authors also recommend identifying particular directions of interest, based on prior knowledge of the underlying process.

Variograms which are unaffected by a change in angle θ will also lead to absence of effect on the fourth-root scale. The mean and covariance descriptions of $\hat{\gamma}^*$ given in Section 2 therefore provide an opportunity for a quantitative assessment of the evidence for anisotropy. In this case $\hat{\gamma}^*(h, \theta)$ can be estimated by constructing bin means d_b over a two-dimensional grid of values of (h, θ). Evidence for anisotropy then rests on comparing an estimate based on (h, θ) with an estimate based on h alone. Smoothing techniques, as demonstrated by Bowman (2007), offer a simple means of doing this by borrowing strength across the bins to exploit smoothness but without making any further modelling assumptions. Denoting the vector of bin means by $\mathbf{d}$, fitted values based on smoothing can be expressed as $\hat{\gamma}_1^* = S_1\mathbf{d}$ and $\hat{\gamma}_0^* = S_0\mathbf{d}$, where S_1 and S_0 are smoothing matrices which incorporate distance and angle (S_1) and only distance (S_0). Since angle lies on a cyclical scale, this feature should be incorporated into the angle component of the smoothing matrix S_1. This is easily done with a two-dimensional p-spline basis where, for each distance index i, the penalty $\sum_{j=1}^{r}(\alpha_{ij} - \alpha_{i,p+1-j})^2$ requires the coefficients of the first r basis functions for angle to equal those of the last r basis functions. Here p is the number of basis functions in each dimension and r is the

number of overlapping basis functions, which is determined by the order of the spline basis and which takes the value $r = 3$ when the spline order is 3.

Under the assumption of isotropy, the difference between the smooth estimators, $(S_1 - S_0)\mathbf{d}$, has mean 0. Global evidence for anisotropy can then be quantified through the test statistic

$$\mathbf{d}^T S^T \hat{V_0}^{-1} S\mathbf{d}, \tag{5}$$

where $S = S_1 - S_0$ and V_0 is the covariance matrix for the isotropic case. The monotonicity and non-negativity constraints discussed in Section 2 are imposed in the estimation of V_0 but these are not adopted in the construction of S_0 and S_1, as use of the observed data here would resulting in the loss of a simple linear form for the smooth estimates. The use of smoothing matrices means that statistic (5) no longer has a simple χ^2 distribution. However, there are well known results on quadratic forms in normal random variables which allow distributional calculations to be performed computationally, to a high degree of accuracy. An exact representation in terms of a weighted sum of χ^2 random variables forms the basis of an algorithm developed by Imhof (1961). However, calculations based on matching moments with a single shifted and scaled χ^2 distribution are described by Bowman and Azzalini (1997) and are shown to be a simple but very effective approximation. Specifically, the test statistic (5) can be represented as $\mathbf{e}^T Q\mathbf{e}$, where $\mathbf{e}$ has a $N(\mathbf{0}, V_0)$ distribution and Q is a known symmetric matrix. The sth cumulant of $\mathbf{e}^T Q\mathbf{e}$ is $2^{s-1}(s-1)!\,\mathrm{tr}\,\{(V_0 Q)^s\}$, from which the moments may be computed and matched to an $a + b\chi^2_c$ distribution, and from which in turn the p-value is very easily calculated.

