Title	A simulation comparison of estimators of spatial covariance parameters and associated bootstrap percentiles
Authors(s)	Kelly, Gabrielle E., Menezes, Raquel
Publication date	2018-09
Publication information	Kelly, Gabrielle E., and Raquel Menezes. “A Simulation Comparison of Estimators of Spatial Covariance Parameters and Associated Bootstrap Percentiles” 8, no. 6 (September, 2018).
Publisher	UCLA Department of Statistics
Item record/more information	http://hdl.handle.net/10197/10492

Journal of Environmental Statistics

September 2018, Volume 8, Issue 6. http://www.jenvstat.org

A simulation comparison of estimators of spatial covariance parameters and associated bootstrap percentiles

Gabrielle E. Kelly
School of Mathematics and Statistics
University College Dublin
Ireland

Raquel Menezes
Department of Mathematics and Applications
Minho University- Azurém
Portugal

Abstract

A simulation study is implemented to study estimators of the covariance structure of a stationary Gaussian spatial process and a spatial process with t-distributed margins. The estimators compared are Gaussian restricted maximum likelihood (REML) and curve-fitting by ordinary least squares and by the nonparametric Shapiro-Botha approach. Processes with Matérn covariance functions are considered and the parameters estimated are the nugget, partial sill and practical range. Both parametric and nonparametric bootstrap distributions of the estimators are computed and compared to the true marginal distributions of the estimators.

Gaussian REML is the estimator of choice for both Gaussian and t-distributed data and all choices of the Matérn covariance structure. However, accurate estimation of the Matérn shape parameter is critical to achieving a good fit while this does not affect the Shapiro-Botha estimator. The parametric bootstrap performed well for all estimators although it tended to be biased downward. It was slightly better than the nonparametric bootstrap for Gaussian data, equivalent to it for t-distributed data and worse overall for the Shapiro-Botha estimates.

A numerical example, obtained from environmental monitoring, is included to illustrate the application of the methods and the bootstrap.

Keywords: Gaussian random field, variogram, restricted maximum likelihood, variogram curve-fitting, Shapiro-Botha estimation, spatial bootstrap.

1. Introduction

Estimation of covariance parameters through the variogram is usual practice in spatial analysis. In addition to assessing spatial dependence, the variogram is also commonly used for spatial prediction i.e. kriging. Standard methods for estimating the variogram include obtaining an empirical variogram by Matheron's method or the variogram cloud and then fitting this empirical variogram by a theoretical variogram model. Diggle and Ribeiro (2007, Section 5.3.3) argue against the use of the empirical variogram for inference and they favour estimation methods based on the likelihood function. Such estimation methods are most fully developed for spatial Gaussian linear models. A third method that has been found popular is the Shapiro-Botha nonparametric estimator of the variogram that avoids specification of a variogram form (Shapiro and Botha 1991).

However, even in likelihood approaches, properties of estimators, such as standard errors, although theoretically available via the observed Fisher information in the case of likelihood estimation, are not available in practice. Diggle and Ribeiro (2007, Section 5.4.2) report that quadratic approximations to the log-likelihood surface are often poor and standard errors obtained from the Hessian matrix unreliable. In addition Warnes and Ripley (1987) give examples of spatial Gaussian processes where the profile likelihood of covariance parameters estimated by maximum likelihood can be multimodal. The bootstrap can be employed to provide properties of estimates for all methods - likelihood and variogram curve-fitting. The performance of the bootstrap in this context needs to be explored and this will be done in a simulation study. Bonat and Ribeiro (2016) argue that inferences based on standard errors do not make sense when analyzing small data sets, especially for the covariance parameters. The profile likelihoods may be highly asymmetric. Therefore percentiles of the bootstrap distribution will be used to characterize these profile likelihoods in the simulation study.

Zimmerman and Zimmerman (1991) carry out a comparison of spatial variogram estimators that includes curve-fitting methods to an empirical variogram, maximum likelihood and restricted maximum likelihood. Only Gaussian data are considered and two variogram models - linear and exponential. The focus is on kriging predictors and standard errors are obtained by a plug-in method. No standard errors for the parameters of the variogram model are provided. Tang, Schucany, Woodward, and Gunst (2007) propose a parametric spatial bootstrap, and in a simulation study, the coverage of confidence intervals from the proposed method is estimated and compared to an existing spatial bootstrap method proposed by Solow (1985). Variogram models are fit using weighted least squares curve-fitting estimators. Confidence intervals for the mean of the process only are considered.

Here we extend both of these studies by considering a wide class of variogram models - the Matérn family, processes with t-distributed marginals as well as Gaussian and three estimators of the variogram. Our focus is on estimation of variogram parameters and percentiles of the marginal distributions of these estimators.

The paper is organized as follows. Section 2 provides a theoretical background and introduces the variogram estimators and two bootstrap methods. Section 3 describes the simulation study and gives the results and summary of the results. Section 4 describes an application example and conclusions are given in Section 5.

2. Theoretical background

2.1. The variogram and the Matérn family

The variogram of a spatial stochastic process $\{S(s): \ s \epsilon R^2\}$ with mean $\mu(s)$ at location s, is the function

$$V(s, s') = \frac{1}{2}\mathsf{VAR}(S(s) - S(s'))$$

and the covariance function is denoted as $\gamma(s, s')$, which for a stationary isotropic Gaussian process is $\gamma(u)$ where $u = ||s - s'||$. The variance of a stationary process is a constant $\sigma^2 = \gamma(0)$. The correlation function is then defined as $\rho(u) = \gamma(u)/\sigma^2$. In this case the variogram simplifies to $V(u) = \sigma^2(1-\rho(u))$ and can also be defined as $V(u) = 1/2E[(S(s) - S(s-u))^2]$. The Matérn family of correlation functions is given by

$$\rho(u) = [2^{\kappa-1}\Gamma(\kappa)]^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi) \tag{1}$$

in which $K_{\kappa}(\cdot)$ denotes a modified Bessel function of the second kind of order κ, $\phi > 0$ is a scale (range) parameter with the dimension of distance and $\kappa > 0$, is a shape parameter determining the analytic smoothness of the underlying process $S(s)$. We denote a set of geostatistical data by $(s_i, z_i), i = 1, \ldots, n$ where z_i is the measured value associated with location s_i, i.e. $Z_i = S(s_i) + e_i, i = 1, \ldots, n$ where the e_i are mutually independent with mean 0 and variance τ^2 called the nugget effect. Then Z(s) has correlation function

$$\mathrm{Corr}(Z(s), Z(s')) = \begin{cases} 1, & s = s' \\ \sigma^2/(\sigma^2 + \tau^2)\rho(||s - s'||), & s \neq s' \end{cases}$$

where $\rho(\cdot)$ is the correlation function of $S(s)$. In this case σ^2 is called the partial sill while $\sigma^2 + \tau^2$ is called the sill.
In what follows attention is restricted to stationary isotropic spatial processes only. We consider estimators of the nugget, partial sill and range.

2.2. Estimating spatial covariance parameters by curve-fitting the empirical variogram

Two widely used empirical variograms are the Matheron and variogram cloud. We consider here estimation from the variogram cloud only, as Jin and Kelly (2017) recommend. Given any two locations s_i and s_j in R^2, let

$$h_{ij} = s_i - s_j; \ T_{ij} = 1/2(Z(s_i) - Z(s_j))^2$$

then

$$E(T_{ij}) = V(h_{ij}, \theta),$$

with θ denoting the set of covariance parameters and V denoting the variogram. Thus generalized estimating equations (GEE) estimates of V with working independence structure can be calculated as the ordinary least squares estimates in the model

$$T_{ij} = V(h_{ij}, \theta) + \delta_{ij}, \ \delta_{ij} \sim i.i.d.(0, \phi)$$

Schabenberger and Gotway (2005, chap. 4). We note that this is the same as fitting the variogram model by ordinary least squares to the variogram cloud consisting of $\{1/2(Z(s_i) - Z(s_j))^2\}$. Based on the results of Zimmerman and Zimmerman (1991) and Jin and Kelly (2017), weighted least squares curve-fitting was not considered here.

