A comparison of approaches for valid variogram achievement

Raquel Menezes[1,*], Pilar Garcia-Soidán[2] and Manuel Febrero-Bande[3,**]

[1] Department of Mathematics for Science and Technology, University of Minho, 4800-058 Guimarães, Portugal.
[2] Department of Statistics and O.R., University of Vigo, Campus A Xunqueira, Pontevedra 36005, Spain.
[3] Department of Statistics and O.R., University of Santiago de Compostela, Campus Sur, Santiago de Compostela 15771, Spain.

Summary

Variogram estimation is a major issue for statistical inference of spatially correlated random variables. Most natural empirical estimators of the variogram cannot be used for this purpose, as they do not achieve the conditional negative-definite property. Typically, this problem's resolution is split into three stages: *empirical variogram estimation*; *valid model selection*; and *model fitting*. To accomplish these tasks, there are several different approaches strongly defended by their authors. Our work's main purpose was to identify these approaches and compare them based on a numerical study, covering different kind of spatial dependence situations. The comparisons are based on the integrated squared errors of the resulting valid estimators. Additionally, we propose an easily implementable empirical method to compare the main features of the estimated variogram function.

Keywords: Spatial dependence, Empirical variogram, Valid model, Fitting criteria, Non-parametric estimation.

1 Motivation

It is well known that variogram analysis provides a useful tool for summarizing spatial data and that it may be used to measure spatial dependence between samples. However, its main contribution is related to inference procedures, when used to estimate the value of the spatial variable at an unsampled location. The approach differs from classical regression in that local features can affect the solution. Bearing in mind that some measurements in the vicinity of the point investigated, or sometimes elsewhere, are more closely related to the unknown true value than others, a natural approximation to consider is a *weighted mean*.

The estimation of a variogram plays a decisive role, as it is commonly used to find the optimal values of the weights. This is indeed the strategy used by the popular *kriging* methods (see, e.g., Stein (1998)).

Suppose $\{Z(s) : s \in D \subset I\!R^d\}$ is a spatial random process, where D is a bounded region with positive d-dimensional volume. This process is intrinsically stationary if its first moment is constant and the variance of the difference between two variables is a function of the difference between their locations:

(i) $E[Z(s)] = \mu(s) = \mu \quad \forall\ s \in D;$
(ii) $Var[Z(s_i) - Z(s_j)] = 2\gamma(s_i - s_j) = 2\gamma(h), \quad \forall\ s_i, s_j \in D.$

The function $2\gamma(.)$ defines the variogram function and $\gamma(.)$ is termed the semivariogram.

In theory, when the lag distance h is zero the semivariogram should also be zero. In practice, however, it can be significantly different from zero, possibly reflecting local effects or sampling error. This non-zero value is coined *nugget effect*. Another important characteristic of the semivariogram deals with the fact that, as the separation distance h increases, this function eventually approaches a constant value, known as the *sill*. When a semivariogram has a sill, it means that there is a distance beyond which the correlation between variables is zero; this distance is called the *range*.

---

* Corresponding author. *E-mail address*: rmenezes@mct.uminho.pt. Partially supported by Grant ref.5.3/N/189.015/01 from PRODEP.

** Partially supported by Grants BFM2002-03213 and PGIDIT03PXIC20702PN from Ministerio de Ciencia y Tecnologia, Spain in collaboration with Foundation for the Regional European Development and Xunta de Galicia, Spain.

The first proposal for a variogram estimator for a stationary process is due to Matheron (1963). This estimator is based on the method of moments and it is given by

$$2\widehat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2 \tag{1}$$

where $N(h) = \{(s_i, s_j) : s_i - s_j = h,\ h \in I\!R^d\}$ and $|N(h)|$ is the total number of pairs in $N(h)$. Matheron's estimator is unbiased, however it presents some drawbacks such as being badly affected by atypical values due to the squared term in the summand of (1).

Cressie and Hawkins (1980) have minimized this weakness, by working with square-root absolute differences and, under a Gaussianity assumption, have produced the estimator

$$2\overline{\gamma}(h) = \frac{\left\{\frac{1}{|N(h)|} \sum_{N(h)} |Z(s_i) - Z(s_j)|^{\frac{1}{2}}\right\}^4}{0.457 + \frac{0.494}{|N(h)|}} \tag{2}$$

Robustness to outliers is normally considered an important characteristic for any estimator. In this regard, some other robust empirical estimators have been proposed in addition to (2). For instance, Genton (1998*a*) proposes a variogram estimator based on the highly robust scale estimator of Rousseeuw and Croux (1992,1993), denoted below by Q_{N_h}. The theory of M-estimators of scale is used to derive robustness properties. The resulting estimator is $2\widehat{\gamma}(h) = (Q_{N_h})^2, h \in I\!R^d$ where

$$Q_{N_h} = 2.2191\{|(Z(s_i + h) - Z(s_i)) - (Z(s_j + h) - Z(s_j))| : i < j\}_{(k)}$$

The k value is equal to $\binom{[N_h/2]+1}{2}$, where $[N_h/2]$ denotes the integer part of $N_h/2$, and is used to compute the k^{th} quantile of all sorted $|.|$ values. One may note that Q_{N_h} does not rely on any location knowledge and is thus said to be location-free, in contrast to Matheron's estimator.

Unfortunately, all these estimators are advised not to be used for inference and prediction. They may fail the conditionally negative-definite property which may lead to absurd negative values for the mean square prediction errors, as proved in Cressie (1993). They are thus classified as non valid estimators.

In our search for valid estimation procedures and in our posterior comparison study, we have decided to focus on the isotropy case, with $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h,\ h \in I\!R\}$. This restriction can be relaxed by assuming geometric anisotropy or by fitting a different semivariogram in each of several directions.