A simulation study was conducted in order to investigate the properties of this test and the results are shown in Table 3. The strength of anisotropy was controlled by two parameters, following Diggle and Ribeiro (2007). Direction was controlled by the anisotropy angle, defined by the direction with greater spatial continuity, and the anisotropy ratio, defined as the ratio between the ranges of the directions with greatest and smallest spatial continuity. Both regular grids and irregular patterns of point locations in $[0, 1]^2$ were used, the latter being created by uniform random sampling. Two different values of the process range parameter ϕ were used. Large values were specified as isotropy cannot be expressed clearly in data with small range correlation. Sample size was set to 100 and the number of simulations to 500. The degrees of freedom for smoothing were set to 12, to provide a good degree of flexibility when estimating the variogram surface over distance and angle while maintaining good control over variance. When smoothing under the smaller model, which depends only on distance, it is important to match the degree of smoothing used under the larger model, depending on distance and angle. Since all smoothing techniques incur bias, the use of comparable degrees of smoothing in the two models aims to cancel as far as possible the biases in the components of the test statistic (5) under the null hypothesis. Theoretical results on bias for p-spline smoothing are available only in the simplest of cases (Claeskens et al., 2009). However, a simple argument based on the numbers of p-spline coefficients in one-dimensional and two-dimensional settings, namely p and p^2, suggests that a comparable degree of smoothing is achieved by using in the one-covariate model the square-root of the number of degrees of freedom used in the two-covariate model. For

			regular		irregular	
angle	ratio	ϕ:	0.25	0.35	0.25	0.35
0	1		0.018	0.016	0.034	0.030
0	5		0.700	0.826	0.222	0.348
$\pi/4$	5		0.318	0.462	0.204	0.294

Table 3: Simulated power for a test of isotropy. The data are simulated from a Gaussian process over a regular square grid with sample size 100, using an exponential semivariogram.

12 degrees of freedom, this corresponds to 4 (rounded up) degrees of freedom for the smaller model.

The empirical power of the test of isotropy slightly undershoots the nominal size (0.05) when the underlying process is isotropic (ratio = 1). This may well arise from the use of an estimated covariance matrix which is not reflected in the distributional calculations. With strong anisotropy (ratio = 5), and with a regular grid structure which is aligned to the direction of anisotropy, the power of the test is high, reaching 0.826 when the range parameter is $\phi = 0.35$. We should not expect a test such as this to have very high power at modest sample sizes or with weak anisotropy. As the material in Section 2 demonstrated, estimation of a single variogram is already a process subject to a great deal of variability, so detecting changes in the variogram surface across a further variable is a demanding task. Nonetheless, the simulation results are encouraging.

Notice that when the orientation of a regular sampling grid does not align with the direction of anisotropy, or an irregular sampling pattern is used, the power of the test is much lower. Clearly, this is a design issue which suggests the benefit of careful consideration of likely directions of anisotropy before the locations of the sampling points are decided, as pointed out by Guan et al. (2004).

As an illustration, the left hand panels of Figure 5 were constructed from 100 data points which were simulated from an exponential semivariogram with range parameter 0.25 and with anisotropy along a north-south direction (ratio 5). The top left panel shows directional (north-south and east-west) empirical semivariograms, from which it is difficult to assess evidence for anisotropy. The lower left panel shows a smoothed variogram on the fourth-root scale, as a function of angle as well as distance, with variogram height coded by colour. Angle is represented as radial direction from the origin and for each angle the variogram is represented by the height of the surface along the radius. As the vector difference between a pair of sampling points does not have a sign associated with it, angles may be defined over $(0, \pi]$; however, it is helpful to construct the display over $(0, 2\pi]$ by reflection. The lower values in a north-south direction close to the origin suggest anisotropy, as the greater spatial continuity in this direction gives the variogram a slower rate of increase. The red contours mark the distance, in units of standard error, between the variogram as a function of angle and the standard variogram which takes no account of angle. Since anisotropy causes a slower rate of increase in one particular direction, it is negative differences which are