2.3. Estimating spatial covariance parameters by likelihood methods

Here we assume a model

$$Z \sim N(D\beta, \sigma^2 R(\phi) + \tau^2 I)$$

where D is an $n \times p$ matrix of covariates, β is the corresponding vector of regression parameters and R depends on a scalar or vector-valued parameter ϕ. The log-likelihood function is

$$L(\beta, \tau^2, \sigma^2, \phi) = -0.5\{n\log(2\pi) + \log\{|(\sigma^2 R(\phi) + \tau^2 I)|\} + (y - D\beta)^T(\sigma^2 R(\phi) + \tau^2 I)^{-1}(y - D\beta)\},$$

maximization of which yields the maximum likelihood (ML) estimates of the model parameters. A variant of maximum likelihood estimation is restricted maximum likelihood (REML) estimation as described in Diggle and Ribeiro (2007, chap.5). This can reduce the small-sample bias of ML estimates. This is done using the **geoR** (Ribeiro Jr. and Diggle 2001) and **geostatsp** (Brown 2015) packages in R (R Development Core Team 2012) and parameter values are found by numerical optimization using the R function optim. In general the algorithm converged as it was called with default options that were equal to the model parameter values. In theory standard errors may be obtained from the information matrix. However the information matrix was rarely available, requiring inversion of large covariance matrices, and in addition can be unstable as outlined in the Introduction, so variability is assessed by bootstrap methods.

2.4. Estimating spatial covariance parameters by a nonparametric Shapiro-Botha fit to the empirical variogram

Here one firstly computes binned estimated variogram values, denoted by $\hat{\psi}_k = \hat{\psi}(h_k), k = 1, \ldots, K$, where K is the total number of bins. One can use the classical Matheron estimator

$$\hat{\psi}(h_k) = 1/2|N(h_k)|^{-1} \sum_{N(h_k)} (Z(s_j) - Z(s_l))^2$$

where $N(h_k)$ is the set of pairs (s_j, s_l) a lag distance h_k apart. However, this may not yield a non-negative definite estimate of the variogram functions. Shapiro and Botha solve this by considering a certain class of nonparametric isotropic non-negative definite functions $\mathcal{G}$ (Shapiro and Botha 1991), and performing the fit by weighted least squares through quadratic programming. Here this is done using the **npsp** package in R (Fernández-Casal 2016) using the methods of Fernández-Casal, González-Manteiga, and Febrero-Bande (2003). The fit is done using a weighted least squares criterion, finding the function $\bar{g} \epsilon \mathcal{G}$ such that:

$$\sum_{k=1}^{K} w_k[\hat{\psi}_k - \bar{g}(r_k)]^2 = \inf_{g \epsilon \mathcal{G}} \sum_{k=1}^{K} w_k[\hat{\psi}_k - g(h_k)]^2.$$

It is shown by Cressie (1985) that approximately optimal (in a certain statistical sense) weights are given by $w_k = n_k g(h_k)^{-2}$ where n_k is the number of pairs used to estimate the variogram

at the kth lag. The procedure is iterative, with $w_k = 1$ (OLS) used for the first step and with the weights recalculated at each iteration until convergence: There is an option in the **npsp** package to make the fitted model monotone by the use of additional constraints, and thus estimates of the sill and practical range are readily available.

2.5. Bootstrap methods

Proceeding from the estimation of the covariance parameters, nugget, sill and range, bootstrap methods provide estimates of the percentiles of the marginal distributions of the estimators.

Spatial bootstrap (SB)
Direct application of the nonparametric bootstrap method fails to provide valid resamples whenever there is correlation in either time series or spatial data. When this bootstrap is applied to correlated data, it randomizes either the residuals or the observations and destroys the correlation pattern inherent in the joint distribution. Solow (1985) proposes a spatial bootstrap (SB) method to obtain spatial re-samples from the original data. He calculates the predicted spatial error process as

$$\hat{\delta} = \{Z(s_1) - \hat{\mu}(s_1), \ldots, Z(s_n) - \hat{\mu}(s_n)\}$$

where n is the sample size and $\hat{\mu}(s)$ estimates the mean. Processes with constant mean only were considered. The variogram or covariance matrix is estimated from $\hat{\delta}$ and the SB method is based on the Cholesky decomposition given by $\hat{\sum}_{SB} = \hat{L}_{SB}\hat{L}_{SB}^T$ where $\hat{L}$ is a lower triangular $n \times n$ matrix. The SB method then uses the Cholesky decomposition matrix inverse, $\hat{L}^{-1}$ to de-correlate the spatial error sequence

$$(\hat{\epsilon_1}, \hat{\epsilon_2}, \ldots, \hat{\epsilon_n}) \equiv \hat{\epsilon} = \hat{L}_{SB}^{-1}\hat{\delta}$$

and then centers the $\hat{\epsilon}$ to obtain

$$\tilde{\epsilon_i} = \hat{\epsilon_i} - \sum_{j=1}^{n} \hat{\epsilon}_j$$

for $i = 1, \ldots, n$. The de-correlated and centered residuals $\tilde{\epsilon_1}, \tilde{\epsilon_2}, \ldots, \tilde{\epsilon_n}$ are bootstrapped to provide the resampled residuals $\epsilon_{SB}^* = (\epsilon_1^*, \ldots, \epsilon_n^*)$. The SB resample is obtained by transforming to re-correlate the bootstrapped residuals

$$Z_{SB}^* = \hat{\mu} + \hat{L}_{SB}\epsilon_{SB}^*. \tag{2}$$

However, the SB method may not be valid because the predicted spatial error process $\hat{\delta}$ may not retain the same correlation structure as the original spatial data. Confidence intervals based on SB tend to suffer from significant under coverage because the de-correlated residuals are not sufficiently uncorrelated in practice (Tang *et al.* 2007).

Parametric spatial bootstrap (PSB)
Tang *et al.* (2007) introduce a parametric spatial bootstrap (PSB) method. They adapt a parametric approach in Sjöstedt-de Luna and Young (2003) and the spatial bootstrap method in Solow (1985) to obtain valid resamples from a Gaussian process or a lognormal process. This method accounts for the spatial correlation in the data by estimating the correlation structure and then imposing it in the re-samples. The proposed method offers a correct way

to obtain confidence limits for a statistic $\hat{T}$. The parametric spatial bootstrap (PSB) approach begins with a random sample of uncorrelated errors:

$$\epsilon^*_{PSB} = (\epsilon^*_1, \epsilon^*_2, \ldots, \epsilon^*_n) \text{ where } \epsilon^*_j \sim N(0,1) \text{ for } j = 1, \ldots, n.$$

The variogram parameter $\hat{\theta}_{SB}$ and resulting covariance matrix $\hat{\sum}_{SB}$ are estimated from the residuals of the original spatial data Z(s). Next, the spatial re-samples are transformed by

$$Z^*_{PSB} = \hat{\mu} + \hat{L}_{SB}\epsilon^*_{PSB} \tag{3}$$

where $\hat{L}_{SB}$ is the Cholesky decomposition of $\hat{\sum}_{SB}$. The statistic of interest, $\hat{T}^*$, is calculated from Z^*_{PSB}. The above procedures are repeated B times to estimate the sampling distribution of $\hat{T}$.