2 Traditional three stages

A common approach to achieving a valid variogram estimator is to approximate an empirical variogram by some theoretical model known as valid. The idea is to select, within the families of valid variograms, a function which captures the underlying spatial dependence of the available data. Traditionally, these type of approaches are accomplished through three distinct stages:

1. *Compute an empirical* parametric or non-parametric semivariogram (typically non valid);
2. *Choose a theoretical model* among the family of valid parametric or non-parametric semivariograms;
3. Estimate the semivariogram by *fitting the theoretical model to the empirical* semivariogram.

Some authors prefer to group these three stages into two, called *variogram estimation* and *variogram fitting* stages, the latter performing the stages 2 and 3 simultaneously (see, e.g., Cressie (1993)). In contrast, we argue that, when possible, three separate stages allow a better classification of the existing approaches. The output of stage 2 is a vague valid candidate and its complete specification is only obtained from stage 3.

Before giving details about the complete approaches that were examined, we make some generic comments on each of the previously listed stages. We shall point out some references, if we think they introduce a relevant idea for the implementation of these tasks.

2.1 Stage 1 – Empirical variogram estimation

The word "empirical" means *based on observation or experiment*. The estimation of the empirical variogram always, unsurprisingly, begins with the observed data, whichever estimator selected. Examples include those estimators introduced in section 1, when slightly modified to suit isotropy requirements.

A non-parametric approach for the empirical estimation of γ can also be considered. Nadaraya-Watson's kernel estimator uses the weighted average

$$\widehat{\gamma}_g(h) = \frac{\sum_i \sum_j w_{ij}[Z(s_i) - Z(s_j)]^2}{\sum_i \sum_j w_{ij}} \quad \text{where} \quad w_{ij} = K\left(\frac{h - \|s_i - s_j\|}{g}\right) \quad (3)$$

K is a symmetric, zero-mean and bounded density function, with compact support $[-C, C]$, and g is a bandwidth parameter. In Garcia-Soidán, Febrero-Bande and Gonzalez-Manteiga (2004) several properties of this estimator are studied and an asymptotically optimal bandwidth parameter obtained.

With Matheron's estimator, only pairs (s_i, s_j) such that $\|s_i - s_j\| = h$ are used to compute a specific $\widehat{\gamma}(h)$. If data is not regularly spaced, Matheron's estimator can be adapted to consider a tolerance region around h. For kernel estimator, all pairs are used and they are all given a particular weight: the weights are at their maximum when the distance between two points is close to h, and zero values if $\left|\frac{h-\|s_i-s_j\|}{g}\right| > C \Longleftrightarrow \|s_i - s_j\| \notin [h - gC, h + gC]$.

2.2 Stage 2 – Valid model selection

The aim of this stage is to find a negative-definite function which, as a measure of spatial dependence, is in some sense closest to the sample data. The notion of "in some sense" is considered in detail at stage 3. At this stage we are concerned with questions such as the choice of exponential vs spherical families, or parametric vs non-parametric estimators. The most common methods used to pick a valid family are based on graphical tools, with model selection reduced to approximating the estimated variogram curve by one from the valid family. In recent years, some alternatives have been suggested.

Maglione and Diblasi (2001) propose a statistical method for choosing a valid model for the variogram. The test statistic for their approach is based on smoothed random variables which reflect the underlying spatial variation. The distribution of this test statistic, which is a ratio of quadratic forms, can be approximated by a shifted chi-square distribution and is used to verify the *distance* between the underlying model for the variogram and the one in the null hypothesis.

Gorsich and Genton (2000) propose a method for the selection of a valid parametric model via the derivative of a non-parametric variogram estimate, without assuming a prior model. The basic idea of their proposal is to avoid choosing among valid parametric variogram models, as they may look similar, and to choose instead among their derivatives, as they are often quite different. These derivatives should be compared with the one obtained from the non-parametric variogram estimate based on the spectral representation of positive definite functions.

The first non-parametric approach to the selection of a valid model appeared in Shapiro and Botha (1991). Key result behind these approaches is *Bochner's theorem*, which states that a covariance function $c(h)$ is positive definite iff it has the following form (Cressie (1993)):

$$c(h) = \int_0^\infty \Omega_d(ht) F(dt)$$

where $\Omega_d(x) = (2/x)^{(d-2)/2}\,\Gamma(d/2) J_{(d-2)/2}(x)$ is a basis for functions in $I\!R^d$, $F(dt)$ is a nondecreasing bounded function, Γ is the gamma function, and J_v is the Bessel function of the first kind of order v.

This theorem, together with the relation $\gamma(h) = c(0) - c(h)$, are employed to represent the family of non-parametric valid variograms. To allow the numeric evaluation of γ, $F(t)$ should be considered a step function with a finite number m of positive jumps $p_1, ..., p_m$ at points $t_1, ..., t_m$. A valid non-parametric estimator can then be given by

$$\widehat{\gamma}(h_i) = \Sigma_{j=1}^{m}\ p_j\,(1 - \Omega_d(h_i t_j)) \tag{4}$$

2.3 Stage 3 – Model fitting

The classical goodness-of-fit criteria may be used to complete the specification of the final variogram. Possible choices are the *minimum variance or norm quadratic unbiased* (MIVQU or MINQU), the *maximum likelihood* (ML) and the *least squares* (LS) criteria. Those based on LS are known as being less limited in scope and by requiring the fewest distributional assumptions about $Z(s)$. In matrix notation, a LS minimizing problem is written as

$$min\left\{(\widetilde{\gamma} - \gamma_\theta)^T\, W^{-1}\,(\widetilde{\gamma} - \gamma_\theta)\right\}$$

where $\widetilde{\gamma}$ identifies an empirical estimator and γ_θ identifies a valid model whose exact form is known except for the unknown parameter θ. The W matrix is a weight matrix. If W is an identity matrix, then one has the *ordinary least squares* (OLS) criterion. If $W = V$ where V is the variance-covariance matrix, then one has the *generalized least squares* (GLS) criterion. If matrix V is reduced to its diagonal, then the resulting criterion is called *weighted least squares* (WLS). In Cressie (1985), he considers WLS as a pragmatic compromise between GLS efficiency and OLS simplicity and suggests $w_j = \frac{|N(h_j)|}{\gamma(h_j)^2}$, where the unknown γ should be approximated by γ_θ through an iterated procedure.

