

Geographical Information Systems Principles of Ordinary Kriging Interpolator

¹J. Negreiros, ¹M. Painho, ²F. Aguilar and ²M. Aguilar

¹ISEGI-New University of Lisbon, Campus de Campolide, Lisboa, Portugal

²Almeria University, La Cañada de San Urbano, Almeria, Spain

Abstract: The aim of this study resumes the main linear stochastic view for spatial interpolation within Geographical Information Systems (GIS) and still unknown by major GIS users: Ordinary Kriging. To review the geostatistical background to involve complex spatial tasks is, thus, central. It starts with the main concepts of the regionalized data nature, exploratory data analysis and distribution standardization since Mother Nature does not follow, most of the times, the Gaussian curve. Sampling considerations follows next while a deep variography inspection is presented later. Cressie's automatic fitness and Kriging equation system are mentioned, as well. It is expected that this article might be used by Geographical Information System (GIS) users to get acquainted with a more complex but better interpolated technique with two major features: BLUE (Best Linear Unbiased Estimator) and BUE (Best Unbiased Estimator) if data holds a Normal distribution.

Key words: Spatial interpolation, Variogram, Kriging, cross-validation, Kriging variance.

INTRODUCTION

Spatial analysis includes an expanding range of methods which address different spatial issues, from remote sensing to spatial error uncertainty. Each of these methods focuses on geographically raw data correlated by statistical methods. Spatial interpolation, in general and stochastic Kriging, in particular, is then the issue addressed in this study.

The existing knowledge of spatial behavior and the ability to estimate resources have been translated into several approaches (Armstrong, 1998). Based on geological processes with extremely complicated parameters to represent them is an option. An additional model is the creation of a polynomial deterministic term plus a random error component that is uncorrelated from one place to another. Certainly, a persistent uncorrelated error between several distances, assuming error randomness, instead of translating this factor as a function that varies from place to place, creates a difficult problem to resolve (Armstrong, 1998).

The regionalized theory states that unknown regions are correlated to sampled ones on the basis of location distance, that is, there is a variance rate between samples over space in a physical continuity context. If traditional statistical methods assume that all data are independent, spatial samples taken close to one another may be

expected to have more similar values than samples taken farther apart, which is known as Tobler's First Law of Geography (Tobler, 1970). This phenomenon is usually referred to as spatial dependence, spatial continuity or autocorrelation (Griffith, 1993; Soares, 2000).

According to Matheron (1965), the regionalized theory treats spatial properties as the sum of three components: (1) Deterministic structural component, drift or trend; (2) Stochastic or random term with local variation; (3) Gaussian independent spatial residual. Hence, spatial data equals large-scale variation plus small-scale plus error. Data transformation and removal trend concerns the first factor. The variogram, $\gamma(h)$, involves the second issue while error measurement is closely linked with the nugget-effect, C_0 .

As an illustration purpose, the Pb contamination default dataset (98 samples) of a commercial geosoftware called GS+[®] (Gamma Design[®]) has been used along this study. All geocomputation presented here was achieved with GeoMS[®] from the Technical University of Lisbon, Portugal. To highlight Kriging interpolation technique becomes, hence, the goal of this study.

THE INNER SELF OF REGIONALIZED SPATIAL DATA

Under the regionalized assumption, local variation is not generally unstructured but it is spatially dependent at

some scale because points within a given distance apart depend on one another in a statistical sense (Armstrong, 1998). The study of such correlation is called structural analysis or variogram, the tool that quantifies Tobler's Law. Any variable is stationary when its distribution is invariant under translation, that is, the stationary region is the one where scientists want to make estimates (Ella, 2000). For Journel (1996), stationary provides a preliminary model decision for uncertainty about the sample value at a location $z(x)$. For any space lag (h) increment, in a particular direction, the distribution of $z(x_1)$, $z(x_2)$... $z(x_n)$ observations is the same as that of $z(x_1+h)$, $z(x_2+h)$... $z(x_n+h)$, where $z(x)$ is regarded as the sum of three terms: (1) mean, drift or trend, (2) random dependent variable plus (3) error.

This stationary assumption requires the first two moments (mean and variance) to be constant. This spatial homogeneous region makes random functions homogeneous and self-repeating in space. The variance of the random spatial variable equals covariance at zero lag distance ($\sigma^2 = C(0)$) while the covariance between $z(x)$ and $z(x+h)$ becomes $\sigma_{z(x),z(x+h)} = \sigma^2 f(h) = C(h)$. This means that the stationary covariance only depends on the distance and direction between two samples and not on their locations, the second order stationary assumption. Hence, the expected difference between values at two locations separated by h distance is zero.