This PSB method does not obtain the errors ϵ^* by de-correlating the spatial (or spatio-temporal) error process as in the SB algorithm. Instead, the errors are independently generated from a standard normal distribution and transformed. More importantly, because $\hat{\delta}$ may not retain the original spatial (or spatio-temporal) structure, the covariance matrix is estimated directly from the original data. The theoretical foundation is given in Sjöstedt-de Luna and Young (2003).

3. Simulation study.

3.1. Simulation description

A sample of 100 points was taken uniformly distributed over the square $[0,1] \times [0,1]$. There is considerable variance among authors as to what is an adequate sample size for variogram estimation. This is discussed in Jin and Kelly (2017) and as in that study a sample of 100 is considered here. This also reflects the fact that there are now many applications including epidemiology and environmental data, for example, where sample sizes are relatively small.

Data sets were generated with Gaussian data, $Z(s_i), i = 1, \ldots, n$, with constant mean using one of the Matérn variogram models described in Section 2. Models were chosen whereby the practical range was equal to one half the maximum distance between the points as recommended in Jin and Kelly (2017). The practical range is defined as distance u at which the correlation function of $S(s)$ process is 0.05. In each case the practical range was 0.5753. The values of the range parameter ϕ were adjusted so that all variograms had the same practical range. The practical range is approximately 3ϕ, 3.9985ϕ and 6.4171ϕ with $\kappa = 0.5, 1.0$ and 3.0, where κ is the smoothness parameter of the Matérn variogram. $\kappa = 0.5$ corresponds to a process that is not differentiable while $\kappa = 1.0$ or 3.0 reflect increasing smoothness. These κ values are suggested in Diggle and Ribeiro (2007, Section 5.4).

The partial sill (σ^2) was set to 1 and a nugget (τ^2) of 0 or 1/3 was chosen. In all cases the nugget was estimated rather than assumed fixed. The change in nugget:sill ratio expresses different degrees of continuity in the spatial variation of the simulated data.

As Zimmerman and Zimmerman (1991) state, a comprehensive study of variogram estimators requires comparison for several variogram models and true values of the parameters. The simulation study was extended to include a model with t-distributed margins instead of Gaussian. Note that if $Z = (Z_1, \ldots, Z_n)$ is an independent random sample from any distribution with zero mean and unit variance then $Y = HZ$ has zero mean and covariance matrix HH^T,

irrespective of the marginal distribution of the Z_i (Diggle and Ribeiro 2007, chap.5). Thus data with t-distribution (10 df) marginals but a specified Matérn covariance structure were generated and fitted.

For each variogram and set of variogram parameters, 500 independent data sets were generated. For each generated data set, we obtained estimates of τ^2, σ^2 and ϕ, and from these we computed the quantiles of their marginal distributions i.e. we used the simulated distribution to approximate the true marginal distributions. Three estimation methods were considered: REML with Gaussian likelihood, variogram fitting to the variogram cloud by OLS and nonparametric variogram fitting by Shapiro-Botha.

For each generated data set, 500 bootstrap samples were obtained as described in Section 3 using the PSB method and the quantiles of the bootstrap distribution calculated based on B=500 samples. We then found the median of these bootstrap quantiles over simulations. In the case of data with t-distribution marginals, the SB was also applied.

The simulation study plan is shown in Table 1. The Matérn $\times$ 3 indicates three values of κ were considered $\kappa = 0.5, 1.0$ and 3.0.

Table 1: The simulation study plan

Data	Theoretical Variogram	Nugget	Fitted variogram	Variogram fit method	Bootstrap
Gaussian	Matérn $\times$ 3	2 values	true Matérn	REML, OLS	PSB[a][1]
Gaussian	Matérn $\times$ 3	2 values	Exponential	REML, OLS	PSB[2]
Gaussian	Matérn $\times$ 3	2 values	Nonparametric	Shapiro-Botha	PSB, SB[b][3]
t_{10}-marginals	Matérn $\times$ 3	2 values	true Matérn	REML, OLS	PSB,SB
t_{10}-marginals	Matérn $\times$ 3	2 values	Exponential	REML, OLS	PSB,SB
t_{10}-marginals	Matérn $\times$ 3	2 values	Nonparametric	Shapiro-Botha	PSB, SB

[1]Corresponding to simulation results in Figures 1 and 3

[2]Corresponding to simulation results in Figures 2 and 4

[3]Corresponding to simulation results in Figures 5 and 6

[a] PSB:Parametric spatial bootstrap

[b] SB:Nonparametric spatial bootstrap

3.2. Simulation results

In this section we present the results of the simulations, according to the plan of Table 1, in graphical form. Both a parametric and nonparametric bootstrap are considered as described in section 2.5.

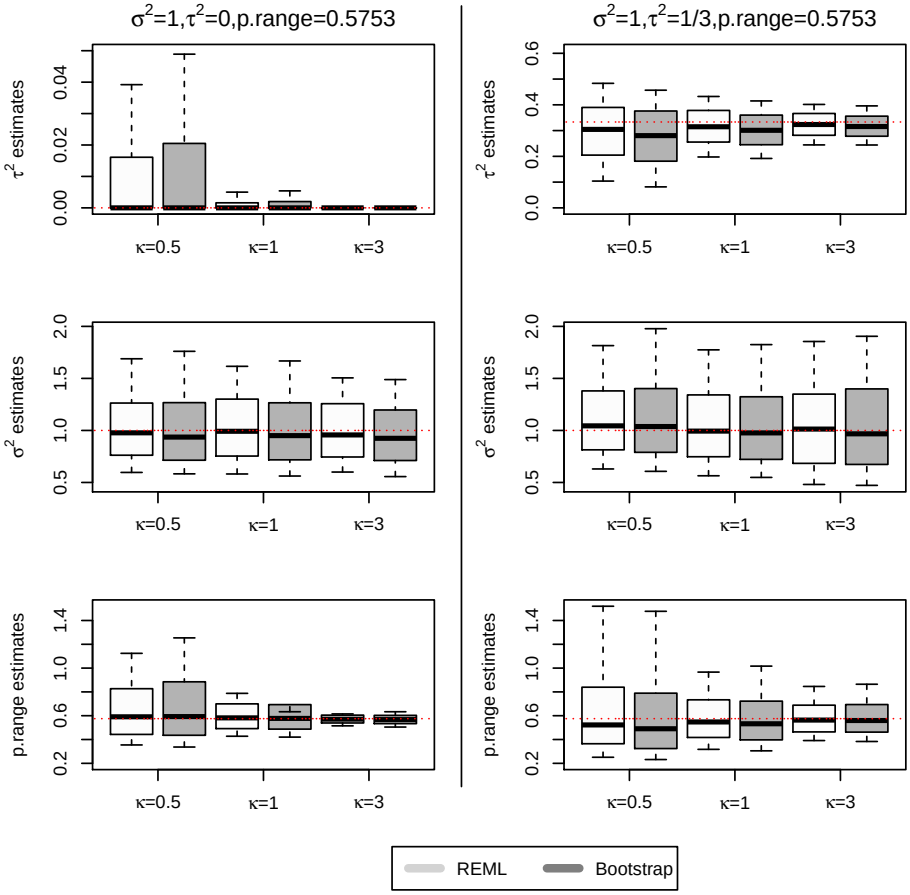

Figure 1: Results of 500 simulations of samples of size N=100 from a Gaussian process with a specified Matérn variogram and associated parameters - nugget τ^2, partial sill σ^2 and practical range. Likelihood fit with REML where variogram with true κ of the Matérn fitted. Boxplots are based on the median, quartiles, 10th and 90th percentiles of the marginal distribution of a parameter estimator (REML) over the 500 simulations and they represent the 'true' parameter distribution. The associated bootstrap boxplots are based on corresponding estimates (each estimate based on 500 bootstrap replications). Bootstrap values are medians of the bootstrap estimates over 500 simulations. Parametric (Gaussian) bootstrap. The dashed line denotes the true parameter value.