Genton (1998*b*) refuses to accept WLS as the solution for GLS complexity and proposes an explicit formula for the covariance structure V, calling the resulting method **GLSE**. The basic idea is to obtain a generic covariance structure by using, iteratively, the correlation structure of Matheron's estimator in the independent case.

2.4 Existing combinations of the previous stages

Next, we shall introduce some existing complete approaches to reach our target: a valid variogram estimator. All of them result from distinct combinations of previous stages.

We shall begin with a mandatory reference, **Zimmerman and Zimmerman (1991)**, where seven different approaches are compared through a Monte

Carlo simulation study. This comparative study, in spite of being considerably exhaustive, is somehow restricted in scope as it only involves parametric techniques. In fact, these seven approaches are mainly distinguishable by their third stage, being four LS-based, two ML-based and one using a modified MIVQU.

Shapiro and Botha (1991) are indeed the pioneers on selecting a valid model in a non-parametric space. They combine the Matheron's estimator at stage 1, a broad class of *permissible variograms* at stage 2 and at the last stage, a WLS fitting criterion where the optimization problem is reduced to a quadratic programming problem. Following Christakos (1984), they define $f(h),\ h \in I\!R^d$ as a *permissible* semivariogram function, if it is continuous (except possibly at the origin), $f(h) = f(-h),\ f(h) \geq 0$ for all h, and $-f(h)$ is conditionally nonnegative definite. The resulting valid variogram estimator then fulfills equation (4). This approach was evaluated by Cherry, Banfield and Quimby (1996), where they conclude that this "non-parametric method is faster, easier to use and more objective than parametric methods".

Gribov, Krivoruchko and Ver Hoef (2000) suggest a new method of computing the empirical variogram of Matheron. The squared differences $[Z(s_i) - Z(s_j)]^2$ are binned into K distinct bins, and point estimations of the semivariogram at K points are obtained. The complicating issue is how to best bin the data. They introduce the notion of logarithmic increases in the size of tolerance regions against the traditional fixed size. This new concept allows better results in estimation near the origin. They also propose to use a kernel method to assign, within a given bin, weighted values depending on how close a value is to the center of the bin. This requires fewer elements per bin than the recommended minimum of 30 pairs of the classic guideline of Journel and Huijbregts (1978), as well as the weights' presence minimizes a possibly existing unequal distribution of lags.

In respects to the *model fitting* stage, they propose a modified WLS[1] procedure. They split this stage into two steps. At step 1, typically with two iterations, they consider logarithmic lag sizes. At step 2, a default lag size obtained from the range estimate in step 1 is used instead.

The last approach included in our survey is the one proposed by **Garcia-Soidán et al. (2004)**. These authors introduce a non-parametric technique in an additional stage. They propose the usage of the non-parametric empirical estimator given by equation (3) together with the *permissible* function of Shapiro and Botha. The empirical $\widehat{\gamma}_g(h)$ and the theoretical curve are fitted through a re-iterated WLS criterion. The former is shown to have desirable properties, such as asymptotically unbiasedness and consistency.

[1]This algorithm is included into the Geostatistical Analyst extension to GIS ArcInfo/ArcView8.1 (Krivoruchko, 1999).

Table 1: Taxonomy of existing approaches for valid $\widehat{\gamma}$ achievement. Bold identifies those approaches selected for the comparative study.

Approaches	Stage 1	Stage 2	Stage 3
Zimmerman	Matheron(1)	P model	OLS
and	Cressie-Haw.(2)	P model	WLS
Zimmerman	**Matheron**	**P model**	**WLS**
(1991)	Matheron	P model	WLS-Delfiner(1974)
	—	P model	ML
	—	**P model**	**REML**
	Matheron	P model	OLS+MIVQU
Shapiro and Botha (1991)	Matheron	NP function(4)	WLS
Gribov *et al.* (2000)	Matheron-modif.	P model	WLS-modified
Garcia-Soidán *et al.* (2004)	**NW kernel**(3)	**NP function**(4)	**WLS**

One should have in mind that an important issue of kernel estimation is the selection of the bandwidth parameter, g. These authors address the problem by asymptotically minimizing the *mean square error* (MSE) or the *mean integrated square error* (MISE), in order to derive the local and the global bandwidth, respectively. Both expressions involve the unknown function $\gamma(h)$. For the purpose of the bandwidth derivation, a simple parametric approach, like the first one presented by Zimmerman and Zimmerman (1991) (see Table 1), may be used to estimate $\gamma(h)$. This isolated parametric estimation can even be improved by being incorporated into an iterated non-parametric procedure.

Outside the boundary, the bias of the Nadaraya-Watson estimator (3) is of the order g^2; however, the latter order amounts to g for distances h close to 0. Then, proceeding as in Kyung-Joon and Shucany (1998), we may denote by $\hat{\gamma}_{q,g}(h)$ the estimator obtained by substituting a boundary kernel H_q for the symmetric one K in (3), where $q = \min\left\{hg^{-1}, C\right\}$ and

$$H_q(z) = \frac{K(z) - rL(z)}{1-r}, \quad z \in [-C, q]$$

where $r = c_{1,K}c_{0,L}(c_{0,K}c_{1,L})^{-1} \neq 1$ and $c_{i,G} = \int_{-C}^{q} z^i G(z)$. This particular selection of the boundary kernel H_q produces a semivariogram estimator $\hat{\gamma}_{q,g}(h)$ that makes it negligible the term of order g in the bias and preserves the same convergence orders for all $h > 0$, as shown in Garcia-Soidán et al. (2004).

3 Simulation study

In order to analyze the performance of the previous approaches for valid $\widehat{\gamma}$ achievement, simulations of spatial data in $I\!R^2$ were carried out for different kind of dependence situations. We considered the exponential and the spherical semivariogram models. Additionally, the wave model was also considered, because of its atypical irregular behaviour.

