ELSEVIER

Computers & Geosciences 34 (2008) 1073–1079

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation ☆

Eulogio Pardo-Iguzquiza*, Mario Chica-Olmo

Department of Geodynamics/CEAMA, University of Granada, 18071 Granada, Spain

Received 23 January 2007; received in revised form 13 September 2007; accepted 17 September 2007

Abstract

In many modern applications of geostatistics in the earth sciences, the empirical information is abundant and with complete spatial coverage (e.g. satellite sensor images). In these cases, a critical characteristic of spatial variability is the continuity of the random field that better models the natural phenomenon of interest. Such continuity describes the smoothness of the process at very short distances and is related to the behaviour of the semivariogram near the origin. For this reason, a semivariogram model that is flexible enough to describe the spatial continuity is very convenient for applications. A model that provides such flexibility is the Matern model that controls continuity with a shape parameter. The shape parameter must be larger than zero; a value larger than 1 implies a random field that is m-times mean square differentiable if the shape parameter is larger than m. A package of computer programs is provided for performing the different steps of a geostatistical study using the Matern model and the performance and implementation are illustrated by an example.

Keywords: Differentiable random field; Shape parameter; Spectral simulation; Weighted least squares

1. Introduction

In many applications of geostatistics in geosciences, the empirical information is sparse such as rain gauges in a river basin, piezometers in an aquifer and drill holes in mining exploration. The continuity of the random field that models the spatial variability of the phenomenon under study is difficult to assess: a nugget variance is often used and a linear behaviour of the semivariogram model close to the origin is assumed for practical purposes. The spherical and exponential models have been the most popular between practitioners. However, in modern applications where experimental data are gathered with complete spatial coverage as in remote sensing, image analysis and some geophysical scanning techniques, there is an abundance of empirical data. In these cases, apart from the spatial variability for medium to large distances described by the range, there is the practical interest of being able to better characterize the variability for short distances. That behaviour of the semivariogram close to the origin is related to the continuity and differentiability of the random field and its local

☆Code available from server at http://www.iamg.org/CGEditor/index.htm.

*Corresponding author. Tel.: +34 958240780.
E-mail address: pardoiguzquiza@yahoo.es (E. Pardo-Iguzquiza).

0098-3004/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2007.09.020

smoothness. Thus a semivariogram model that is flexible enough for characterizing the smoothness of the random field that better models the spatial variable under study is of great interest. This flexibility must allow the model to go without discontinuity from non-differentiable random fields, to random fields that are differentiable once, twice, etc. until the model that can be differentiated an infinite number of times. The Matern semivariogram model (Matern, 1960; Stein, 1999) is such a model where the smoothness of the random field is controlled by a shape parameter. This model is also known by other names as the K-Bessel model (Chilès and Delfiner, 1999) and the Whittle–Matern model (Guttorp and Gneiting, 2006). It is defined as

$$\gamma(\mathbf{h}) = \sigma^2 \left\{ 1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|\mathbf{h}|}{\alpha} \right)^{\nu} K_{\nu}\left(\frac{|\mathbf{h}|}{\alpha} \right) \right\} \qquad (1)$$

where $\sigma^2 > 0$ is the variance, $\alpha > 0$ is scale (or range) parameter, $\nu > 0$ is shape parameter, $\Gamma(.)$ is the gamma function, $K_\nu(.)$ is the modified Bessel function of the second kind and order ν and $|\mathbf{h}|$ is the norm of vector $\mathbf{h}$.

For $\nu = 0.5$, the Matern model is the exponential model and the Gaussian model is a limiting case of the Matern model when ν tends to infinity. The case $\nu = 1$ has also been recognized as a significant model (Whittle, 1954). The Matern model has been applied to different areas such as hydrology (Rodriguez-Iturbe and Mejia, 1974), soil science (Minasny and McBratney, 2005), topography (Handcock and Stein, 1993) and geophysics (Handcock and Wallis, 1994), among others.

The Matern semivariogram is a bounded model, reaching the sill asymptotically. Fig. 1 shows a variety of Matern semivariograms for different values of the shape parameter. In Fig. 1 the scale parameter is kept constant with a value of one and the shape parameter is made variable. A nugget variance can be included in the model and the model can be anisotropic in the scale parameter with the usual convention for anisotropy as may be seen in Isaaks and Srisvastava (1989).