- Figure 1 shows the REML estimator performed better for higher values of the shape parameter κ than lower and for nugget $\tau^2 = 0$ rather than 1/3. The median values of the REML estimators were close to theoretical values. The bootstrap boxplots of estimates were very similar to the 'true' REML boxplots ('true' in the sense of based on 500 simulations).

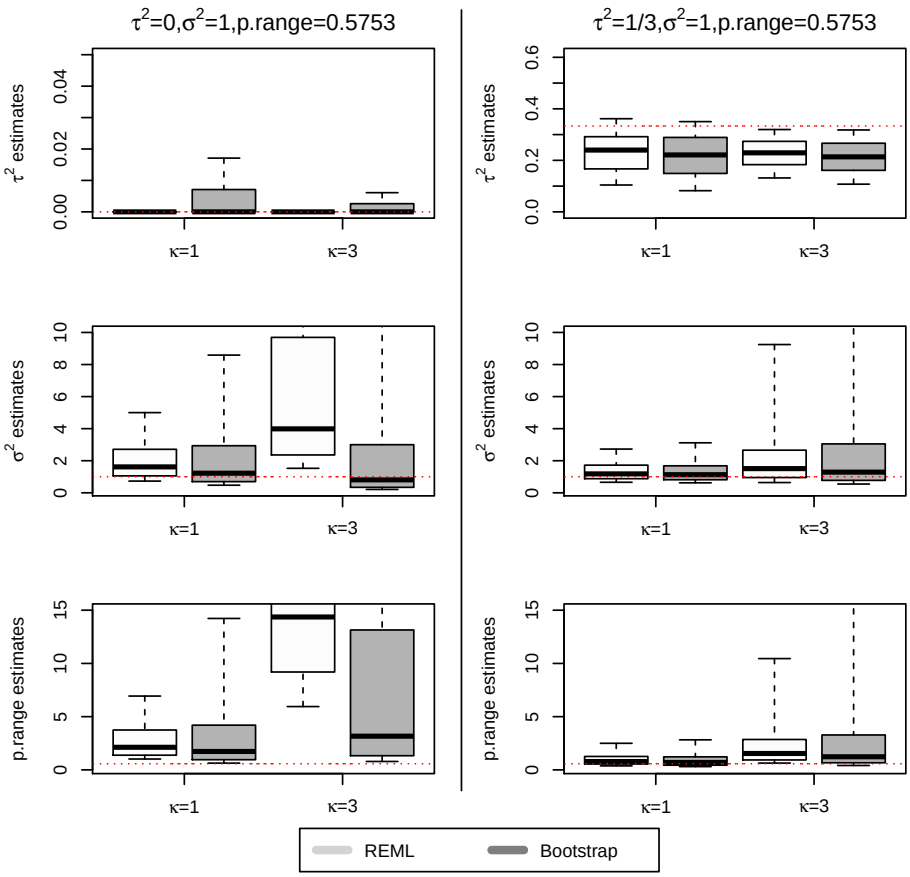

Figure 2: Same as Figure 1, except exponential variogram fitted.

- Figure 2 shows results where the fitted variogram κ is not the same as the theoretical one used to generate the data. Our goal here is to to analyze the mis-specification of the variogram model, as this is commonly an unknown function in real data applications and can be difficult to estimate (Diggle and Ribeiro 2007, chap. 5). A parametric bootstrap with $\kappa = 0.5$ was then calculated also. Both REML and bootstrap underestimated the nugget, while estimates of partial sill and range were too large and far from theoretical values. Results for $\kappa = 1.0$ were now better than for $\kappa = 3.0$ as it was closer to the true κ of 0.5.

In Figures 1- 2 a parametric bootstrap was carried out. One could only expect the nonparametric spatial bootstrap SB to do worse so results are not shown for SB.

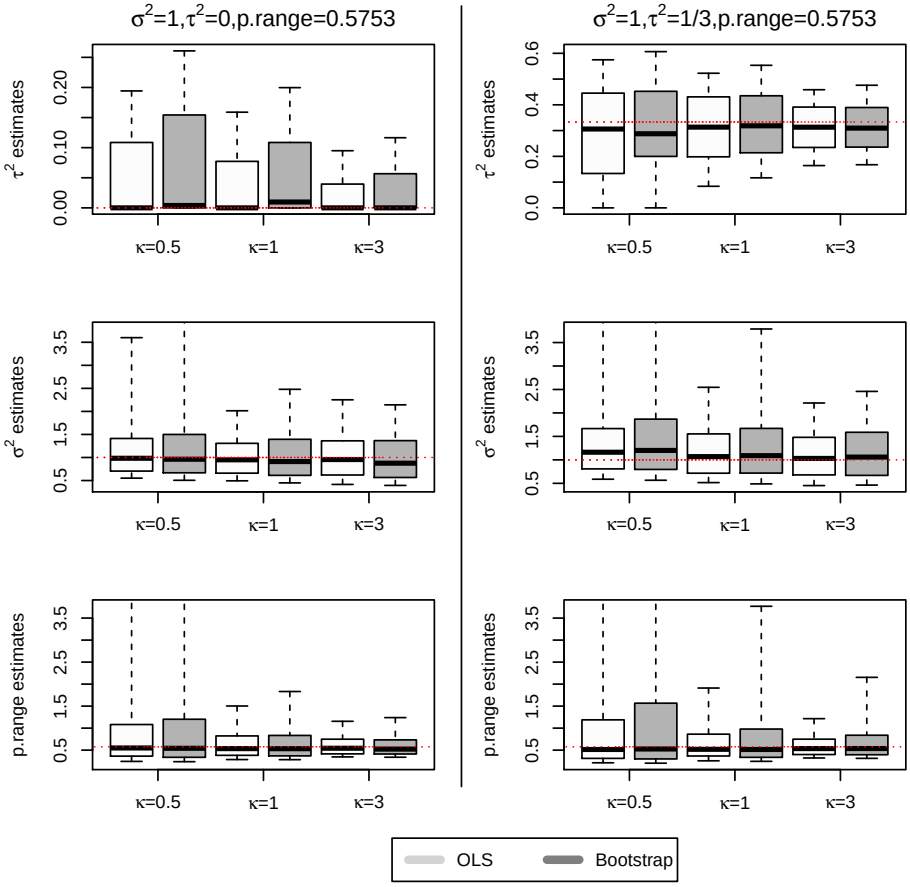

Figure 3: Same as Figure 1, except the estimator is OLS fit to the variogram cloud.

- Figure 3 is the same as Figure 1 but where fitting is to the variogram cloud by ordinary least squares (OLS). The median values of the OLS estimators were smaller than theoretical values for $\tau^2 = 1/3$ and practical range. Results were better for larger values of κ. Results were worse than in Figure 1, the REML fit, as to be expected with inter-quartile ranges (IQR's) larger than REML. The bootstrap boxplots were close to the 'true'.