- Exponential model: $\gamma_e(h,\theta) = \theta_0 + \theta_1 \left[1 - exp(-h/\theta_2)\right],\ h \neq 0$
- Spher. model: $\gamma_s(h,\theta) = \begin{cases} \theta_0 + \theta_1 \left[\frac{3}{2}(h/\theta_2) - \frac{1}{2}\left(h/\theta_2\right)^3\right] & , 0 < h \leq \theta_2 \\ \theta_0 + \theta_1 & , h > \theta_2 \end{cases}$
- Wave model: $\gamma_w(h,\theta) = \theta_0 + \theta_1 \left[1 - \theta_2\ sin(h/\theta_2)/h\right],\ h \neq 0$

In all cases, a uniform distribution on $[0,1] \times [0,1]$ was assumed for spatial locations $s_i = (x_i, y_i),\ i = 1, .., n$, where n represents the sample size. Several data sets were generated with Gaussian data, $Z(s_i),\ i = 1, .., n$, using one of the above semivariogram models. These models' parameters were chosen in such a way that the corresponding curves were comparable according to their *radius of influence* (or range). We have then fixed the values for the nugget θ_0 and θ_1, being 0.25 and 5.0, respectively. The third parameter was the one chosen depending on the model: exponential, $\theta_2 = 0.167$; spherical, $\theta_2 = 0.5$; and wave, $\theta_2 = 0.113$. With this selection, the theoretical semivariograms have a sill of 5.25 and a range (referred to the minimum value for which the semivariogram reaches either the sill or 95% of the sill, in case that the range is not finite) of 0.5. More precisely, the wave model oscillates around the sill value and, consequently, the 0.5 value identifies the global maximum of the corresponding semivariogram function.

3.1 Comparing empirical estimators

The aim of our first exercise was to compare the three main empirical estimators used at stage 1 of the approaches included in Table 1, given in expressions (1), (2) and (3). For data generation, we took sample size $n = 200$ and we started by selecting the exponential model.

Unusual estimated values were obtained by estimators (1) and (2) for the largest lags. Additionally, some of them did not have the recommended minimum of 30 pairs. Therefore, in posterior simulations, we have decided to only consider the first 55% of lags. One may note that this guideline still is less *conservative* than the one proposed by Journel and Huijbregts (1978), specifying that the largest used lag, h_k, should be less than or equal to half of the largest existent lag. As non-parametric estimation requires more lags

than those empirically obtained, we have also decided to consider a larger number of lags, equally spaced, within interval $[\ min(h_k)\ ,\ 0.55 * max(h_k)\ ]$. Following these considerations, Figure 1 shows the obtained data, as well as two more graphs assuming the spherical and the wave models for data generation. All graphics included in this paper use the following notation: lines are used to represent a valid estimator; and isolated symbols, e.g. small squares, are used for empirical estimates.

Figure 1 demonstrates the behaviour of the estimator when one sample is considered, although it will depend strongly on the sample variability. For this reason, we include a second study where 100 independent samples are considered. For each one, the *integrated squared error* (ISE) between each of the three empirical estimators and the theoretical semivariogram, given by

$$ISE = \int \left[\widehat{\gamma}(h) - \gamma(h)\right]^2 dh,$$

was approximated numerically through the trapezoid rule. $\widehat{\gamma}(h)$ represents an empirical estimator and $\gamma(h)$ represents the theoretical curve. This simulation was repeated for the previous models: exponential, spherical and wave.

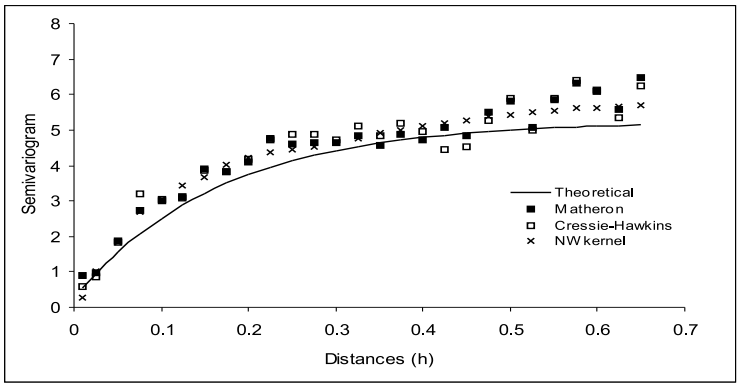
The results are summarized in the boxplot in Figure 2, through the quartiles of the found ISE values. If one compares the *median*'s values associated to the three estimators, then the best performance is clearly achieved by the non-parametric estimator, using the Nadaraya-Watson kernel. Another advantage of this non-parametric estimator is that it is a continuous function. In contrast, estimators (1) and (2) propose point values of the semivariogram for given distances h, making them discontinuous. Most analyses requires knowledge about estimations in a continuous range of $\gamma(h)$.

We conclude by bringing attention to the different orders of magnitude of the ISE values for each theoretical model, being the lowest values associated to the exponential curve and the largest ones to the wave curve.

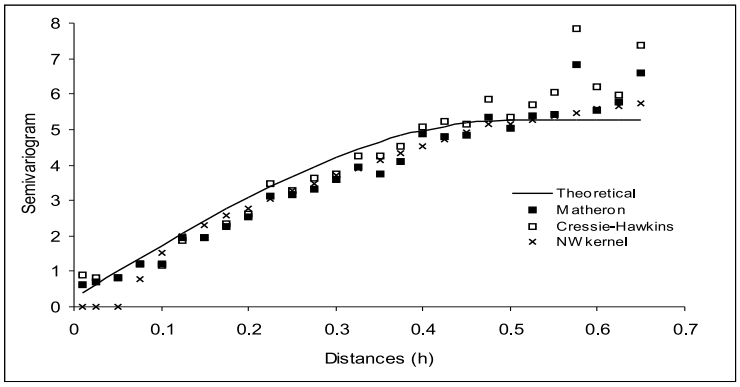
3.2 Comparing complete approaches

We highlight three approaches (marked in bold) from Table 1, which we consider the most representative of the existing alternatives. For two of the approaches, a valid model is chosen within the space of parametric families. They are identified as the *parametric approaches* (P), one of which uses WLS as fitting criterion and the other REML. The third approach, introduced by Garcia-Soidán et al. (2004), is referred to as the *non-parametric approach* (NP).