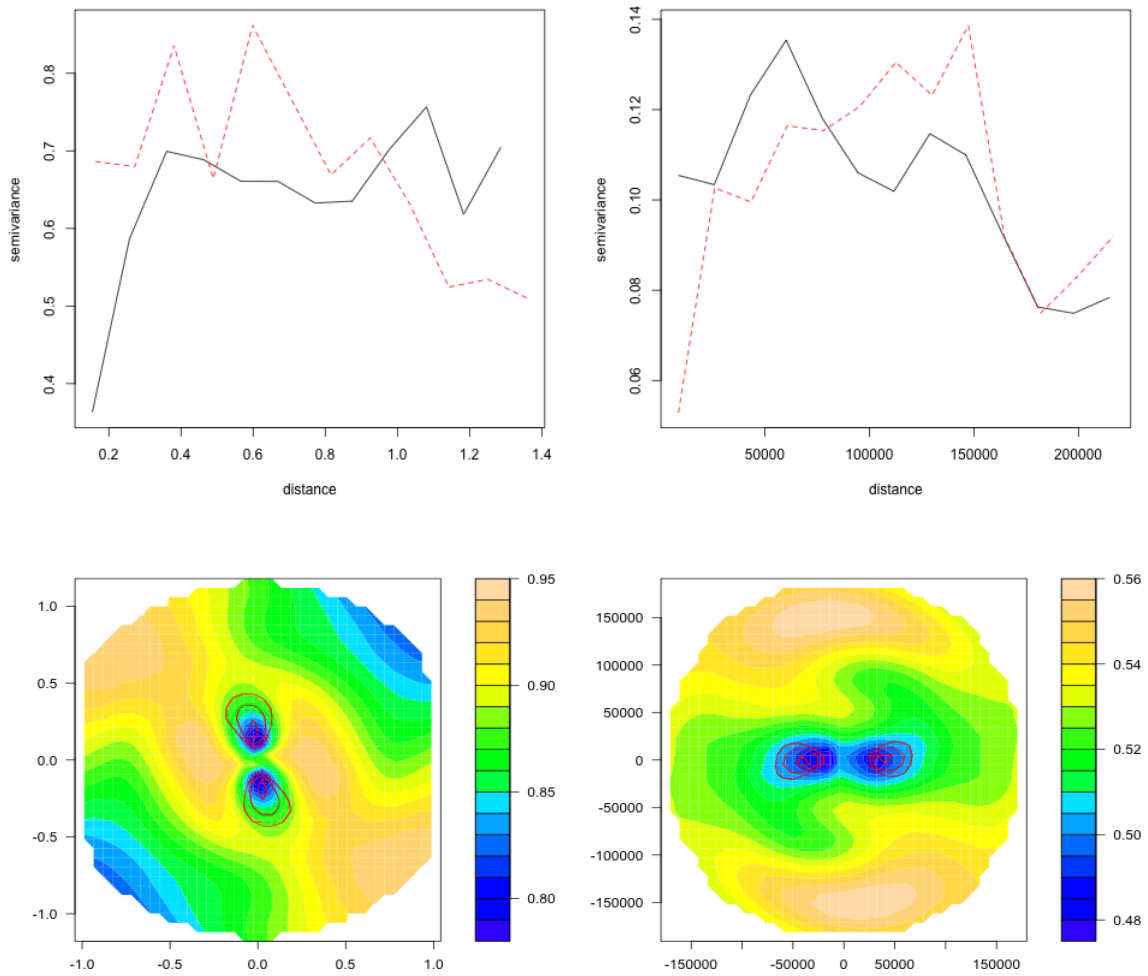

Figure 5: The top left hand panel shows directional (east-west, black full line, and north-south, red dashed line) semivariograms for 100 data points which were simulated from an exponential variogram with range parameter 0.25 and with anisotropy along a north-south direction (ratio 5). The lower left hand panel shows a smoothed estimate of the semivariogram on the fourth-root scale in a polar co-ordinate representation. The red contours at -2, -3 and -4 mark the distance, in units of standard error, between the variogram as a function of both distance and angle and the standard variogram which takes no account of angle. The right hand panels repeats these plots for the log(Hg) mosses data. Here the standard error contours are at -2, -2.5 and -3.

of interest. The contours at -2, -3 and -4 strengthen the evidence for a slower rate of increase of the variogram in a north-south direction. The statistical significance of this feature is represented by a p-value of < 0.001 from the isotropy test.

