Computers & Geosciences 35 (2009) 1139–1150

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

MLMATERN: A computer program for maximum likelihood inference with the spatial Matérn covariance model ☆

Eulogio Pardo-Igúzquiza [a,*], Kanti V. Mardia [b], Mario Chica-Olmo [a]

[a] Departamento de Geodinámica/CEAMA, Avenida Fuentenueva s/n, Universidad de Granada, 18071 Granada, Spain
[b] Department of Statistics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

ARTICLE INFO

Article history:
Received 15 November 2007
Received in revised form
11 June 2008
Accepted 3 September 2008

Keywords:
Geostatistics
Kriging
Simulation
Semivariogram
Sparse data
Smoothness parameter
Gaussian process

ABSTRACT

The Matérn covariance scheme is of great importance in many geostatistical applications where the smoothness or differentiability of the random field that models a natural phenomenon is of interest. In addition to the range and nugget parameters, the flexibility of the Matérn model is provided by the so-called smoothness parameter which controls the degree of smoothness of the random field. It has been the usual practice in geostatistics to fit theoretical semivariograms like the spherical or exponential, thus implicitly assuming the smoothness parameter to be known, without questioning if there is any theoretical or empirical basis to justify such assumption. On the other hand, if only a small number of sparse experimental data are available, it is more critical to ask if the smoothness parameter can be identified with statistical reliability. Maximum likelihood estimation of spatial covariance parameters of the Matérn model has been used to address the previous questions. We have developed a general algorithm for estimating the parameters of a Matérn covariance (or semivariogram) scheme, where the model may be isotropic or anisotropic, the nugget variance can be included in the model if desired, and the uncertainty of the estimates is provided in terms of variance–covariance matrix (or standard error-coefficient of correlation matrix) as well as likelihood profiles for each parameter in the covariance model. It is assumed that the empirical data are a realization of a Gaussian process. Our program allows the presence of a polynomial trend of order zero (constant global mean), one (linear trend) or two (quadratic trend). The restricted maximum likelihood method has also been implemented in the program as an alternative to the standard maximum likelihood. Simulation results are given in order to investigate the sampling distribution of the parameters for small samples. Furthermore, a case study is provided to show a real practical example where the smoothness parameter needs to be estimated.

1. Introduction

In geostatistics, spatial covariance and semivariogram models play a key role. The covariance model may be known from theoretical considerations, but more often must be inferred from the experimental data. Often, in practice one uses a covariance model whose behavior close to the origin is fixed in advance rather than being estimated from the experimental data. For example, when fitting a spherical or exponential model, one assumes that the random field, used to model the spatial variable of interest is not differentiable, or on the other hand, if a Gaussian semivariogram model is used, one is assuming that the random field can be differentiated any number of times.

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

* Corresponding author. Tel.: +34 958 240780.
E-mail address: pardoiguzquiza@yahoo.es (E. Pardo-Igúzquiza).

0098-3004/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2008.09.009

However, there are practical problems where the smoothness of the random field is of great interest. For example, when the geovariables are sampled intensively as for a computer-scanned image in petrology or in a high-resolution image in remote sensing. Additionally, a differentiable model is necessary for kriging estimation of directional derivatives and hence the gradient of a random field (Pardo-Igúzquiza and Chica-Olmo, 2007).

A flexible covariance model for modelling the smoothness, then differentiability, of a random field is the Matérn model (Whittle, 1953; Matérn, 1960), which can be defined as:

$$C(\mathbf{h}) = \begin{cases} \sigma^2 & |\mathbf{h}| = 0 \\ \left(\dfrac{(\sigma^2 - \delta^2)}{2^{v-1}\Gamma(v)}\right)\left(\dfrac{|\mathbf{h}|}{\alpha}\right)^v \mathrm{K}_v\left(\dfrac{|\mathbf{h}|}{\alpha}\right) & |\mathbf{h}| > 0 \end{cases} \quad (1)$$

where $\sigma^2 > 0$ is the total variance (sill); $\delta^2 \geqslant 0$ is the nugget variance; $\alpha > 0$ is the range (also known as the scale parameter); $v > 0$ is the smoothness parameter (also known as the shape parameter); $\Gamma(\cdot)$ is the gamma function; $\mathrm{K}_v(\cdot)$ is the modified Bessel function of the second kind and order v; $\sigma^2 - \delta^2$ is the total variance minus nugget (partial sill); $\mathbf{h} = (x, y)$ is the distance vector with two components in the plane.