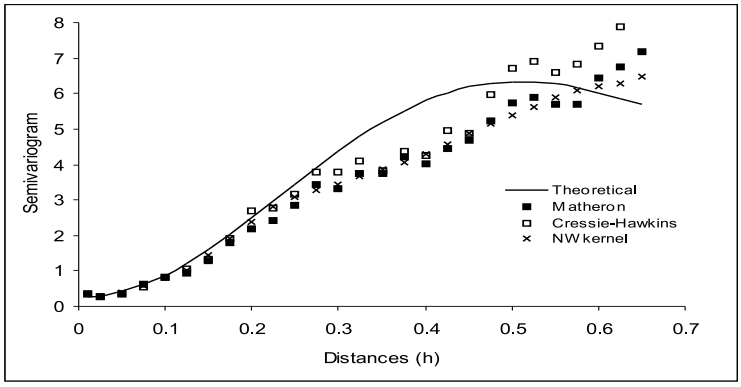
The superior results of the Nadaraya-Watson kernel estimator, when compared to the Matheron's estimator, led us not to include Shapiro and Botha (1991) in our numerical study. Gribov et al. (2000) was also excluded as,



a) Exponential

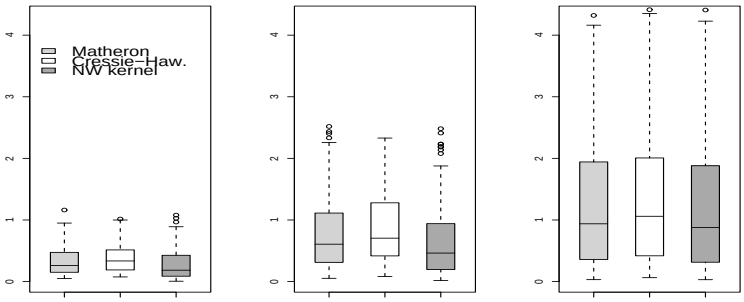


b) Spherical



c) Wave

Figure 1: Three empirical estimators and the associated theoretical curve. Data simulated with three distinct models. Sample size equals 200.



a) Exponential b) Spherical c) Wave

Figure 2: Boxplot of the evaluated ISE from three empirical estimators, using data simulated from three distinct models. The simulation consisted of 100 replications, each with a sample size of 200.

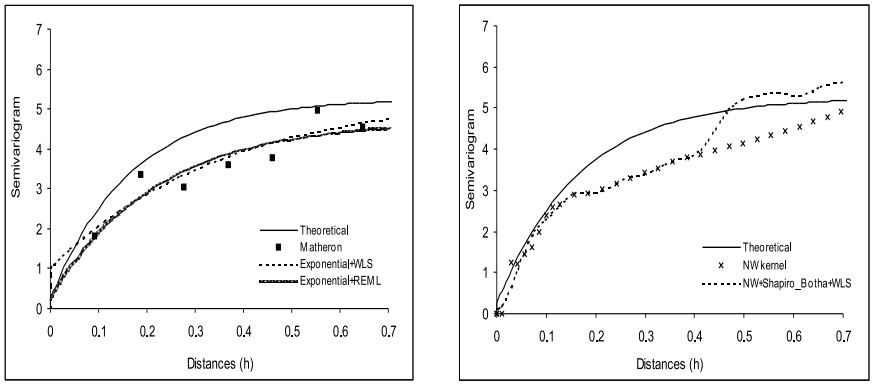
under isotropy, their main contribution is mainly reduced to the usage of weights within a given bin. In this case, the kernel estimator does not differ much from their proposal and may be indeed a better choice.

Under the NP approach, we preferred to asymptotically minimize the MSE to derive a local bandwidth parameter. For this purpose, the symmetric Epanechnikov kernel was employed. Additionally, as the bandwidth derivation needs itself an estimation of the semivariogram, the available WLS parametric estimation was used for this purpose. Near the semivariogram endpoint 0, a specific asymmetric boundary kernel was constructed from the Epanechnikov kernel and the quartic kernel.

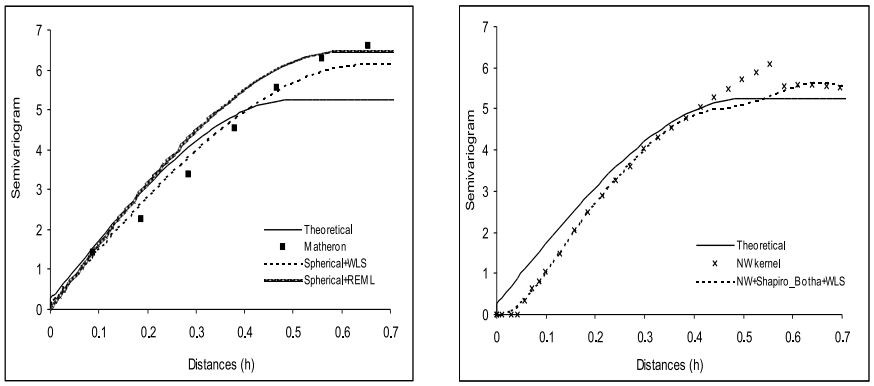
For the implementation of the REML fitting criterion, we used the geoR library from R, which provides several functions for geostatistical analysis as explained in Ribeiro Jr and Diggle (2001). Excluding this particular case, we used Fortran to implement our numerical study.