The right hand panels repeat this process for log(Hg) from the March 2006 mosses data. Although the shading and standard error contours suggest some evidence of anisotropy in an east-west direction, the global isotropy test indicates that the evidence in not entirely convincing (p-value of 0.082). An isotropic model may therefore reasonably be adopted.

5. Assessing stationarity

Of the simplifying assumptions for the variogram discussed in Section 1, the remaining one is stationarity, where the variogram $2\,\gamma(\mathbf{h}, \mathbf{s})$ loses its dependence on location and reduces to the simpler form $2\,\gamma(\mathbf{h})$. There are very few tools to assess the adequacy of this assumption. Fuentes (2005) describes a spectral approach which requires the sampling points to lie on a regular grid. Jun and Genton (2012) use asymptotic distributions to compare the spatial covariances between specified sub-regions of the spatial domain. Again, the existence of an estimated covariance matrix for the variogram on the fourth-root scale, as discussed in Section 2, creates an opportunity to compare stationary and non-stationary estimates of the variogram.

In the test of isotropy, the key was to associate an angle with each point pair and then to smooth the variogram across both distance and angle. For stationarity, the analogue is to associate a location with each point pair and then to smooth the variogram across both distance and location. Location may be identified simply by the average of the two points involved in each pair. Variogram bins can then be constructed over the three-dimensional space defined by location and inter-point distance. A three dimensional p-spline basis over this space allows the construction of a smoothing matrix S_1, while a one-dimensional basis smooths the variogram bins over distance but not location, with an associated smoothing matrix S_0. A test statistic of the same form as (5) can then be constructed. A p-value is then available through the quadratic form methods described in Section 4. Following the arguments presented in Section 4, the degrees of freedom for the three-dimensional and one-dimensional smoothing procedures were set to 20 and 3 respectively, with the latter being the integer ceiling of $20^{1/3}$.

The upper panels of Figure 6 illustrate this procedure using a set of simulated data. Non-stationarity was induced by the simple device of simulating from a stationary process over a 15×15 regular grid on $[0, 1]^2$ with exponential variogram, using variance 1 and range parameter 0.15, and then down-weighting the simulated values by the function $1 - w \exp\{-0.5||\mathbf{x} - (0.5, 0.5)^T||^2/0.2^2\}$ with w set to 0.5. This mimics simulation from a process whose variance is small near the point $(0.5, 0.5)^T$ and which increases radially towards the edges of the spatial region. The top left panel shows a plot of the observed data, from which the suitability of a stationarity assumption is difficult to assess. The other two panels show the variogram, on a fourth-root scale, at two specific