Note that the semivariogram is given by

$$\gamma(\mathbf{h}) = \sigma^2 - C(\mathbf{h}).$$

For $v = 0.5$, the Matérn model is identical to the exponential model and the limit case as v tends to infinite is the Gaussian model. The case $v = 1$ has also been used in practice (Rodriguez-Iturbe and Mejia, 1974). If the smoothness parameter is larger than one, $v > 1$, then the random field is $[v]$ times mean square differentiable, where $[v]$ is the integer part of v.

The isotropic model given in Eq. (1) can be extended to the anisotropic case by using the usual geometric anisotropy transformation i.e. by introducing an anisotropy ellipse with three parameters: the two ellipse semi-axis and the anisotropy angle. In the program, these three parameters are named rangeX, rangeY and anisotropy angle parameter. The angle parameter is the angle between rangeX and the X-axis, measured counterclockwise from the X-axis. Thus, the maximum number of parameters for the Matérn covariance model considered in this paper is six: total variance, nugget variance, smoothness parameter, rangeX, rangeY and anisotropy angle.

It is a common practice in geostatistics to have a parameter for defining the distance beyond which two locations are no longer correlated. This parameter is the range for a spherical model or the practical range for models where the correlation tends to zero as $|\mathbf{h}| \to 0$. An example of the latter case is the "practical range" of an exponential covariance model, which is usually taken to be three times its range parameter.

For the Matérn covariance model, the "practical range" defined as the distance for which the value of the correlation between two locations is equal to 0.05 is a function of both the range parameter α and the smoothness parameter v. This dependency leads to high negative correlation for estimates of α and v, so that if one of those two parameters is on average underestimated, the other will be on average overestimated.

The basic model adopted in this paper for the random field is to assume that it has two components: a deterministic component known as the trend and which accounts for the long correlations (i.e. low frequency variation) and a zero-mean second-order stationary stochastic component or residual, which models the short correlations (i.e. high frequency variation). Given a multivariate normal sample vector $\mathbf{Z}$, the log-likelihood with parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ may be written as (Mardia, 1980; Mardia and Marshall, 1984; Mardia, 1990)

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{Z}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)| - \frac{1}{2\sigma^2}(\mathbf{Z} - \mathbf{F}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)^{-1}(\mathbf{Z} - \mathbf{F}\boldsymbol{\beta}) \quad (2)$$

where n is the number of experimental data; $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\theta}^*)$ is the covariance parameters, so that $\boldsymbol{\theta}^*$ is the vector of covariance function parameters apart from the variance, i.e. $\boldsymbol{\theta}^*$ is the vector of correlation parameters; $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)$ is the the $n \times n$ correlation matrix (it will be also denoted by $\boldsymbol{\Sigma}$ if there is no ambiguity); superscript T denotes matrix transpose; $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)^{-1}$ is the inverse of the correlation matrix; $|\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)|$ is the determinant of the $n \times n$ correlation matrix; $\mathbf{Z}$ is the $n \times 1$ vector of experimental data; $\mathbf{F}$ is the $n \times p$ matrix of known basis function for the trend. In our implementation, we have used the two-dimensional monomials $\{1, x, y, x^2, y^2, xy\}$ from the spatial location of $\mathbf{Z}$. The first monomial represents a trend of order zero, the next three represents a linear trend and the last six represent quadratic trend; $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown trend coefficients.

For any given set of correlation parameters $\boldsymbol{\theta}^*$, it can be shown from Eq. (2) that the maximum likelihood estimates of trend coefficients $\boldsymbol{\beta}$ and variance σ^2 are given by (see for example Mardia and Marshall, 1984)

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{Z} \quad (3)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{Z} - \mathbf{F}\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}^{-1}(Z - F\hat{\boldsymbol{\beta}})}{n}. \quad (4)$$

By substituting Eqs. (3) and (4) into Eq. (2), the ML estimates of the correlation parameters $\boldsymbol{\theta}^*$ can be obtained by numerical maximization of the following expression:

$$L'(\boldsymbol{\theta}^*; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{Z}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2} + \frac{n}{2}\ln(n) - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{n}{2}\ln((\mathbf{Z} - \mathbf{F}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{F}\hat{\boldsymbol{\beta}})). \quad (5)$$

Alternatively, we may want to use the restricted maximum likelihood (REML) estimates instead of ML estimates which filters out the effect of the polynomial trend. In REML, the expression to be maximized is given by

$$L''(\boldsymbol{\theta}^*; \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \mathbf{Z}) = -\frac{m}{2}\ln(2\pi) - \frac{m}{2} + \frac{m}{2}\ln(m) - \frac{1}{2}\ln|\boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{\mathrm{T}}| - \frac{m}{2}\ln((\boldsymbol{\Lambda}\mathbf{Z})^{\mathrm{T}}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{\mathrm{T}})^{-1}(\boldsymbol{\Lambda}\mathbf{Z})) \quad (6)$$