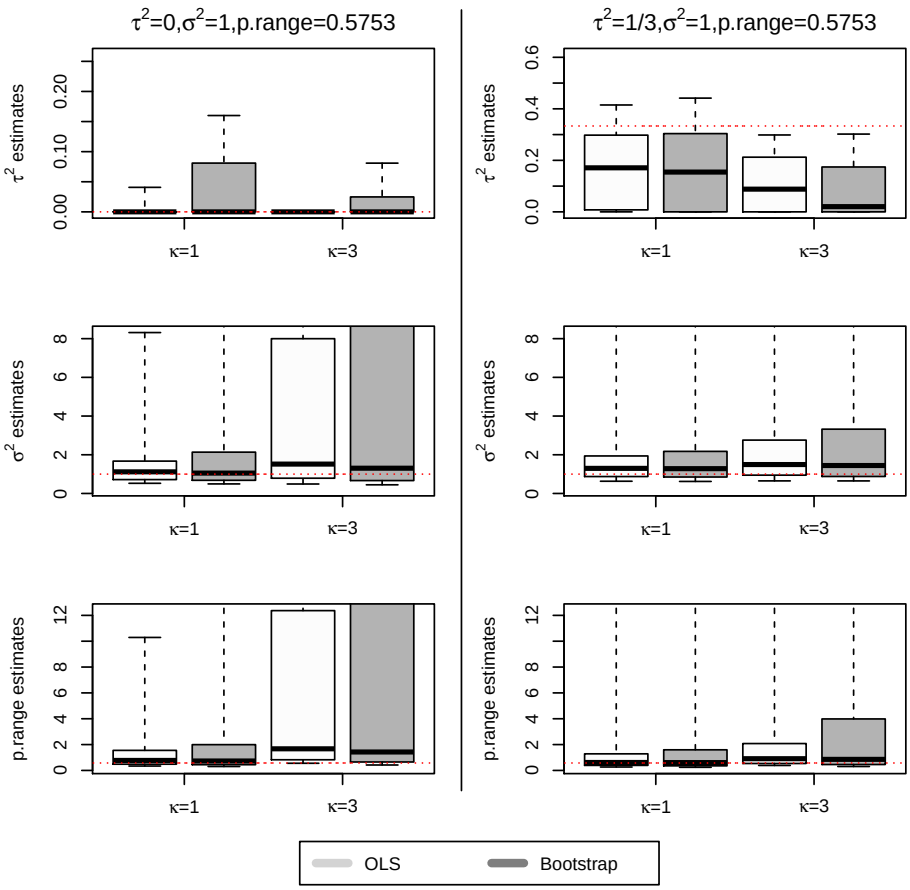

Figure 4: Same as Figure 1, except the estimator is OLS fit to the variogram cloud and exponential variogram fitted.

- Figure 4 is the same as Figure 2 but where fitting is done by OLS. As in Figure 2 results were better for $\kappa = 1.0$ than $\kappa = 3.0$. Estimates were particularly variable for $\kappa = 3.0$. The marginal medians of $\tau^2 = 1/3$ were smaller and those of the partial sill and practical range larger than theoretical values. The bootstrap medians were close to the 'true' but biased downward and the IQR's were larger. Results were generally worse than REML fitting.

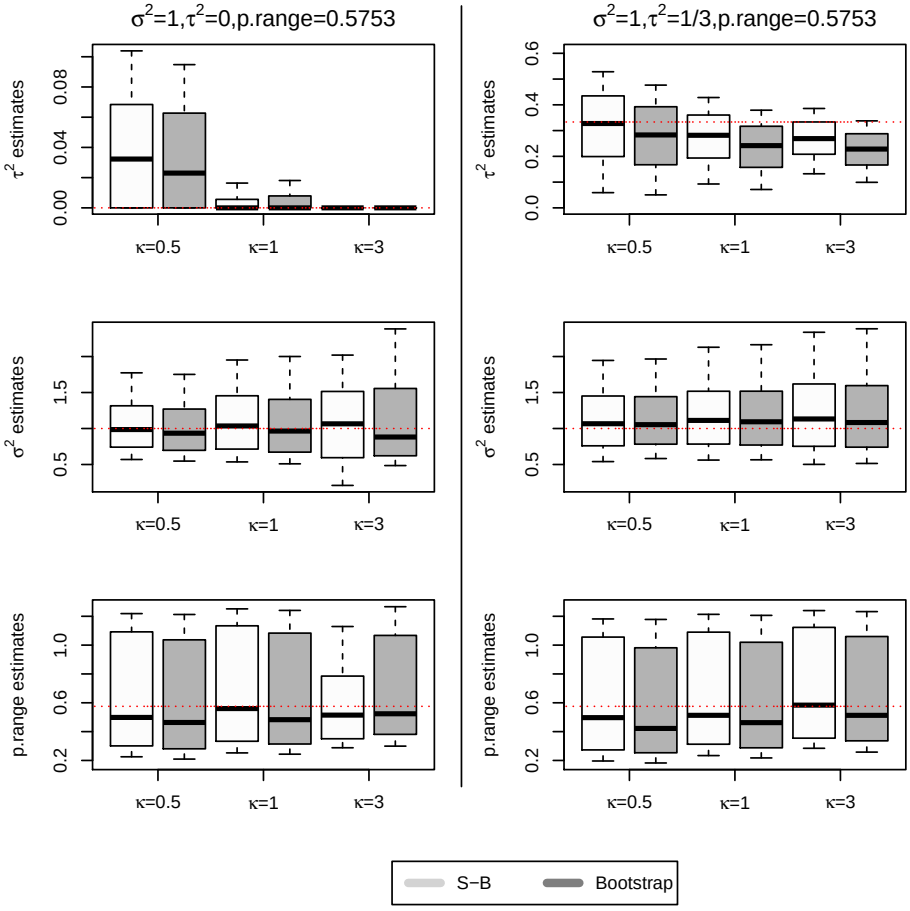

Figure 5: Same as Figure 1, except the estimator is Shapiro-Botha (S-B) fit to the variogram.

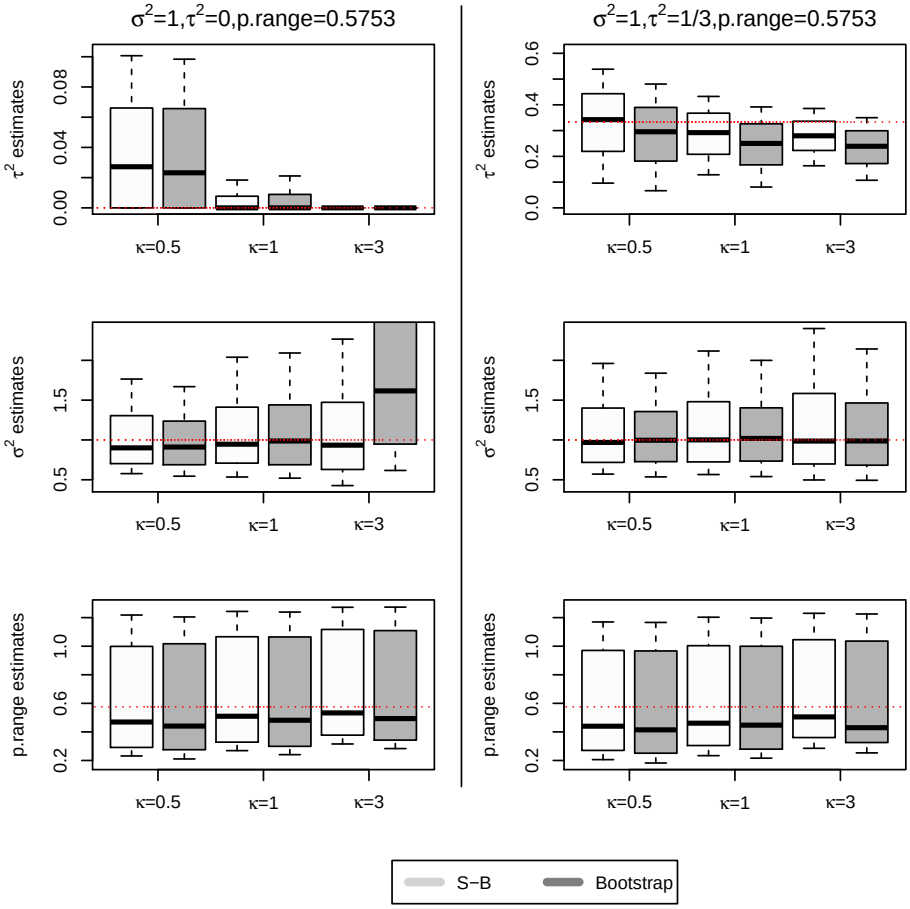

Figure 6: Same as Figure 1, except the estimator is Shapiro-Botha (S-B) fit to the variogram and nonparametric bootstrap.