If $E(\mu(x)) = m$ then $z(x) = m + e(x)$ and $E(z(x)-z(x+h)) = 0$ Covariance at h lag distance comes as $E\{(z(x)-m)(z(x+h)-m)\} = E\{z(x) \times z(x+h) - m^2\} = \text{Cov}(h)$, being $z(x)$ the value at x site, $z(x+h)$ the value at x site plus h lag distance, m represents the overall mean and $E()$ symbolizes the expected value. Yet, if there were a trend or drift over the area then the expected value would change and the mean difference between any two samples would not necessarily be zero. This is why Matheron developed the intrinsic hypothesis, stationary assumption of the difference or variance of differences, by assuming a weakly stationary situation (stationary of order two): the $z(x)-z(x+h)$ increments exist and they are location independent (Armstrong, 1998). Although the covariance does not exist, the variance of the difference between two locations is assumed to be present, it does not depend on the location and $\text{Var}(z(x+h)-z(x)) = E\{(z(x+h)-z(x))^2\} = 2\gamma(h)$, being $E(z(x)-z(x+h)) = 0$. That is, differences between sites are merely a displacement function between them, the intrinsic hypothesis.

The $\gamma(h)$ is called the dissimilarity average function and it is the key tool for the structural interpretation of

phenomena as well as for estimation and uncertainty. Mathematically, the variogram is defined as $\gamma(h) = 0.5 \times \text{Var}(z(x+h)-z(x))$. Since, the mean of $z(x+h)-z(x)$ is zero for stationary and intrinsic variables then the variance is just the mean square difference. Consequently, the variogram becomes as follows: $\gamma(h) = 0.5 \times E\{(z(x+h)-z(x))^2\}$. Conceptually, the variogram indicates how different the values are as distance increases in a specific direction. Statistically and on the basis of the method-of-moments (Cressie, 1993), the variogram estimator becomes as it is shown in Eq. 1 being $N(h)$ the number of point pairs separated by h distance lag, $z(x_i)$ the sample value at the i th location while $z(x_i+h)$ is the value at the i th site plus h lag:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

Two major parameters describe this tool: the range or lag distance at which all successive values are independent of each other and the sill, the variogram value corresponding to the range. Therefore, the range is a length measure of heterogeneity correlation. There is an equivalence between the variogram $\gamma(h)$ and the covariance $C(h)$, but no covariance exists for a non-stationary situation. Yet, the variogram always exists and this makes it more generally useful. Other useful relationships can be drawn: (1) $C(0) = \sigma^2$, the covariance between the same sample equals the variance; (2) $C(h) = C(-h)$, the covariance between two given samples does not depend on their positions; (3) $|C(h)| \leq C(0)$, the module of the covariance is always below or equal to the variance; (4) $C(\infty) = 0$, the covariance between samples further apart an infinite distance equals zero; (5) $\gamma(h) = \gamma(-h)$, the variogram computed in both directions is the same; (6) $\gamma(h) = C(0) - C(h)$, the corresponding variogram is obtained by turning down the covariance upside down (Fig. 1). The proof of the relationship between the variogram and the covariance starts out from the variogram definition: $2\gamma(h) = E\{(z(x+h)-z(x))^2\} = E\{(z(x+h)-m)^2 + (z(x)-m)^2 - 2 \times (z(x+h)-m) \times (z(x)-m)\} = E\{Z^2(x+h)\} + E\{Z^2(x)\} - 2E\{z(x+h) \times z(x)\} = \sigma^2(z(x+h)) + \sigma^2(z(x)) - 2C(h) = 2C(0) - 2C(h)$. Clearly, this relation is only possible with bounded variograms and only stationary regionalized variables hold this feature. Unbounded ones come from non-stationary situations with infinite dispersion. In this case, the trend has to be defined as a vicinity position function in the field.

