ELSEVIER

Computers & Geosciences 31 (2005) 361–370

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

EMLK2D: a computer program for spatial estimation using empirical maximum likelihood kriging ☆

Eulogio Pardo-Igúzquiza[a], Peter A. Dowd[b,*]

[a]*Department of Mining and Mineral Engineering, University of Leeds, Leeds LS2 9JT, UK*
[b]*Faculty of Engineering, Computer and Mathematical Sciences, University of Adelaide, Adelaide, SA 5006, Australia*

Received 13 August 2003; received in revised form 15 August 2004; accepted 1 September 2004

Abstract

The authors describe a Fortran-90 program for empirical maximum likelihood kriging. More efficient estimates are obtained by solving the estimation problem in the 'Gaussian domain' (i.e., using the normal scores of the experimental data), where the simple kriging estimate is equivalent to the maximum likelihood estimate and to the conditional expectation. The transform to normality is done using the empirical cumulative probability distribution function. A Bayesian approach is adopted to ensure a conditionally unbiased estimate, which is obtained as the mean of the posterior distribution. The posterior distribution also provides a complete specification of the probability of the variable and thus provides the basis for a more realistic evaluation of uncertainty by various methods: inverting Gaussian confidence intervals, confidence intervals measured from the posterior distribution, variance measured from the posterior distribution or intervals obtained using the likelihood ratio statistic. A detailed case study is used to demonstrate the use of the program.

Keywords: Geostatistics; Linear kriging; Bayesian estimation; Maximum likelihood; Posterior distribution; Conditional unbiasedness

1. Introduction

Linear kriging (simple kriging, ordinary kriging and universal kriging) is a well-known spatial estimator of point and integral values that provides unbiased estimates with minimum estimation variance. One advantage of linear kriging is that, as it makes no distributional assumptions, it is a distribution-free estimator. The corresponding disadvantage is that it delivers only an estimate and the associated estimation variance. This is of no consequence when the experimental data follow a multivariate Gaussian distribution because the kriging estimator then coincides with the conditional expectation, which is the most efficient estimator available (linear or otherwise) Matheron (1975), Papoulis (1984); Rivoirard (1994). For this reason it has often been suggested that the original data (original domain) are transformed to normal scores (Gaussian domain) and the estimation is done in the Gaussian domain before back-transforming the estimates to the original domain. Whilst the transform of the data to the Gaussian domain can be achieved by using the empirical cumulative probability density function, the back-transform of the estimates to the original domain is more difficult as the inverse of the transform of the data to the Gaussian domain does not

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

*Corresponding author. Tel.: +61 8 8303 4700; fax: +61 8 8303 4361.

E-mail address: peter.dowd@adelaide.edu.au (P.A. Dowd).

0098-3004/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2004.09.020

generally ensure conditionally unbiased estimates, i.e., if Y^* is a minimum variance unbiased estimate in the Gaussian domain and Z^* is the estimate in the original domain, conditional unbiasedness implies that (Chilès and Delfiner, 1999):

$$\mathrm{E}\{Z^*|Y^*\} = \mathrm{E}\{Z|Y^*\}. \tag{1}$$

In general, however, conditional unbiasedness is not achieved by deriving the estimator Z^* from Y^* by using the inverse of the transform of the data to the Gaussian domain, i.e.

$$Z^* = \varphi^{-1}(Y^*) \tag{2}$$

will be conditionally biased when $\varphi(.)$ is the transform of the data to the Gaussian domain.

For a specified probability distribution a functional transform of the data can be used and it is then possible, in principle, to determine the back-transform of the Gaussian estimates that yields conditionally unbiased estimates in the original domain. For the lognormal distribution, for example, conditionally unbiased estimates are obtained by using the back-transformation:

$$Z^* = \exp(Y^* + \tfrac{1}{2}\sigma_{SK}^2), \tag{3}$$

which requires the estimation variance, σ_{SK}^2, of Y^*, which is provided by simple kriging and which is equal to the variance of the conditional distribution in the Gaussian domain (see, for example, Dowd, 1982).