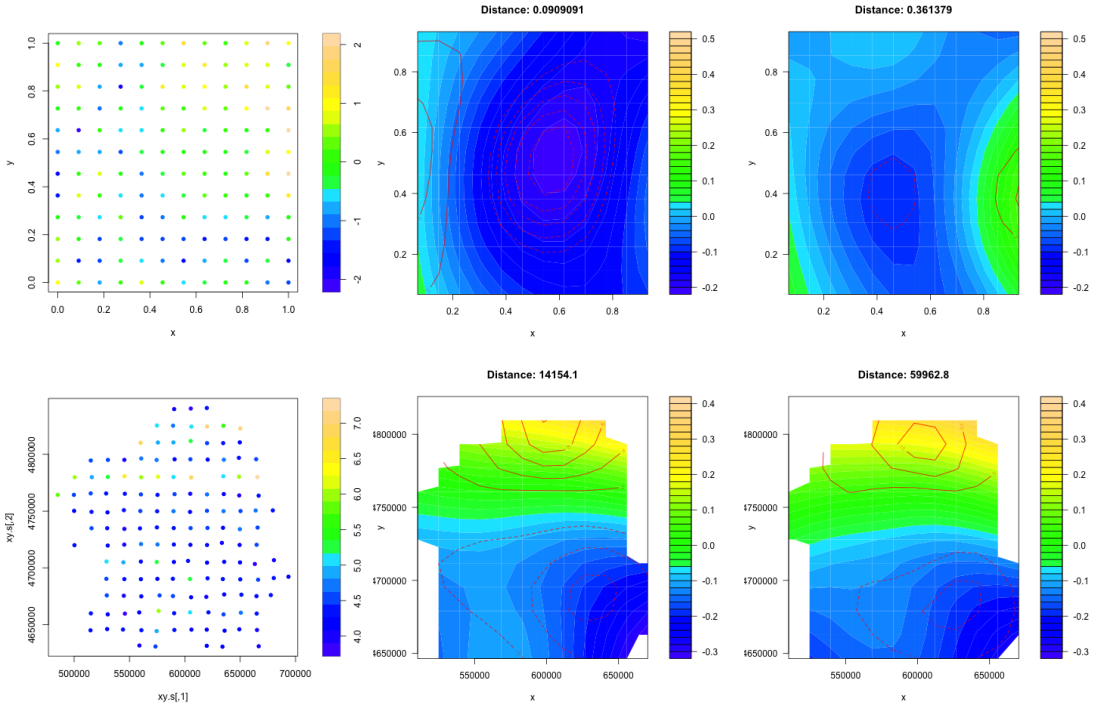

Figure 6: The top row of panels displays simulated data (left) with non-stationary covariance and the smoothed empirical semivariogram as a function of location, evaluated at two specific distances (middle and right). The bottom row of panels shows similar plots for the September log(Hg) measurements from the mosses data.

			regular			irregular		
n	weight	ϕ:	0.05	0.15	0.25	0.05	0.15	0.25
100	0.0		0.052	0.096	0.112	0.068	0.120	0.116
100	0.7		0.254	0.254	0.234	0.310	0.312	0.206
100	0.9		0.484	0.400	0.316	0.504	0.410	0.254
225	0.0		0.084	0.092	0.116	0.142	0.160	0.118
225	0.7		0.546	0.378	0.258	0.560	0.436	0.308
225	0.9		0.768	0.546	0.338	0.674	0.480	0.350

Table 4: Simulated size and power for the test of stationarity. The data are simulated from a Gaussian process over a regular square grid with sample size 100, using an exponential semivariogram and with 500 simulations in each case.

distances, 0.09 and 0.36, smoothed over spatial location. The most effective form of display is created by animation over variogram distance but in printed form a selection of plots at specific distances provides an adequate substitute. If stationarity applies then, for a particular distance, the values of the variogram should be identical at all spatial locations. The upper panels in Figure 6 show lower estimated variogram values near the centre of the plot and this pattern persists over all distances. Together with a p-value of 0.009, this gives very clear and effective indication of the non-stationarity which is known to be present in this simulated dataset.

Table 4 records the results of a simulation study to investigate the size and power of the test, using the same weight function employed above to induce non-stationarity, with a variety of values of w, sample size, covariance parameters and grid structure, and with 500 simulations in each case. The size of the test is slightly higher than the target of 0.05 for large spatial covariance, particularly with irregular sampling points. However, despite the fact that the identification of second-order patterns in modestly sized datasets is a challenging task, high power is achievable when the non-stationarity is strong.

The test was also applied to the September 2006 log(Hg) measurements from the mosses data and significant evidence of non-stationary was found ($p < 0.001$). The two distance plots in the lower middle and lower right hand panels of Figure 6 suggest lower variance in the southern part of the sampling region and higher variance in the north. This is consistent with the plot of the data in the lower left hand panel of Figure 6, where higher values are observed in the north.