Figure 3 shows an example of results obtained with the three selected approaches, when using each of our theoretical models for data simulation and a sample size $n = 50$. The correct specification of the theoretical variogram is considered: if data is generated with a given model then this same model is the one elected at stage 2. On the left side, are the valid estimators resulting from the P approaches. On the right side, the final valid estimator is given by a *permissible* function of Shapiro and Botha when fitted through WLS to the NW kernel's estimations.

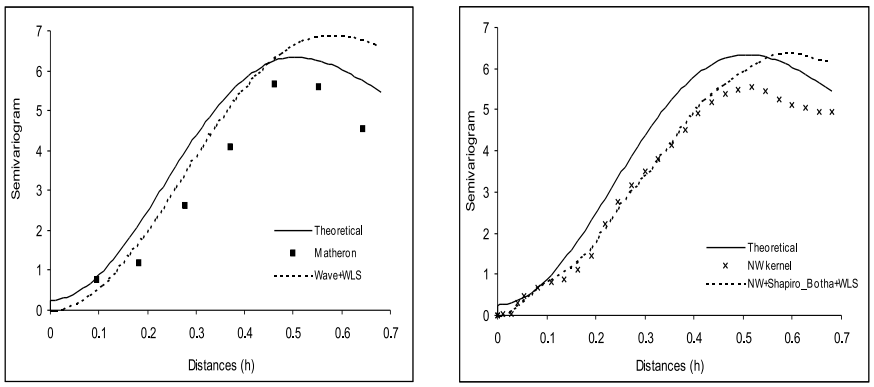
The wave semivariogram, see e.g. Figure 3c, causes problems in achieving a valid estimator through REML fitting criterion, because the Choleski facto-



a) Data simulated with an exponential model



b) Data simulated with a spherical model



c) Data simulated with a wave model

Figure 3: Approaches to achieving a valid $\widehat{\gamma}(h)$: the 2 parametric approaches are on the left and the non-parametric approach is on the right. Data simulated with three distinct models. Sample size equals 50.

Theoretical model		n = 50 Parametric model EXP	SPH	WAV	n = 200 Parametric model EXP	SPH	WAV
E	Mean($\text{ISE}_{P_{wls}}$)	0.61	**0.58**	0.79	0.21	0.20	0.38
X	Mean($\text{ISE}_{P_{reml}}$)	**0.54**	0.63	1.00	0.37	0.77	1.42 *
P	Mean(ISE_{NP})	0.61	0.63	**0.74**	**0.19**	**0.19**	**0.22**
S	Mean($\text{ISE}_{P_{wls}}$)	**0.96**	**0.85**	0.95	0.34	0.26	0.44
P	Mean($\text{ISE}_{P_{reml}}$)	1.10	1.10	1.12	0.63	0.69	1.60 *
H	Mean(ISE_{NP})	1.14	0.87	**0.88**	**0.32**	**0.25**	**0.24**
W	Mean($\text{ISE}_{P_{wls}}$)	0.84	0.99	1.35	0.89	0.68	0.70
A	Mean($\text{ISE}_{P_{reml}}$)	2.22	2.98	N/A	3.31	4.77	N/A
V	Mean(ISE_{NP})	**0.62**	**0.95**	**1.20**	**0.62**	**0.67**	**0.55**

Table 2: Mean values of the obtained ISE for the three approaches (P_{wls}, P_{reml} and NP) chosen to achieve a valid $\widehat{\gamma}(h)$. Data simulated with three theoretical models. Total number of replicas is 100 and each sample size is either 50 or 200. For each combination of one theoretical and one parametric model, bold identifies the lowest mean when comparing the three approaches. * For these two cases about 80 replicas were used, as for the remaining replicas the variance-covariance matrix was non-positive definite, not allowing the Choleski factorization.

rization of the variance-covariance matrix was required and this matrix was typically non-positive definite.

Next, we will cover different kinds of spatial dependence situations. The data were generated using any of our three elected models, exponential, spherical or wave, and we supposed that any of these models could be chosen as the best guess by the user at stage 2. The idea is to analyze how the wrong selection of a parametric model affects our approaches. It is worth noting that even the NP approach is expected to be somehow affected by this error, through the procedure of bandwidth derivation.

Table 2 shows the mean values of the evaluated ISE for 100 independent data sets. The errors associated with the P approaches, WLS and REML fitting criteria, are denoted by $ISE_{P_{wls}}$ and $ISE_{P_{reml}}$ respectively, and the errors associated with the NP approach are denoted by ISE_{NP}. Two different sample sizes, $n = 50$ and $n = 200$, were considered.

The NP approach is, in general, preferable, as it provides smallest mean of

ISE values in 55.5% of the cases considered for samples of size $n = 50$ and 100% for $n = 200$. More precisely, the results achieved for the NP approach exceed, at least 5%, those obtained for the second best approach in 44.4% and 77.8% of the total cases for $n = 50$ and $n = 200$, respectively.

The P_{wls} approach seems competitive with NP (i.e. not more than 5% inferior or even superior) in 44.4% of the observed cases for samples of size $n = 50$ and 22.2% for $n = 200$. As regards the P_{reml} approach, it should be avoided for larger samples, as well as when the wave model is involved on the procedure for valid $\widehat{\gamma}(h)$ achievement. The P_{reml} approach presents the best behaviour when the exponential model is correctly specified and $n = 50$.