- Figures 5-6 concern the Shapiro-Botha fit to Gaussian data. The results are very good in that the median values of estimates are close to the theoretical values although estimates of $\tau^2 = 1/3$ were a bit too small and estimates of practical range also too small for $\kappa = 0.5$. The lower quartiles of the marginal distributions were lower and the upper quartiles higher than the corresponding ones in Figure 1 and Figure 3 although results were close generally to those in Figure 3. The bootstrap boxplots were shifted downward from the true for all parameters. Results did not improve as κ increased. Note that Figures 5 and 6 have similar marginal distributions. This is as it should be as the data were generated with the same seed, but the Figures were not amalgamated as the PSB and SB did not always converge for the same re-samples or same simulations and the latter were omitted until 500 simulations each with 500 bootstraps was achieved. This was necessary to make automation possible.

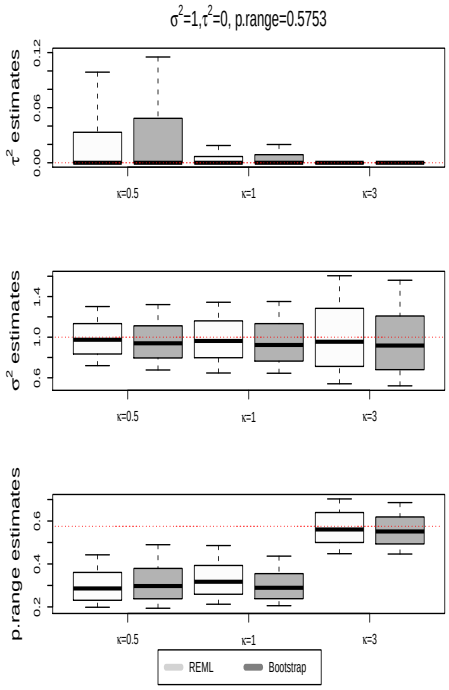

Figure 7: Same as Figure 1, except the estimator is likelihood fit REML with κ estimated.

- REML estimation was repeated for Gaussian data with Matérn variograms where now the parameter κ is estimated. The results are shown in Figure 7. The results for $\kappa = 0.5$ and $\kappa = 1.0$ underestimated the practical range while those for $\kappa = 3.0$ were as good as where κ was assumed known. The parametric bootstrap boxplots were close to the marginal boxplots but slightly below in the case of the partial sill and practical range. OLS estimation where κ is estimated could not be automated. For some bootstrap iterations κ became extremely large and the simulation was aborted.

The foregoing simulations studies done for Gaussian data, and represented in Figures 1-7, were all repeated for t-distribution marginals data with both bootstrap methods SB and PSB being carried out. The results are not included in this manuscript but are available on request. Fitting by REML with Gaussian likelihood the results were not as good as for Gaussian data. IQR were larger for both REML and bootstrap than in Gaussian case. The medians of the marginal distributions of the nuggets, partial sills and practical ranges for $\tau^2 = 1/3$ were too large. The bootstrap estimates were also too large and shifted slightly below REML values, with bootstrap IQR smaller than true REML. Estimates were worse for $\tau^2 = 1/3$ than $\tau^2 = 0.0$. The best results were for $\kappa = 3.0$. Results for SB were as good as for PSB.
The wrong Matérn variogram (i.e. $\kappa = 0.5$) was also fitted to t-distribution data. For REML fit and the parametric bootstrap estimates of nugget are good with the bootstrap percentiles slightly below true REML. Results for REML for $\kappa = 3.0$ in particular are very poor for partial sill and range with the bootstrap estimates reflecting this. Results were quite similar to the corresponding fit to Gaussian data.

For the OLS fit to t-distributed data, median estimates of nugget both OLS and bootstrap are too large with bootstrap quartiles a little below the 'true'. Estimates of partial sill are too high also for both OLS and bootstrap while median values of the practical range sampling distribution are very close to theoretical values. The bootstrap accurately reflected variability of the estimates although the bootstrap quartiles were a little below the 'true' for all parameters. SB results are similar to PSB. When an exponential variogram was fit by OLS results were very poor for $\kappa = 3.0$ as for Gaussian data but the bootstrap quartiles were close to the 'true'. Overall OLS performed worse than REML.

When Shapiro-Botha was fit to t-distributed data, marginal estimates of nugget and partial sill are too large but quite similar to results from OLS. The bootstrap IQR are similar to the true marginal IQR for all estimates but with all bootstrap quartiles below those of true. Results did not improve as κ increased. The SB is similar but slightly worse than PSB. The Shapiro-Botha fit performed worse than REML both for Gaussian data and t-distribution data but was as almost as good as OLS.
In summary,

1. The distributions of all parameter estimators- REML, OLS and Shapiro-Botha, were skewed to the right as is typical for covariance parameters.
2. Results for all methods of estimation were worse when a nugget was present i.e. $\tau^2 = 1/3$ than $\tau^2 = 0.0$.
3. Results were better for larger values of the shape parameter κ for REML and OLS estimators (except in case where wrong κ was fitted).
4. The bootstrap estimates accurately reflected the estimator being bootstrapped although tended to be biased downward. The parametric bootstrap is better than the nonparametric bootstrap.
5. In terms of fitting method, REML (with Gaussian likelihood) outperforms OLS and Shapiro-Botha even for t-distributed data.
6. If the wrong Matérn κ is fitted, REML and OLS perform very poorly while this is not an issue for the Shapiro-Botha method. Similarly, REML performs poorly for small values of κ, where κ is estimated, while again this is not an issue for the Shapiro-Botha method.

All simulations were carried out on a 20 node, 356 core HPC Cluster. The study was limited to 500 simulations and each simulation with 500 bootstrap replications due to the computation time required. Each set of parameter values in Figure 1 takes approximately 105-125 h while those in Figures where fitting is done by OLS take 10-40 h. Therefore, they were run in parallel on a cluster. In contrast, each set of parameter values in Figure 5, for example, took about 30 min.