6. Discussion

The determination of the sample properties of the empirical variogram, on a robust scale, enables the implementation of several very useful inferential procedures. Specifically, the construction of confidence bands, a method of variogram comparison and testing procedures to assess the suitability of simplifying hypotheses such as isotropy

and stationarity have been introduced and their performances evaluated in practice. The key step is the estimation of the covariance matrix of the empirical variogram on the fourth-root scale and here the use of p-splines to promote smoothness, non-negativity and monotonicity proved to be very helpful. Experimentation showed that only light smoothing should be applied at this stage and the recommendation of setting the degrees of freedom to $0.8B$ (where B is the number of variogram bins) while empirical, proved to be a useful and workable solution (details in the supplementary material). In terms of direct estimation of the variogram, smoothing does not necessarily produce estimators which are conditionally negative definite. Where this condition is required, it can be induced by use of the Shapiro–Botha projection, as described in general by Shapiro and Botha (1991) and in the context of smoothing by García-Soidán et al. (2003).

While the addition of approximate confidence bands to an empirical variogram is a valuable step, more focussed issues such as the evidence for differences between two variograms can also be addressed by combination of the associated variance matrices, firstly in the evaluation of a simple test statistic and secondly in the construction of a reference band for equality. The assumptions of isotropy and stationarity can also be assessed by suitable comparisons of the empirical variograms constructed with and without these particular properties. In addition to its role in stabilising the estimates of the variance matrix, smoothing allows the borrowing of strength across the variogram surface when this is a function not only of distance but also of direction (isotropy) or location (stationarity). Selection of the degrees of freedom at this stage has been guided by a simple dimensionality argument, using a small number of degrees of freedom for the one-dimensional case and the square (isotropy) or cube (stationarity) of this for the two- and three-dimensional spaces associated with the larger models. This is again an empirical strategy but the results have shown this to be effective in its performance and it has the merit of simplicity, in the absence of more complex theory or effective methods of automatic selection of smoothing parameters in this setting. There is a further advantage in allowing the application of simple quadratic form methods to evaluate the significance of the test statistics, without the need to reflect the variability of more complex methods of smoothing in the distributional calculations.

The introduction to the paper illustrated the high degree of variability which is present in a variogram and so high power cannot be expected when applying these tests with small sample sizes. Nonetheless, the simulations have demonstrated the ability of the procedures to detect marked differences, anisotropy and non-stationarity when these are present, providing very useful diagnostic procedures which can inform decisions on subsequent modelling. Where significant evidence of these characteristics is identified by the tests, graphical procedures are available to investigate their nature. These graphs provide very informative follow-up procedures to the global tests.

Software to implement the methods discussed in the paper will shortly be available in the `sm` package (Bowman and Azzalini, 2003) for the `R` statistical computing environment (R Development Core Team, 2011).

Acknowledgement

The work of Rosa Crujeiras was supported by Project MTM2008-03010 of the Spanish Ministry of Science and Innovation. The work of Adrian Bowman was partially supported by the same project and also benefitted by the discussions held under the auspices of project MTM2008-02901 from the Spanish Ministry of Science and Innovation. We would also like to thank Prof. Andrés Prieto for his computational assistance and the Ecotoxicology Research Group of University of Santiago de Compostela for providing the mosses datasets. The helpful comments of an associate editor and two referees are gratefully acknowledged.

Baczkowski, A., Mardia, K., 1987. Approximate lognormality of the sample semivariogram under a gaussian process. Communications In Statistics-Simulation and Computation 16 (2), 571–585.

Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, Florida.

Bollaerts, K., Eilers, P. H. C., van Mechelen, I., 2006. Simple and multiple p-splines regression with shape constraints. British Journal of Mathematical & Statistical Psychology 59, 451–469.

Bowman, A., 2007. Comparing nonparametric surfaces. Statistical Modelling 6, 1–21.

Bowman, A., Azzalini, A., 1997. Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford.

Bowman, A., Azzalini, A., 2003. Computational aspects of nonparametric smoothing with illustrations from the sm library. Computational Statistics & Data Analysis 42 (4), 545–560.

Bowman, A., Young, S., 1996. Graphical comparison of nonparametric curves. Applied Statistics-Journal of the Royal Statistical Society Series C 45 (1), 83–98.

Claeskens, G., Krivobokova, T., Opsomer, J. D., 2009. Asymptotic properties of penalized spline estimators. Biometrika 96 (3), 529–544.