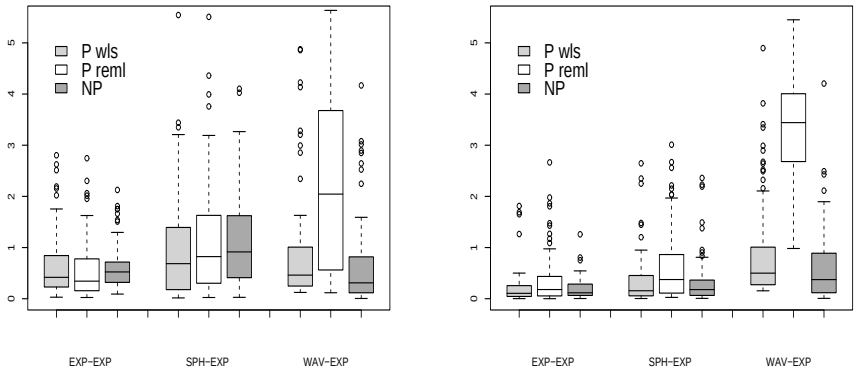
The boxplot in Figure 4 shows more detailed information about previous ISE values, for one particular situation: the exponential curve was elected as the parametric model. This illustrates a likely situation as this family is one of the most popular, making it a strong candidate for election at stage 2. This boxplot contains three different groups of boxes: the first one, labelled EXP-EXP, stands for data simulated with an exponential model, whereas the second, SPH-EXP, and third, WAV-EXP, represent two cases of wrong specification, as data was simulated with a spherical and wave model, respectively.

In this boxplot, the NP approach shows the lower dispersion, measured in terms of the interquartile range, even when the median value of its evaluated ISE is worst. Another interesting conclusion is that the larger median values and the larger interquartile ranges are normally associated to the smaller sample size, i.e. $n = 50$. The exception is the P_{reml} approach, as its median value is degraded by a large sample size.

From the boxplot, it is also evident that the NP approach is the preferred choice in the presence of the wave model. Otherwise, one of the two P approaches might be acceptable.

The last comparison of estimates included in our simulation study involves important features typically associated with the semivariogram function and introduced in section 1: *nugget*, *sill* and *range*. Table 3 summarizes the median values (P_{50}) and the mean square errors (MSE) of their corresponding estimators, comparing the outcome results from P_{wls}, P_{reml} and NP approaches. The correct specification of the theoretical model was always considered.

The nugget effect's estimator is given by $\widehat{\theta}_0$, for the P approaches, and by $\widehat{\gamma}(0)$, otherwise. These estimates should be compared with the theoretical value of 0.25. In respect to the remaining features of $\widehat{\gamma}(h)$, sill and range, we proposed an empirical method extended to the family of non-parametric valid semivariograms. With our proposed method, the derivative of $\widehat{\gamma}(h)$ is



a) Sample size $n = 50$ b) Sample size $n = 200$

Figure 4: Boxplot of the evaluated ISE from the three approaches (P_{wls}, P_{reml} and NP) chosen to achieve a valid $\widehat{\gamma}(h)$. Data was simulated with exponential, spherical and wave models, but the exponential model was estimated.

used to estimate a sill approximation and range.

Under a wave model, we compare the global maximum of $\widehat{\gamma}(h)$ obtained from the NP approach, against $\widehat{\theta}_0 + (1 - (\sqrt{2}\pi)^{-1}\sin(\sqrt{2}\pi))\ \widehat{\theta}_1$ obtained from the P approaches. The corresponding range is defined as $\sqrt{2}\pi\widehat{\theta}_2$. Under the exponential and the spherical models, the range's estimators are $3\widehat{\theta}_2$ and $\widehat{\theta}_2$, respectively. All three models share a theoretical range of 0.5. For these last models, the estimated sill approximation is specified as the maximum of $\widehat{\gamma}(h)$ or 95% of this value, when considering a finite number of lags. The P approaches use $\widehat{\theta}_0 + 0.95\widehat{\theta}_1$, for the exponential model, and $\widehat{\theta}_0 + \widehat{\theta}_1$, for the spherical one.

In terms of MSE values, the NP approach offers the best estimators for the semivariogram's features, as it provides lowest values in 77.8% of the total evaluated MSEs. The P_{wls} approach always presents the worst MSE values. In terms of median values, however, the best results are not necessarily associated with the NP approach. More precisely, the nugget effect seems to be under-estimated when a spherical or a wave model is used. As well as, this same approach seems to over-estimate the sill approximation when sample size is equal to 200. An aspect also worth mentioning is that, overall, the performance of each estimator improves as sample size increases.

A final remark about the numerical study is related to the computational cost of the three approaches chosen to achieve a valid estimator. The CPU

Model	n	Approach	$\widehat{Nugget}$		$\widehat{Sill Approx}$		$\widehat{Range Approx}$	
			P_{50}	MSE	P_{50}	MSE	P_{50}	MSE
EXP	50	P_{wls}	0.31	0.71	5.58	7.74	0.57	0.98
		P_{reml}	*0.30*	0.29	5.15	4.36	*0.55*	0.20
		NP	0.16	**0.06**	*5.15*	**3.93**	0.39	**0.02**
	200	P_{wls}	0.40	0.58	*5.08*	8.76	*0.49*	1.15
		P_{reml}	0.20	0.03	4.74	**2.60**	0.47	0.06
		NP	*0.20*	**0.02**	7.87	9.31	0.52	**0.01**
SPH	50	P_{wls}	0.23	0.29	5.65	8.46	*0.52*	0.12
		P_{reml}	*0.23*	**0.07**	5.38	7.81	0.54	0.07
		NP	0.00	0.23	*5.30*	**5.70**	0.44	**0.01**
	200	P_{wls}	0.60	0.34	*5.12*	3.33	0.38	0.05
		P_{reml}	*0.32*	**0.04**	4.98	**3.27**	0.33	0.04
		NP	0.17	0.04	7.40	8.41	*0.51*	**0.01**
WAV	50	P_{wls}	*0.37*	0.15	6.64	7.10	*0.49*	0.02
		NP	0.00	**0.05**	*6.30*	**5.14**	0.48	**0.01**
	200	P_{wls}	*0.35*	0.46	*6.43*	8.67	*0.50*	0.03
		NP	0.08	**0.03**	7.24	**6.28**	0.49	**0.01**

Table 3: Summary of the main features of the estimated semivariogram: nugget effect, sill approximation and corresponding range. Data simulated with three theoretical models. 100 replications and each sample size is 50 or 200. Bold identifies the lowest MSE when comparing the chosen approaches, while italic identifies the P_{50} value closest to the theoretical value.