2. Simulation of Matern random fields

MSPECSIM is a Fortran program that implements conditional and non-conditional simulations. The non-conditional simulation is performed by the spectral simulation method using the fast Fourier transform of a covariance model to calculate the spectral density (Borgman et al., 1984; Gutjahr, 1989). A random phase is added to the spectral density and, by inverse fast Fourier transform, the simulated realization is obtained. The program is similar to SPECSIM (Pardo-Igúzquiza and Chica-Olmo, 1994) but with the implementation of the Matern model and with the possibility of conditional simulation. The conditioning is obtained by kriging using a set of experimental data and the normal score transform (Journel and Huijbregts, 1978). An example of simulations with an increase in the smoothness of the random field is shown in

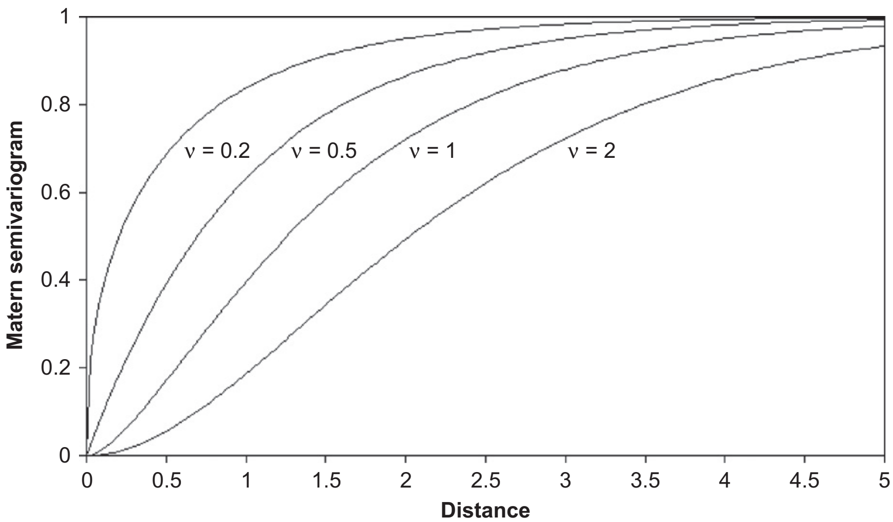

Fig. 1. Examples of Matern semivariograms for different values of shape parameter and a constant scale parameter of one. In all cases, variance is equal to one.