4. Example

In this section we derive an application of the bootstrap methodology to a real data set con-

cerning bio-monitoring of arsenic pollution in the Central Region of Portugal. The measured variable represents the concentrations in moss samples, in micrograms per gram dry weight. The typical procedure, alternative to the more expensive solution of determining the amount of pollutant directly, is to plant the moss and some time later to collect it, which allows the concentration of arsenic (and other heavy metals) to be measured. Further details can be found in Martins, Figueira, Sousa, and Sérgio (2012). The data set was collected in 2006 and it can be represented by $(s_i, Z(s_i)); i = 1, \ldots, n$; with $n = 98$ and $Z(s_i)$ identifying the log-transformed concentration of arsenic (As) at location s_i. As in García-Soidán, Menezes, and Rubiños (2014), we adopted the log-transformation and afterwards, there were still three gross outliers, which were replaced by the average of the remaining values from that year's survey. This lead to a more symmetric distribution. The data are displayed in Figure 8.
We first estimate a deterministic model for E[Z(s)] $= \mu(s)$ with $s = (x, y) \epsilon D \subset R^2$ identifying the easting and northing planar co-ordinate pair in the central region of Portugal zone. We then assume that the random process Z(s) can be modeled as:

$$Z(s) = \mu(s) + S(s) + \epsilon$$

where $S(s) : s \epsilon D$ is a zero-mean strictly stationary random process, ϵ is the nugget effect and $\mu(s) = \beta_0 + \beta_1 \times y$, (y being the northing co-ordinate of the location s) as in García-Soidán *et al.* (2014). The co-ordinates were divided by a 1000. Assuming a Matérn variogram where the shape parameter κ was unknown, the variogram was estimated by REML, curve-fitting to the variogram using OLS and by the Shapiro-Botha estimator of the variogram. For OLS and Shapiro-Botha, $\mu(s)$ was first estimated by a regression model and the observations converted to residuals before calculating the empirical variogram. For REML a Gaussian likelihood was assumed. The variability of estimates was assessed using the PSB, with $\hat{\mu}(s)$ replacing $\hat{\mu}$ in equation (3) and a new $\hat{\mu}(s)$ estimated and residuals obtained at each PSB iteration.
The results of the fit where the 1st and 3rd quartiles of distributions are estimated by the

Table 2: Mosses arsenic data. A Matérn variogram was assumed where the shape parameter κ was unknown. Three estimators of variogram parameters were computed: REML, curve-fitting using OLS and Shapiro-Botha as described in section 2. In each case a parametric bootstrap (PSB) was carried out and 1st and 3rd quartiles of the estimated parameter's distributions estimated.

Parameter	REML (Q[1])	Shapiro-Botha (Q)	OLS (Q)
β_0	-31.52 (-37.52,-24.352)	-37.77(-43.52,-31.57)	-37.77(-43.49,-31.64)
β_1	0.6945 (0.5344,0.8249)	0.8336(0.6942,0.9587)	0.8336(0.6970,0.9588)
nugget τ^2	0.5350 (0.5533,0.8331)	0.5607 (0.5123,0.6854)	0.8012 (0.6762,0.8425)
partial sill σ^2	0.2945 (0.0,0.2889)	0.1697 (0.0,0.3494)	0.0 (0.0,0.2158)
practical range	0.4622 (0.0001,0.2126)	0.1461(0.0,3.3047)	1.7028 (1.7028,2.1638)
shape κ	4.6968 (4.4411,5.0964)	NA	∞ Gaussian

[1]bootstrap first and third quartiles

PSB are shown in Table 2. κ is also estimated in each PSB iteration. REML and Shapiro-Botha estimates show reasonably good agreement with the estimates just inside/outside the corresponding bootstrap quartiles. The OLS results show weaker agreement, with an estimated $\kappa = \infty$ and therefore a Gaussian variogram was fitted. The estimated sill was zero

indicating little spatial structure. Diggle and Ribeiro (2007, chap. 5) note that in their experience REML is more sensitive than maximum likelihood (ML) to the chosen model for $\mu(s)$. Therefore the data were also fitted by ML. Results were almost identical to the REML fit apart from the estimated κ was found to be ∞ i.e. a Gaussian variogram was fitted and the practical range at 0.2193 (0.0,0.1147) was somewhat smaller than for the REML estimate. All fits indicated that the nugget is at least two times the partial sill and thus the error process is mostly white noise with spatial structure only up to a short distance. The trend surface accounts for a substantial proportion of the spatial variation in the data. The simulation results of Figure 7 indicate that the PSB is reliable here since the estimated κ is > 3. The nonparametric bootstrap (SB) gave similar results to PSB but with larger IQR's.
Figure 9 displays the fitted variograms.

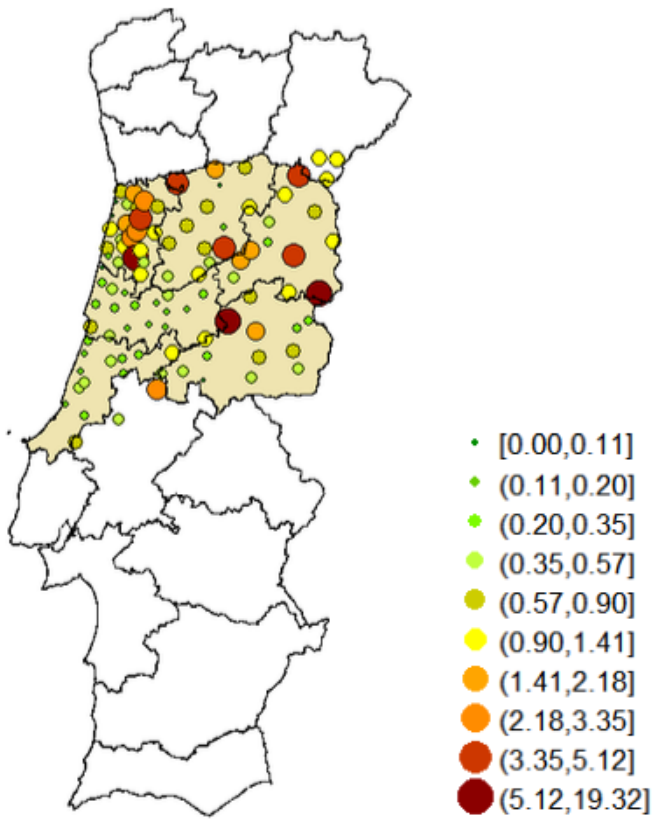

Figure 8: Circle plot for arsenic data from Portugal

5. Discussion

In summary, our results show that REML estimation is superior to curve-fitting to the variogram by OLS for Gaussian random fields underlying a Matérn dependence structure. This is also true for random fields with t-distributed marginals. Likelihood methods are favoured by Diggle and Ribeiro (2007, chap.5), as they point out that the variogram is only one of

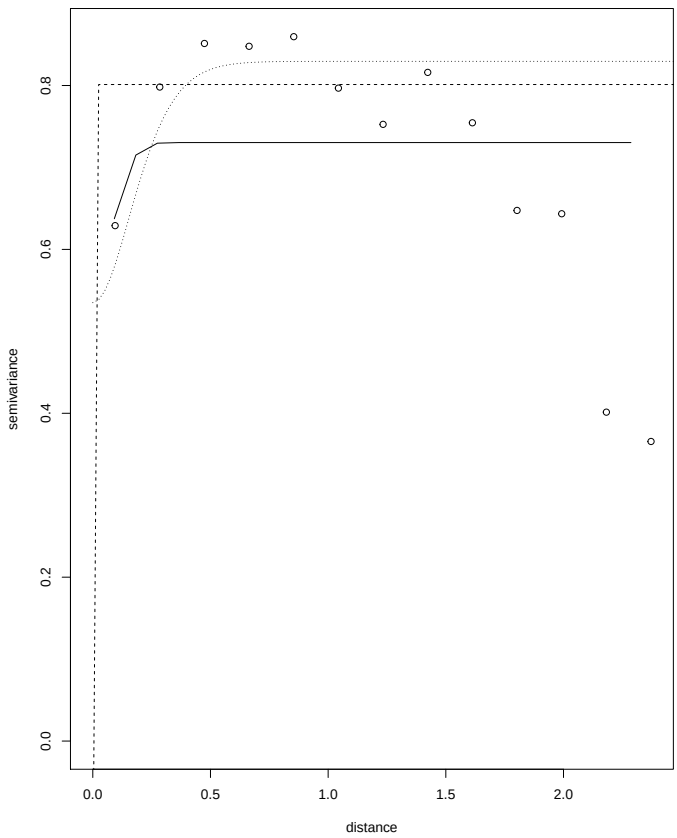

Figure 9: Empirical variogram for the arsenic data (circles), with the the REML fitted variogram (dotted line), the OLS fitted variogram (dashed line) and Shapiro-Botha fitted variogram (solid line)

a number of possible summaries of data while likelihood methods use an explicit model for the original data. In addition, Diggle and Ribeiro (2007, chap. 5) recommend the use of REML estimation over ML estimation, as do Zimmerman and Zimmerman (1991), so ML is not considered here.