A gradual change in $C(h)$ function is notorious near the origin indicating strong spatial continuity, while a

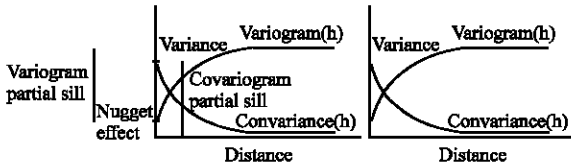


Fig. 1: The variogram and the covariogram function with and without the nugget-effect, C_0

sudden decrease in the variogram points to scale variability. If space is continuous then $E(z(x+h)-z(x))^2$ should be zero when $h = 0$. The main reason for a nugget-effect greater than zero should be measurement error. Yet, it is also admitted that the micro-scale variation leads to a highly irregular variogram pattern at a very short range. Because nothing can be said about variograms at lag distances smaller than $\min\{|z(x+h)-z(x)|\}$ then it is not possible to tell whether micro-scale variation is continuous or not. Therefore, the nugget-effect equals $C_{\text{Micro-scaleVariation}}$ plus $C_{\text{MeasurementError}}$ (Cressie, 1993). According to Armstrong (1998), this dramatic short scale variability has a serious impact on the accuracy of future estimates. On the other hand, a pure nugget-effect variogram concerns a positive constant value for the all lags which means uncorrelated locations for any h distance. As expected, the residual random noise with zero mean, constant variance and zero covariance also exists within this stochastic approach, as is shown in Eq. 2 where $\mu(x)$ represents the deterministic function that describes the structural component of Z at location x ; $\epsilon(x)$ equals the stochastic term whose variance is the nugget-effect; $\gamma_1(h) + \dots + \gamma_n(h)$ represents different variogram structures while ξ defines the spatially independent Gaussian noise term.

$$z(x) = \mu(x) + \epsilon(x) + \xi \quad \text{or} \quad z(x) = \mu(x) + \gamma_1(h) + \dots + \gamma_n(h) + \xi \quad (2)$$

Under a purely theoretical point of view, the crucial question is to decide whether a particular variable can be considered stationary or not. No right or wrong judgment should be made, only appropriate or inappropriate (Isaaks and Srivastava, 1989). In practical situations, the variogram is only used up to a certain distance. The problem is to decide which series of moving neighborhoods can be found with a constant expected value and enough data to give meaningful estimates. The quasi-stationary assumption is, thus, a homogeneity scale and sampling density compromise.

In Fig. 2, it can be appreciated a clearly increasing trend along the distance, a non-stationary situation.

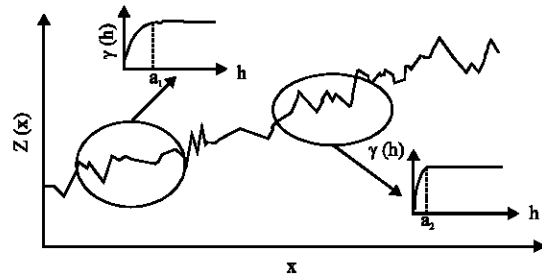


Fig. 2: The long trend versus short range stationary (Soares, 2000) where a_1 and a_2 represents the range until no spatial autocorrelation, $\gamma(h)$, can be found between two samples separated by h lag distance

However, looking at both blow-ups of the central section, fluctuation appears to be stationary in the short run (constant mean with the local search neighborhood). The assumption of its being quasi-stationary does not apply to the entire dataset but only to the search neighborhood under which the estimation model is fitted. These sub-areas meet the quasi-stationary assumption. One common way to test this local stationary pattern is the variogram cloud: selecting pairs of points with large variance at small distances reflects mean and variance discontinuity (Soares, 2000). The mean-variance computation in a moving window context can also help to find spatial heterogeneity.

If there is a spatial trend, then Ordinary Kriging (OK) variogram is no longer the appropriate tool to model spatial patterns. Some alternatives are possible: (1) Universal Kriging where the trend and the residual local are estimated at the same time, taking into consideration the restricted neighborhood of Kriging stationary; (2) stratified regionalization; (3) simple Kriging with varying local mean where interpolation are based on secondary information and external drift (residuals of the secondary linear variable function) and (4) to model the global trend component using a multi-linear coordinate regression to remove the drift component before OK is carried out on autocorrelated residuals. However, if data residuals are not normally distributed, then interpolation will be biased (Beckers, 1998). Because this detrended process requires parameter estimation, to detrend with lower-order global polynomials is, hence, recommended. Goovaerts (1997), for instance, considers this approach unreliable because the estimated residuals strongly depend on the algorithm used to estimate the trend component. For this researcher, OK with a local search neighborhood achieves better outcomes. Theoretically, the main reason to detrend within Kriging is to satisfy stationary assumptions although, quite often, the stationary decision for the

entire dataset is taken because the GIS user cannot afford to split data into smaller subsets.