execution times[2] were recorded for each sample, without considering data simulation but just the time needed to implement all existing stages. The results are summarized below

	n = 50	n = 200
P_{wls}	1.3 s	1.5 s
P_{reml}	3.0 s	$\approx$ 30 s
NP	$\approx$ 30 s	$\approx$ 30 s

The lowest computational cost was achieved by the P_{wls} approach, being around 1.3 and 1.5 seconds for $n = 50$ and $n = 200$, respectively. The cost for the P_{reml} approach was around 3 seconds for $n = 50$, being at least 10 to 15 times greater for $n = 200$. With respect to the NP approach, we have

[2]The CPU times, in seconds, were obtained on an Intel Pentium III 850 MHz.

registered CPU times from 27 to 36 seconds for $n = 50$, being the lowest values associated to the spherical data and the greatest to the exponential data. These costs have only shown a slight increase when we moved to sample sizes of $n = 200$. Bear in mind that the heavy costs obtained for the NP approach are usually justified by the optimal bandwidth derivation.

3.3 Concluding remarks

The problem of estimation of the variogram can be analyzed in practice from several points of view. If the aim is just to obtain an approximation of the dependence structure of the spatial data, then the classical and the Nadaraya-Watson kernel provide good estimators that behave better than the robust estimator proposed by Cressie and Hawkins, using as a term of comparison the values estimated for the median and interquartile range of the ISE; however, the robust estimator reduces the range of variation of the ISE.

If we focus on the problem of spatial prediction, we modify the variogram estimators to obtain valid variograms; otherwise, negative mean squared prediction errors may be achieved. From the different alternatives discussed, the valid kernel estimation (referred to as the NP approach) has the best performance for large sample sizes in terms of the values estimated for the ISE, regardless of the parametric model that is considered. In this respect, it is surprising that fitting the correct parametric family does not produce a better fit than the non-parametric method. The misspecification of the parametric family has a second order effect on the kernel estimator, since it affects estimation of values associated to the bandwidth parameter. On the other hand, when considering typical features associated with the variogram (nugget, sill and range), we conclude that the valid kernel estimation provides lowest values of the MSE, although the P approaches prove competitive in the estimation of the corresponding median values.

In general, the results presented here show that a valid semivariogram estimator obtained from a NP approach is a good alternative to those valid estimators obtained from the classic parametric approaches. The NP approach has the additional advantage of avoiding problems associated with using the wrong parametric model, which can occur in many conventional approaches. These advantages become even more evident if sample data underlies an atypical spatial dependence, like the one from the wave model. However, one must be prepared to pay an extra computational cost over the cost associated to a simple P approach like the one that fits a valid model to some empirical estimations through the WLS criterion.

The P approach using REML as fitting criterion is only able to compete with the other methods in the presence of small datasets and, simultaneously,

observed data does not follow a wave-type structure.

Acknowledgements

We thank Patrick Brown and two anonymous referees for many comments that improved earlier versions of the paper. The first author is also grateful to Jonathan Tawn and Paulo Ribeiro for the valuable hints and stimulating discussions.

References

Cherry, S., Banfield, J. and Quimby, W. (1996), 'An evaluation of a non-parametric method of estimating semivariograms of isotropic spatial processes', *Journal of Applied Statistics* **v.17**, 563–586.

Christakos, G. (1984), 'On the problem of permissible covariance and variogram models', *Water Resources Res.* **20**, 251–265.

Cressie, N. (1985), 'Fitting variogram models by weighted least squares', *Journal of Int. Association for Mathematical Geology* **17, n.5**, 563–586.

Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley and Sons Inc., New York.

Cressie, N. and Hawkins, D. (1980), 'Robust estimation of the variogram', *Journal of Int. Association for Mathematical Geology* **12, n.2**, 115–125.

Garcia-Soidán, P., Febrero-Bande, M. and Gonzalez-Manteiga, W. (2004), 'Nonparametric kernel estimation of an isotropic variogram', *J. Statist. Plann. Inference* **121**, 65–92.

Genton, M. (1998*a*), 'Highly robust variogram estimation', *Journal of Int. Association for Mathematical Geology* **30, n.2**, 213–221.

Genton, M. (1998*b*), 'Variogram fitting by generalized least squares using an explicit formula for the covariance structure', *Journal of Int. Association for Mathematical Geology* **30, n.4**, 323–345.

Gorsich, D. and Genton, M. (2000), 'Variogram model selection via nonparametric derivate estimation', *Journal of Int. Association for Mathematical Geology* **32, n.3**, 249–270.

Gribov, A., Krivoruchko, K. and Ver Hoef, J. (2000), 'Modified weighted least squares semivariogram and covariance model fitting algorithm', *Stochastic Modeling and Geostatistics. AAPG Computer Applications in Geology* **2**.

Journel, A. and Huijbregts, C. (1978), *Mining Geostatistics*, Academic Press, London.

Kyung-Joon, C. and Shucany, W. (1998), 'Nonparametric kernel regression estimation near endpoints', *J. Statist. Plann. Inference* **66**, 289–304.

Maglione, D. and Diblasi, A. (2001), 'Choosing a valid model for the Variogram of an isotropic spatial process', *2001 Annual Conference of Int. Association for Mathematical Geology* .

Matheron, G. (1963), 'Principles of geostatistics', *Economic Geology* **58**, 1246–1266.

Ribeiro Jr, P. and Diggle, P. (2001), 'geoR: A package for geostatistical analysis', *R-NEWS* **vol 1, n.2**, ISSN 1609–3631.

Shapiro, A. and Botha, J. (1991), 'Variogram fitting with a general class of conditionally nonnegative definite functions', *Computational Statistics and Data Analysis* **11**, 87–96.

Stein, M. (1998), *Interpolation of Spatial Data - Some Theory for Kriging*, Springer.

Zimmerman, D. and Zimmerman, M. (1991), 'A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors', *Technometrics* **33, n.1**, 77–91.