The nonparametric Shapiro-Botha estimator does not perform as well as REML or OLS for these data. However, when the fitted variogram is not the same as the theoretical one used to generate the data, then both REML and OLS perform poorly while this does not affect the Shapiro-Botha estimator as it does not require specification of a variogram form. In this sense the Shapiro-Botha estimator is robust. Diggle and Ribeiro (2007, chap.5) state the value of κ in the Matérn correlation function is often poorly identified and recommend choosing κ from a discrete set. This is not possible in an automated simulation study. Our results show that if κ is large i.e $\kappa \geq 3$ then it is well estimated by REML but results for REML are quite poor when estimating smaller values of κ.

We note that the curve-fitting to the variogram method by OLS is computationally less expensive than REML estimation. Curve-fitting by OLS took approximately one-third of the time required by REML. Curve fitting by Shapiro-Botha was even faster than that by OLS.

The PSB performed well in that it approximated the quartiles of the sampling distributions of covariance parameter estimates closely, although it tended to be biased downward. Olea and Pardo-Igúzquiza (2011) argue strongly that bootstrap percentile confidence intervals, that do not require distributional assumptions for construction, provide an achieved coverage similar to the nominal coverage. The latter cannot be achieved by symmetrical confidence intervals based on the standard error, regardless if the standard error is estimated from a parametric equation or from bootstrap. They used the SB method to provide confidence intervals for the parameters of a Gaussian process with a spherical variogram fitted using weighted least squares. Just one set of parameter values were examined. Similar to that study, but estimating different models and parameters, we conclude that distributions for estimated variogram parameters tend to be positively skewed. Here it is also clear that bootstrap percentiles can provide confidence intervals for parameters for all three estimators. They become less useful however, if the distribution of the estimator is extremely skewed as happens when the 'wrong' variogram is fitted in the case of REML and OLS.

Moreover our results show, unlike Tang *et al.* (2007), that SB performs almost as well as PSB. For the OLS it performed even better than PSB for t-distributed data and in the case of the Shapiro-Botha estimator performs better overall. However, Tang *et al.* (2007) consider estimation of the mean only and do not consider covariance parameters. O'Rourke and Kelly (2016) employ the PSB in a spatio-temporal variogram setting and all bootstrap iterations converged and gave reasonable values indicating the reliability of the PSB.

The reader should note our results are limited to linear geostatistical models. This is because it is difficult to simulate spatial binary data with a specified covariance structure - see for example Lin and Clayton (2005).

Our simulation is limited to one sample size, as we are primarily concerned with small sample behaviour, and regular grids. Varying grid sizes and densities could be considered while keeping the practical range constant, to enable effects of denseness of the data and sample size to be examined. This was beyond the scope of this study. Neither was it practicable here to consider non-rectangular designs, proved to be more effective designs in Müller and Zimmerman (1999) in the case of spherical variogram forms. However, the results obtained in this work could be used to inform and establish a more extensive study.

References

Bonat W, Ribeiro P (2016). "Practical Likelihood Analysis for Spatial Generalized Linear Mixed Models." *Environmetrics*, **27**, 83–89.

Brown PE (2015). "Model-Based Geostatistics the Easy Way." *Journal of Statistical Software*, **63**(12), 1–24. URL http://www.jstatsoft.org/v63/i12/.

Cressie N (1985). "Fitting variogram models by weighted least squares." *Mathematical Geology*, **17**, 563–86.

Diggle P, Ribeiro P (2007). *Model-based Geostatistics*. Springer, New York.

Fernández-Casal R (2016). ***npsp**: Nonparametric Spatial (Geo)statistics*. R package version 0.5-3, URL http://CRAN.R-project.org/package=npsp.

Fernández-Casal R, González-Manteiga W, Febrero-Bande M (2003). "Flexible Spatio-Temporal Stationary Variogram Models." *Statistics and Computing*, **13**, 127–36.

García-Soidán P, Menezes R, Rubiños Ó (2014). "Bootstrap Approaches for Spatial Data." *Stochastic Environmental Research and Risk Assessment*, **28**, 1207–19.

Jin R, Kelly G (2017). "A Comparison of Sampling Grids, Cut Off Distance and Type Of Residuals in Parametric Variogram Estimation." *Communications in Statistics- Computation and Simulation*.

Lin PS, Clayton M (2005). "Analysis of Binary Spatial Data by Quasi-likelihood Estimating Equations." *Annals of Statistics*, **33**, 542–55.

Martins A, Figueira R, Sousa A, Sérgio C (2012). "Spatio-Temporal Patterns of Cu Contamination in Mosses using Geostatistical Estimation." *Environmental Pollution*, **170**, 276–84.

Müller W, Zimmerman W (1999). "Optimal Designs for Variogram Estimation." *Environmetrics*, **10**, 23–37.

Olea R, Pardo-Igúzquiza E (2011). "Generalized Bootstrap Method for Assessment of Uncertainty in Semivariogram Inference." *Mathematical Geosciences*, **43**, 203–28.

O'Rourke S, Kelly G (2016). "Spatio-temporal Dependence of Irish Sitka Spruce Stands using GIS-based Sum-metric Models." *Forest Science*, **66**, 851–60.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org, URL http://www.R-project.org.

Ribeiro Jr P, Diggle P (2001). "**geoR**: A Package for Geostatistical Analysis." *R-NEWS*, **1**(2), 15–18. ISSN 1609-3631. URL http://cran.R-project.org/doc/Rnews.

Schabenberger O, Gotway C (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman&Hall/CRC.

Shapiro A, Botha J (1991). "Variogram Fitting with a General Class of Conditionally Nonnegative Definite Functions." *Computational Statistics and Data Analysis*, **11**, 87–96.

Sjöstedt-de Luna S, Young A (2003). "The Bootstrap and Kriging Prediction Intervals." *Scandinavian Journal of Statistics*, **30**, 175–92.

Solow A (1985). "Bootstrapping Correlated Data." *Journal of the International Association for Mathematical Geology*, **17**, 769–75.

Tang L, Schucany W, Woodward W, Gunst R (2007). "A Parametric Spatial Bootstrap." *Technical report*, Department of Statistical Science, Southern Methodist University, Dallas, TX 75275, USA.

Warnes J, Ripley B (1987). "Problems with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes." *Biometrika*, **74**, 640–2.

Zimmerman D, Zimmerman M (1991). "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors." *Technometrics*, **33**, 77–91.

Affiliation:

Gabrielle E. Kelly
School of Mathematics and Statistics, University College Dublin
Belfield, Dublin 4, Ireland
E-mail: gabrielle.kelly@ucd.ie
URL: https://maths.ucd.ie/people/kelly_g

Journal of Environmental Statistics http://www.jenvstat.org
Volume 8, Issue 6 Submitted: 2017-12-13
September 2018 Accepted: 2018-09-08