EXPLORATORY (A)SPATIAL DATA ANALYSIS

A first step with Kriging interpolation is assessment of univariate and bivariate descriptive indexes such as distribution graphs (QQ and PP plot, cumulative frequency table and histogram), variance, inter-quartile range, skewness, coefficient of variation, kurtosis, scattergram and Pearson linear regression. Univariate analysis explores each variable in a data set, separately. Briefly, it describes each variable on its own. For instance, the skew refers to the shape of the values distribution of any variable. If the values of the observations are distributed symmetrically around the mean of a variable, that is called a normal distribution. In this case, the mean, median and mode will all coincide. Yet, if the values of most of the observations are lower than the value of the mean, then the distribution is called a negatively skewed or left skewed distribution. In this case, the mode will have a lower value than the median and the mean will have a higher value than the median.

Bivariate statistics tries to highlight the possible correlation between two variables. Either way, both practices treat data as a set of numbers ignoring its spatial nature resulting in a potentially critical loss of information (Glick and Hirsch, 2010). Hence, spatial description (mathematical techniques that includes spatial location and topology of variables) such as nearest neighborhood analysis try to fill this gap by exploring the spatial component such as data posting, flux movement maps, cartograms, choropleth mapping and nearest

neighborhood analysis. Familiarity with the dataset is an asset in any statistical study (Isaaks and Srivastava, 1989).

Figure 3a-c show the Pb study variable for a particular region of 4500×2950 m by showing the histogram, distribution function, Box-plot and main univariate exploratory statistics. Its major evidence relies on the fact that soil content of this contaminant follows a positive asymmetric distribution clearly non normal, leading to the following question: Should the GIS user standardize these values to obtain a Gaussian distribution or should he/she maintain the original ones?

DISTRIBUTION STANDARDIZATION

A major concern regarding random spatial processes is to confirm the Gaussian spatial distribution assumption. Kriging, as a spatial predictor, does not require that data follows a Normal distribution because Kriging is considered the best linear unbiased predictor. However, if data pursues a Normal one then Kriging becomes the best unbiased predictor of all (ESRI, 2001). Definitely, normality of input data is important to obtain interpolation maps with a strong sense of security. If sample values come from a lognormal distribution, for instance, the variance assessment from raw untransformed data is synonymous with outright stupidity (Clark and Harper, 2000). The lognormal distribution variance is a skewness measure, not variability and if the goal is to measure changeability some transformation should be produced for a better-behaved distribution, particularly for a very small sample size.

Data transformation often makes variance to be more constant and allows the correction lacking of

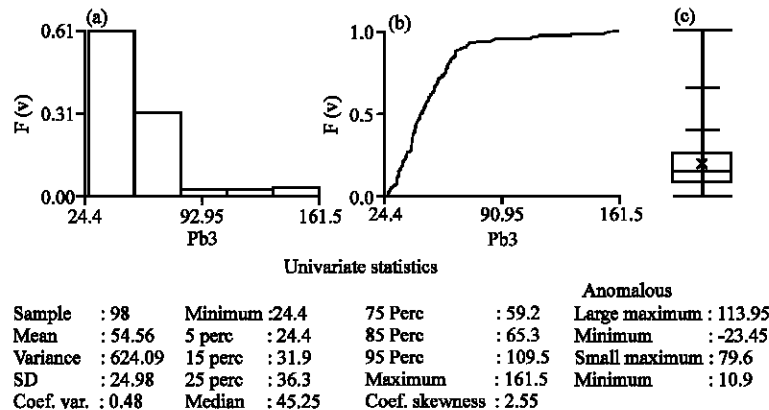


Fig. 3: Descriptive statistics regarding the Pb rhetorical case. (a) Histogram, (b) Distribution function and (c) Box-plot

homoskedasticity (Goovaerts, 1999). In addition, the multiple linear regression model states clearly that the relationship between each independent and dependent variable should also be linear and it can be achieved with data standardization (Clark and Hosking, 1986). Hence, the farther the random distribution is far from the Gaussian one, the worse the optimal linear prediction performs. The Gaussian model is unique in statistics for its extreme analytical simplicity and for being the distribution limit of many analytical theorems, globally known as central limit theorem.