Conditional unbiasedness guarantees that estimates are unbiased in the original domain, which is ultimately the domain of interest; the Gaussian domain is simply an intermediate state for improving the efficiency of the estimates.

As a rule in earth science applications, the multivariate distribution of the experimental data is unknown. Even the marginal probability distribution, inferred from the experimental histogram, is uncertain and thus functional back-transforms, such as that in Eq. (3), have questionable validity especially as most are very sensitive to departures from the distributional assumptions.

An approximation given by Chilès and Delfiner (1999) is

$$Z^* = \varphi(Y^*) + \tfrac{1}{2}\sigma_{SK}^2\varphi''(Y^*), \tag{4}$$

where $\varphi''(Y^*)$ is the second derivative of the Gaussian transform of the data $Y = \varphi(Z)$. The Gaussian transform is, in general, unknown and even if it can be estimated, the estimation of first and second derivatives is much more unstable (unreliable). As an alternative, we propose the method of empirical maximum likelihood kriging (Pardo-Igúzquiza and Dowd, 2004).

2. Methodology

The approach proposed in this paper comprises the following steps, as given in more detail in Pardo-Igúzquiza and Dowd (2004):

- **The original data are transformed to normal scores** using the empirical cumulative distribution function:

$$\hat{F}(z) = \hat{P}\{Z \leqslant z\} = \frac{\aleph(z_i \leqslant z)}{n}, \tag{5}$$

where $\aleph(z_i \leqslant z)$ is the cardinal of the set $(z_i \leqslant z)$, i.e., the number of experimental values less than or equal to z. In practice, the experimental data are sorted into increasing order to provide the order statistics $\{z_{(1)}, z_{(2)}, \ldots, z_{(n)}\}$; the function $\hat{F}(z)$ is then estimated at these values of Z by

$$\hat{F}(z_{(i)}) = \frac{i - 0.5}{n}. \tag{6}$$

The corresponding set of normal scores $\{y_{(1)}, y_{(2)}, \ldots, y_{(n)}\}$ is then given by

$$y_{(i)} = \Phi^{-1}(\hat{F}(z_{(i)})), \tag{7}$$

where $\Phi^{-1}(.)$ is the inverse of the standard cumulative Gaussian distribution function.

For values that do not coincide with the experimental data the normal score value is estimated by linear interpolation:

$$y \in [y_{(i)}, y_{(i+1)}] \Rightarrow z = z_{(i)} + \frac{(z_{(i+1)} - z_{(i)})}{(y_{(i+1)} - y_{(i)})}(y - y_{(i)}). \tag{8}$$

The tails of the transform, which lie outside the Gaussian interval $[y_{(1)}, y_{(n)}]$ corresponding to the original domain interval $[z_{(1)}, z_{(n)}]$, must be modelled. This is done by linear interpolation between the values $[y_{(0)}, y_{(1)}]$ and $[y_{(n)}, y_{(n+1)}]$. The values $y_{(0)}$ and $y_{(n+1)}$ are set to -6.0 and 6.0, respectively, and appropriate corresponding z-values $z_{(0)}$ and $z_{(n+1)}$ must be chosen. For a specific application, chemical or physical constraints may suggest appropriate values; for example, for an ore grade, $z_{(0)}$ cannot take negative values. Values outside the Gaussian interval $[-6.0, 6.0]$ are very unlikely and thus the corresponding z-values should be such that most of the probability of Z lies inside the range they define. In general, the back-transformed values are not very sensitive to these limits because the variability of the estimated values is usually less than that of the original data. It is, nevertheless, advisable to make an informed choice of limit values; for example, Gaussian tails suggest the values:

$$z_{(0)} = -6.0\sigma_z + \mu_z, \tag{9}$$

$$z_{(n+1)} = 6.0\sigma_z + \mu_z, \tag{10}$$