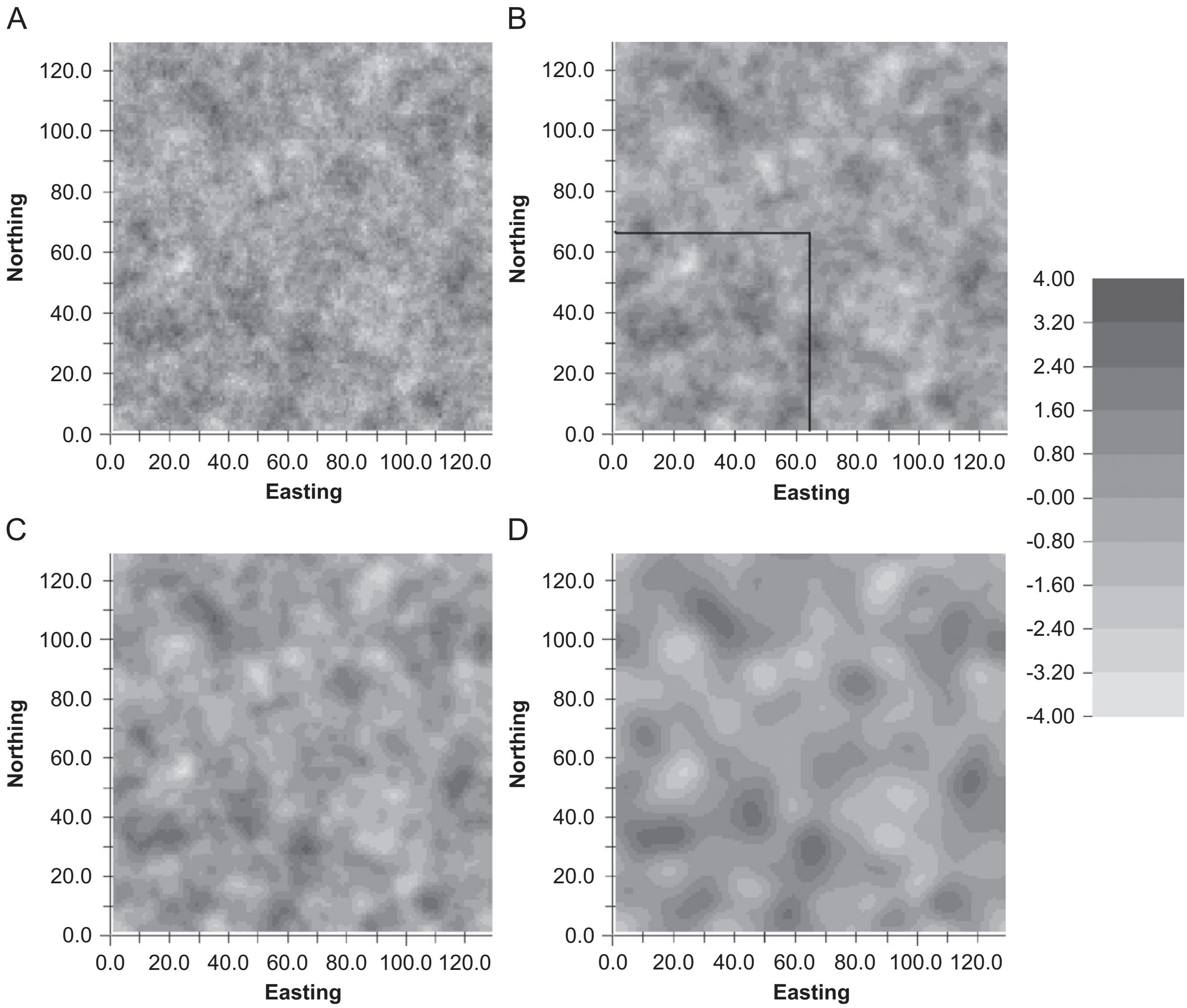

Fig. 2. Realizations of random fields with Matern semivariogram model. Sequence of realizations, from (A) to (D), shows an increase in degree of smoothness. (A) is not differentiable. (B) to (D) are differentiable. Parameters of models (shape parameter, scale parameter, variance) used in generating realizations are (A) 0.5, 4.35, 1; (B) 1.1, 3.36, 1; (C) 2.1, 2.67, 1; (D) Gaussian, 9.40, 1.

Fig. 2A–D, for shape parameters of 0.5, 1.1, 2.1 from Fig. 2A–C respectively, while Fig. 2D represents a realization with a Gaussian semivariogram. Each realization has locations on a square grid with unit interdistance along both the X and the Y directions. In all cases, the scale parameter has been chosen in such way that for the four realizations the semivariogram reaches 99% of the variance at a distance of 20 units. The same random phase has been used for all the realizations (same random seed when running the program) in order to have an equivalent realization where the smoothness is being increased as may be seen in the sequence of Fig. 2A–D. In Fig. 3A–D, each semivariogram model has been represented together with the empirical semivariogram calculated from the realizations of Fig. 2 and for the four main geographical directions, E–W, N–S, NE–SW and NW–SE. The good correspondence between experimental and model seems obvious from the figures. MSPECSIM is as easy to use as SPECSIM, with the parameters being introduced interactively. The only difference is that the program offers the Matern model as an additional option for semivariogram model and if the Matern model is selected, the shape parameter must be entered when requested. If a conditional simulating is desired, the name of an experimental data file will be typed in when required by the program.