The easiest way to check if samples respect the ‘bell’ curve is given by the cumulative probability plot or fractal diagram: the closer the data values are to the 45° line, the better. An additional method is to verify the overall similarity shape between the Gaussian distribution and the sample histogram. Another classical goodness-of-fit test is the χ^2 statistic based on the frequency distribution of classes, as is shown in Eq. 3 being O_j the observed value of class j and E_j the expected value while nc equals the number of classes considered:

$$\chi^2 = \sum_{j=1}^{nc} \frac{(O_j - E_j)^2}{E_j} \tag{3}$$

If the resulting statistic is less than $\chi^2(p = 95\%, df = nc-3)$ critical value, for instance, then the rejection of the null hypothesis becomes true and the observed distribution does not differ from the Normal distribution. The Anderson-Darling, the Ryan-Joiner, the Shapiro-Wilk or the Kolmogorov-Smirnov statistics are other possibilities.

If the histogram is highly positively skewed, then the trick is to persuade samples to act normally (Murteira, 1993). But positively skewed distributions do not imply a lognormal transformation. If data is only skewed, the probability logarithm transformation plot will be curved, kinked or just weird revealing the presence of two or more populations. This may be indicative of different generating events. However, the second order stationary assumption forces data values to be viewed as an outcome of a random function model that has the same mean and same covariance function between pair of values that are h distance away from each other. When more than one population appears to be present in the spatial dataset, it is common practice to look for these multiple population reasons. According to Goovaerts (1997), division of the dataset by soil domains is a possibility.

The straight line of the cumulative probability plot on a logarithmic scale requires a logarithmic transformation adjusted by an additive constant, if necessary. Yet, the issue is that GIS users do not really care what the logarithm mean is. Researchers are only concerned about knowing what the average reef thickness is (Clark and Harper, 2000). However, the lognormal distribution estimation is quite a complex procedure based on Sichel tables. For instance, the Sichel optimal estimator for the arithmetic mean of the lognormal population is as follows: $mean_back_unbiased = e^{\log(\overline{mean})} \gamma_n(v)$, where $\gamma_n(v)$ is the bias correction factor based on natural logarithm variance and n equals the total number of samples. It is imperative to stress that the direct exponential back-transform is biased since the logarithm mean is the logarithm of the median and not the mean. When anti-logged, large values will become exponentially larger while small values will become exponentially smaller and cannot go below zero (Clark and Harper, 2000).

The over and under estimations will no longer cancel out when back-transformed. Therefore, it is not possible to write the regression relationship as $e^{(y)} = e^{(b_0 + b_1x)}$ and expect unbiased answers. As confirmed by Clark and Harper (2000), the best back-transformed estimate would be $e^{(y)} = e^{(b_0 + b_1x + 0.5\sigma^2 + 0.5\sigma^2)}$. For Cressie (1993) back transformation can be accomplished with the following formula, if Lognormal Kriging (LK) is used: $z^*(x) = e^{y^*(x) + 0.5\text{var}(x) - \Psi + 0.5\psi(0)}$, where $z^*(x)$ is the original back-transformed estimate at site x , $y^*(x)$ signifies the LK estimate, $\psi(0)$ represents the nugget-effect, $\text{var}(x)$ equals LK variance and Ψ represents the LaGrange multiplier. Confirming this view, Keranchenko and Bullock (1999) showed from 30 agriculture fields in Illinois, Indiana and Iowa, USA, that for lognormal data, LK produces better results than Ordinary Kriging.

However, the log transformation advantage disappears with an increase in dataset size, particularly with more than 200 data points (Beckers, 1998). In addition, LK plot suggests biased estimates for datasets with a low coefficient of variation (Beckers, 1998). For logarithmic data, Cressie (1993) supports the spatial median polish procedure for detrend and relative variograms, while Olivier (1996) recommends Hermite polynomials for Disjunctive Kriging (DK) or the Normal Score Transformation (NST). This last option works with its equivalent rank from a Normal distribution. After making predictions on the transformed scale, a back-transformation is, then, needed. The goal of NST is to make all random errors for the whole population normally distributed. Therefore, the NST transformation for DK must occur after detrending since the variogram is

calculated on residuals after a trend correction. Controversially, citing D'Agostino et al. (1998), we have accepted the bias introduced by the skewed data because log transformation before the final exponential Kriging estimate would result in another bias and we see no reason to prefer this latter bias.