Cressie, N., 1985. Fitting variogram models by weighted least-squares. Journal of the International Association for Mathematical Geology 17 (5), 563–586.

Cressie, N., Hawkins, D., 1980. Robust estimation of the variogram, I. Journal of the International Association for Mathematical Geology 12, 115–125.

Cressie, N. A., 1993. Statistics for spatial data. Wiley, New York.

Crujeiras, R. M., Fernandez-Casal, R., Gonzalez-Manteiga, W., 2007. Comparing spatial dependence structures using spectral density estimators. Environmetrics 18 (7), 793–808.

Crujeiras, R. M., Fernandez-Casal, R., Gonzalez-Manteiga, W., 2008. An l-2-test for comparing spatial spectral densities. Statistics & Probability Letters 78 (15), 2543–2551.

Davis, B. M., Borgman, L. E., 1979. Some exact sampling distributions for variogram estimators. Journal of the International Association for Mathematical Geology 11, 643–653.

Davis, B. M., Borgman, L. E., 1982. A note on the asymptotic-distribution of the sample variogram. Journal of the International Association for Mathematical Geology 14 (2), 189–193.

Diblasi, A., Bowman, A., 2001. On the use of the variogram in checking for independence in spatial data. Biometrics 57 (1), 211–218.

Diggle, P. J., Ribeiro, P. J., 2007. Model-based Geostatistics. Springer, New York.

Eilers, P., Marx, B., 1996. Flexible smoothing with b-splines and penalties. Statistical Science 11 (2), 89–102.

Emery, X., Ortiz, J. M., 2007. Weighted sample variograms as a tool to better assess the spatial variability of soil properties. Geoderma 140 (1-2), 81–89.

Fernández, J., Real, C., Couto, J., Aboal, J., Carballeira, A., 2005. The effect of sampling design on extensive biomonitoring surveys of air pollution. Science of the Total Environment 337, 11–21.

Freedman, D., Diaconis, P., 1981. On the histogram as a density estimator - l2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 57 (4), 453–476.

Fuentes, M., 2005. A formal test for nonstationarity of spatial stochastic processes. Journal of Multivariate Analysis 96 (1), 30–54.

García-Soidán, P., González-Manteiga, W., Febrero-Bande, M., 2003. Local linear regression estimation of the variogram. Statistics & Probability Letters 64 (2), 169–179.

Guan, Y., Sherman, M., Calvin, J., 2004. A nonparametric test for spatial isotropy using subsampling. Journal of the American Statistical Association 99 (467), 810–821.

Imhof, J., 1961. Computing distribution of quadratic forms in normal variables. Biometrika 48 (3-4), 419–426.

Jun, M., Genton, M. G., 2012. A test for stationarity of spatio-temporal random fields on planar and spherical domains. Statistica Sinica, 1737–1764.

Kim, H., Boos, D., 2004. Variance estimation in spatial regression using non-parametric semivariogram based on residuals. Scandinavian Journal of Statistics 31, 387–401.

Lu, N., Zimmerman, D., 2005. Testing for directional symmetry in spatial dependence using the periodogram. Journal of Statistical Planning and Inference 129, 369–385.

Maglione, D., Diblasi, A., 2004. Exploring a valid model for the variogram of an isotropic spatial process. Stochastic Environmental Research and Risk Assessment 18 (6), 366–376.

Marchant, B., Lark, R. M., 2004. Estimating variogram uncertainty. Mathematical Geology 36, 867–898.

Marchant, B. P., Lark, R. M., 2007. Robust estimation of the variogram by residual maximum likelihood. Geoderma 140 (1-2), 62–72.

Pardo-Igúzquiza, E., Dowd, P., 2001. Variance-covariance matrix of the experimental variogram: assessing variogram uncertainty. Mathematical Geology 33, 397–419.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Shapiro, A., Botha, J., 1991. Variogram fitting with a general class of conditionally non-negative definite functions. Computational Statistics & Data Analysis 11 (1), 87–96.