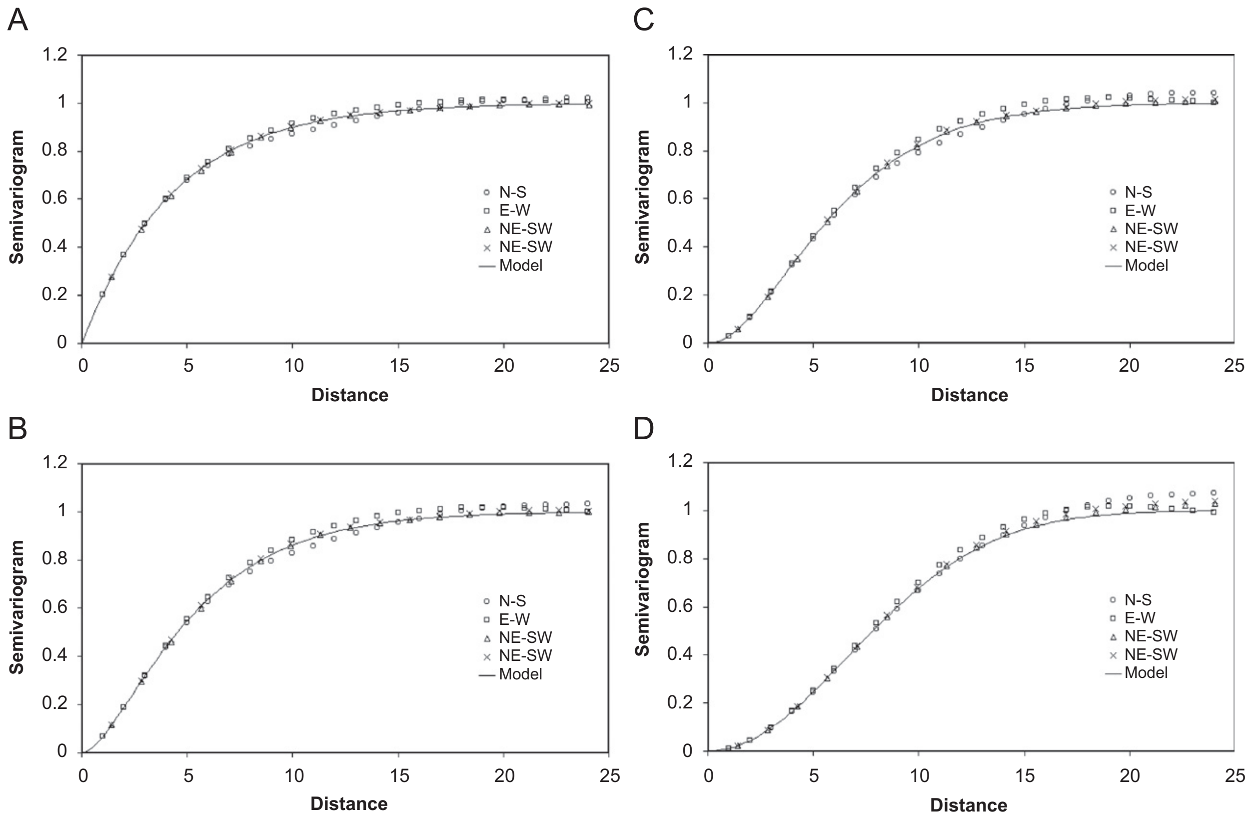

Fig. 3. Experimental semivariograms along main geographical directions for simulated random fields shown in Fig. 2. Theoretical Matern semivariogram models that were imposed to simulation algorithm are also represented as a solid line. Parameters of model (shape parameter, scale parameter, variance) are the same as in Fig. 2.

The spectral simulation method has its advantages and disadvantages (see Chilès and Delfiner, 1999, p. 504). Particularly it is advisable to simulate a random field whose dimension (for example the side length of a square field) is as large as possible with respect to the practical length for which the semivariogram reaches the variance value (or equivalently the covariance value vanishes). From this large simulated field the area of interest can be taken for practical use.

3. Fitting a Matern model by weighted least squares

A sample of 200 experimental data is represented in Fig. 4. The 200 experimental locations were taken at random from the bottom left quarter of the random field shown in Fig. 2B (a realization with theoretical parameters for the Matern semivariogram model of 1.1, 1.0 and 3.36 for the shape parameter, variance and scale parameter, respectively). The omnidirectional semivariogram is shown in Fig. 5 where the number of data pairs for some of the semivariogram lags has been represented. In Table 1 the number of data pairs is given for all the experimental semivariogram lags. It may be seen how the number of data pairs is less than 100 for the first three lags and much larger for the rest of the lags. MVARFIT is a Fortran computer program for fitting a Matern semivariogram model to the empirical semivariogram. There is the possibility of using the five weighting functions described in VARFIT (Pardo-Igúzquiza, 1999) and repeated in Table 2. One can also fit a nugget variance, keep the experimental variance as the estimated variance and fit an anisotropic semivariogram when the experimental semivariogram for more than one direction is used as input file. Table 2 shows the results of running the program with the experimental semivariogram given in Table 1. Fig. 5 shows the experimental semivariogram and the models fitted using the weighting functions 1 and 4 of Table 2. The model fitted by weighting function 1 (i.e. ordinary least squares) tends to give the best global fitting to all the lags and thus the fitting is not so good for the first lags. Contrarily, the weighting function 4 takes into account the value of the semivariogram, which implies a better fitting to the first lags and results in a better estimate of the underlying shape parameter. The experimental semivariogram of Table 1 has been calculated with the provided program VAR2D, a program for estimating the semivariogram by the method of moments. It requires a parameter file as the one shown in Table 3 in the file sam200.par. The content of the parameter file is: file with experimental data (sam200.dat), number of semivariogram directions to consider (1), direction in degree (0.0), number of semivariogram lags (30), basic lag length (1.0 unit), lag length tolerance (0.5 units), angle tolerance (90°, thus the omnidirectional semivariogram is calculated), output file with the results (sam200.var).

Additionally, and for visual fitting of the experimental semivariogram with the Matern model, an interactive graphics computer program, MOD2D, is also provided.

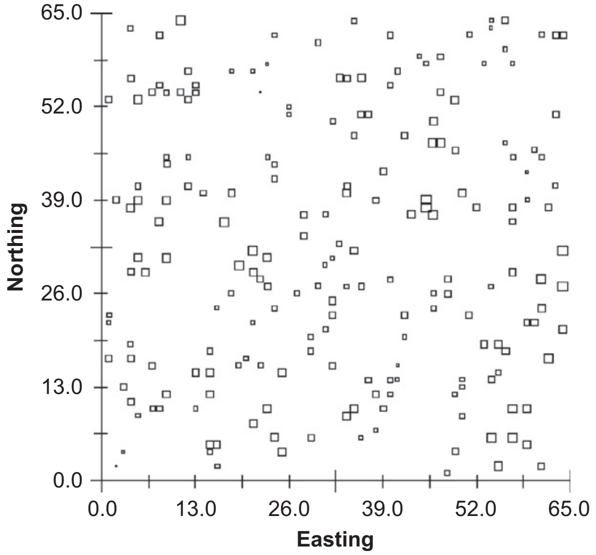

Fig. 4. Pictogram with locations of 200 experimental data sampled from realization of random field shown in Fig. 2B, which corresponds to a Matern semivariogram model with a shape parameter of 1.1, scale parameter of 3.36 and variance of 1. Samples are located at random inside lower left quarter of realization shown in Fig. 2B. In the figure, the size of each square is proportional to value of realization of random field at that location.

4. Kriging with the Matern model

MKRIGE2D is a Fortran program for universal kriging in two dimensions that implements the Matern semivariogram model. The drift order can be zero (then ordinary kriging), one (linear drift) or two (quadratic drift). This and other parameters are specified in a parameter file used as input to the program. An example of parameter file is the file sam200.kri shown in Table 3. The content of the parameter file is (Table 3): file with experimental

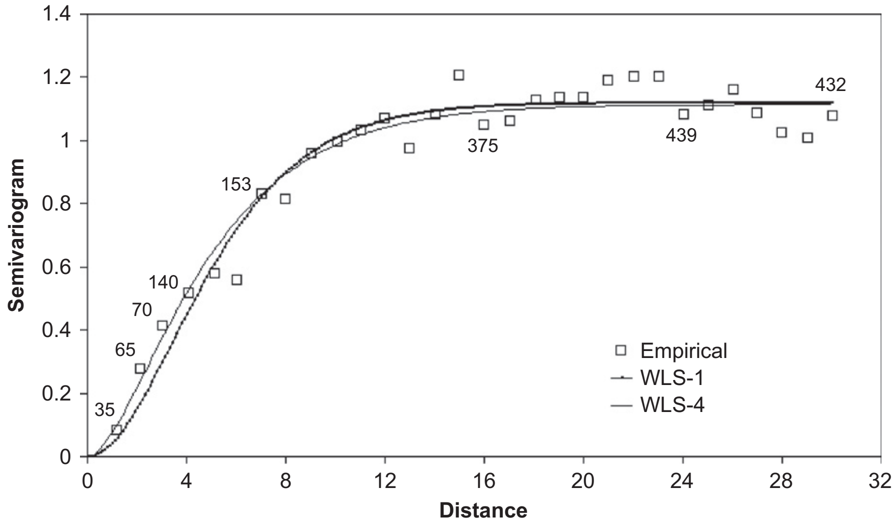

Fig. 5. Empirical semivariogram and models fitted by weighted least squares (WLS) using weighting functions 1 (WLS-1) and 4 (WLS-4) shown in Table 2. Experimental semivariogram has been calculated using 200 experimental data shown in Fig. 4. Number of data pairs for estimating each semivariogram lag ranges from 35 as minimum to 439 as maximum. Number of data pairs for some selected lags is also shown.

Table 1
Empirical semivariogram of experimental data shown in Fig. 4

1	1.048680	
30	0.000000	
1.213024	8.115137E−02	35
2.152536	2.782877E−01	65
3.037153	4.133573E−01	70
4.110816	5.178697E−01	140
5.148042	5.780738E−01	117
6.063357	5.595282E−01	143
7.060555	8.290541E−01	153
8.011979	8.119491E−01	180
9.044687	9.571254E−01	254
10.129390	9.964118E−01	229
11.058030	1.033403	267
12.017810	1.070408	246
13.023180	9.729821E−01	317
14.029380	1.082919	346
15.006060	1.204629	287
16.009420	1.047410	373
17.075580	1.059508	351
18.099700	1.126120	341
19.072090	1.136753	345
19.994900	1.135851	323
20.974590	1.189553	459
22.037690	1.202729	369
23.041730	1.201973	400
24.026460	1.080095	439
25.058080	1.110324	481
26.067530	1.161920	415
27.017640	1.088051	418
28.018900	1.025299	442
29.029090	1.006355	393
30.037220	1.078322	432

The first row indicates the number of directions and experimental variance. The second row contains the number of lags and direction of semivariogram (an arbitrary number in this case as the omnidirectional variogram was calculated). The next lines represent, for each lag, mean distance, mean semivariogram and number of pairs used in the estimation of the semivariogram for that lag. This file is the input file of program MVARFIT. It is also output of program VAR2D and parameter file sam200.par given in Table 3.

data (sam200.dat), the drift order (0), maximum number of experimental data allowed in a kriging neighbourhood (30), minimum number of neighbours for kriging (2), radius of the circle search (40 units), origin of the estimation grid (1,1), interpoint distance of the estimation grid (1,1), number of grid points (64,64), point kriging (1, 2 for block kriging), semivariogram model (sam200.mod), output file with grid coordinates and estimation at each grid location (sam200e.dat) and output file with grid coordinates and estimation variance at each grid location (sam200v.dat), no information on the

Table 3
Examples of different parameter files

Sam200.kri	Sam200.par	Sam200.mod
Sam200.dat	Sam200.dat	0.000000E+00
0	1	1
30	0.0	4
2	30	1.114000
40.0	1.0	2.970000
1.0 1.0	0.5	0.000000E+00
1.0 1.0	90.0	1.000000
64 64	Sam200.var	1.213000
1		
Sam200.mod		
Sam200e.dat		
Sam200v.dat		
0		

First column: parameter file (sam200.kri) for kriging program MKRIGE2D. Second column: a parameter file (sam200.par) for calculating experimental semivariogram from sparse data on plane. Third column: file (sam200.mod) with a semivariogram model. These files are described in the main text. Sam200.dat is file with 200 experimental data shown in Fig. 4. Sam200.var is output variogram file shown in Fig. 5. Matern model represented by sam200.mod can be seen in Fig. 5 as model fitted using WLS-4. The same model is given in Table 2.

Table 2
Results of weighted least-squares fitting using MVARFIT

Weighting function		Estimated shape parameter $\hat{\nu}$	Estimated scale parameter $\hat{\alpha}$	Estimated variance $\hat{\sigma}^2$
1	1.0	1.966	2.650	1.119
2	$N(h)$	2.379	2.041	1.121
3	$1.0/(\gamma(h))^2$	1.769	2.384	1.072
4	$N(h)/(\gamma(h))^2$	2.970	1.114	1.213
5	$1.0/h^2$	4.117	1.165	0.902

Matern semivariogram model is fitted to experimental semivariogram shown in Fig. 5 and Table 1. The first weighting function is constant and equal to 1, which implies ordinary least squares fitting. The rest of weighting functions include number of data pairs for each lag, $N(h)$, semivariogram $\gamma(h)$ for each lag or the distance of each lag h. Number of lags is in the numerator in order to represent uncertainty inversely proportional to this number. On the other hand, semivariogram value and distance appear in denominator in order to give more weight to lags close to origin.

screen (0, 1 if information is desired). The semivariogram model file (sam200.mod in Table 3) has the following content: nugget effect (0.0), number of nested structures (1), kind of structure (4 for the Matern model, 1 for spherical, 2 for exponential, 3 for Gaussian), semivariogram sill (1.114), range or spatial scale parameter (2.97), anisotropy angle in degree (0.0), anisotropy ratio (1.0), shape parameter (1.213). The last five lines should be repeated for each additional nested structure in the case that the number of structures is larger than one. It is assumed that the variance is equal to nugget variance plus sill variance.

5. Discussion and conclusions

The Matern model is of great interest in the geostatistical analysis of spatial processes where the smoothness of the random field model is of relevance. The Matern model introduces flexibility and control in the degree of smoothness of the random field to be used. Furthermore, the exponential model is a particular case of the Matern model and the Gaussian model is a limiting case. Because of this, the use of the Matern model has been encouraged by some authors such as Stein (1999). The model is known since Whittle (1954) and Matern (1960), but its use in applications has been limited. In fact, in case studies with just a few sparse experimental data, it is often not possible to infer with reliability the behaviour of the semivariogram close to the origin and a linear behaviour is assumed implicitly by the use of the spherical or exponential models. However, nowadays there is an availability of spatial information with complete coverage of a given area and in all these cases there is enough information of variability at short distances for inferring the smoothness of the random field model. Because of this the purpose of this paper is to describe a package of computer programs that are provided for geostatistical analysis with the Matern model. The examples can be used to check the implementation of the programs.

Acknowledgements

The first author acknowledges a ‘‘Ramon y Cajal’’ Grant from the Ministry of Science and Education of Spain. We are grateful for the financial support given by the Spanish MCyT (Project CGL2006-06845/CLI) and Junta de Andalucía (Group RNM122). We would like to thank the reviewers for their constructive criticism.

References

Borgman, L., Taheri, M., Hagan, R., 1984. Three-dimensional frequency-domain simulations of geological variables. In: Verly, G., David, M., Journel, A.G., Maréchal, A. (Eds.), Geostatistics for Natural Resources Characterization. Reidel, Dordrecht, pp. 517–541.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 695pp.

Gutjahr, A., 1989. Fast Fourier transforms of random fields. Technical Report No. 4-R58-2690R. Los Alamos Grant, Department of Mathematics, New Mexico Institute of Mining and Technology, Socorro, NM, 106pp.

Guttorp, P., Gneiting, T., 2006. Studies in the history of probability and statistics XLIX on the Matern correlation family. Biometrika 93 (4), 989–995.

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

Handcock, M.S., Wallis, J.R., 1994. An approach to statistical spatio-temporal modeling of meteorological fields. Journal of the American Statistical Association 89, 368–378.

Isaaks, E.H., Srisvastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, Oxford, 561pp.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600pp.

Matern, B., 1960. Spatial Variation. Meddelanden fran Statens Skogsforskningsinstitut, 49, no. 5 (second ed., 1986, Lecture Notes in Statistics, no. 36). Springer, New York.

Minasny, B., McBratney, A.B., 2005. The Matern function as a general model for soil variograms. Geoderma 128 (3–4), 192–207.

Pardo-Igúzquiza, E., 1999. VARFIT: a Fortran-77 program for fitting variogram models by weighted least squares. Computers & Geosciences 25 (3), 251–261.

Pardo-Igúzquiza, E., Chica-Olmo, M., 1994. SPECSIM: a program for simulating random fields by an improved spectral approach. Computers & Geosciences 20 (4), 597–613.

Rodriguez-Iturbe, I., Mejia, J.M., 1974. The design of rainfall networks in time and space. Water Resources Research 10 (4), 713–728.

Stein, M.L., 1999. Interpolation of Spatial Data. Some Theory for Kriging. Springer, New York, 247pp.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 439–449.