SAMPLING CONSIDERATIONS

Statistically, samples should be independent, representative of a very large population and truthful. Regarding probabilities properties with two independent events, $P(e_1, e_2) = P(e_1) \times P(e_2)$ and $P(e_1 \cup e_2) = P(e_1) + P(e_2)$, geographical sampling is rarely concerned with events that are not mutually exclusive. If sampling with replacement constitutes a random sample, it is common practice not to evaluate the same element twice due to time and money reasons and to evaluate as many elements as possible. For instance, Fig. 4 presents a reflex of this strategy. It can be seen that Pb soil content follows a preferential East-West spatial continuity, an anisotropic phenomena.

In addition, samples like people or animals are not stationary in space but, when surveyed, they are considered to be for a particular instance of time. In fact, the measurements collected at different time's t_i can be interpreted as a particular realization of temporally random functions k but spatially correlated (D'Agostino *et al.*, 1998). Hence, data are assumed to come from the same distribution with no significantly different statistical properties. When dealing with the nitrate concentration measures of May (47 wells), July (28 wells) and November (27 wells) for the Luca Plain, Italy, D'Agostino *et al.* (1998) use CoKriging (CK) and reach an improved spatial distribution estimation variance (16% on average) when compared to Ordinary Kriging.

However, if spatial variability and process scales do not match, then it is complicated to combine various datasets for spatial analysis since the measurement of the sampling interval must capture the same distribution nature for the spatial phenomena. In practical terms, the variogram should be the same (Soares, 2000). According to Burrough and Mc Donnell (1998), the survey of fifty points taken in the Netherlands to measure soil surface temperature was carried out using two different supports: a Daedalus airborne multispectral scanner and a Heimann ground radiometer. It became clear that the two variogram structures show little correlation to each other. As expected, spatial variation varies inversely with sampling support size: samples hold greater variances than blocks while their distribution is less symmetric, making the support effect another conditional bias for incorrect

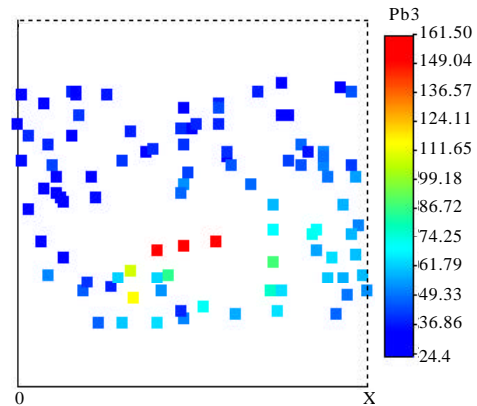


Fig. 4: Example of spatial distribution concerning the 98 samples of Pb

threshold values. Regarding variogram regularization, Armstrong (1998) demonstrates that the variogram of a new regularized variable is the corresponding variogram of other available supports ($\gamma_{2,2}(h)$, for $2m \times 2m$ and $\gamma_{3,3}(h)$, for 3×3 m) plus the Journel and Huijbregts (1978) table factor. The affine correction or indirect lognormal corrections are other possibilities for this upscaling operation (Isaaks and Srivastava, 1989).

Concerning sampling size, no rule could be provided although, quite often, time and money dictate sample size. Nevertheless, some recent studies have shown that it is possible to compute the final estimation uncertainty (e.g., mean expected value for the variable to estimate) using both confidence intervals and theoretical reliability calculation by means of a non-parametric approach where no particular distribution is assumed (Aguilar *et al.*, 2007a, b). At last, the common sense dictates that the greater the available samples the more reliable are the final estimations. As a rule of thumb, a sample size of at least 30 is employed in most situations (Griffith, 1996), although the central limit theorem guarantees normality for a large number of observations. Gaudard *et al.* (1999) argued that Bayesian methodology reduces the number of sample points required for the estimation of a reliable variogram but paid off by the greater computational complexity.

But in a spatial context, the problem is even larger because spatial sampling distribution can be critical for the analysis that follows. Clustered samples will not improve reliability as dispersed ones do. In fact, Ordinary Kriging (OK) variance takes this factor into consideration. Looking more closely at the OK variance formulae (Eq. 12 and 13), there is a system balance between getting the samples close to the estimated one and keeping them as far from one another as possible. While the second term calculates how closely the samples are related to one

another, the last one assesses the relationship between samples and estimations. Hence, sample density can turn out to be only a rough guide to quality interpolation.

Uniform or systematic sampling imposes a limit on the minimum distances among points, the distance interval in which capturing spatial dependency is most important. Hence, random location provides better estimates on the variogram for distances close to zero but yielding a less representative depiction of a surface (over and under samples for different sub-regions). According to Griffith and Layne (1999), if a choice is to be made between these two sampling types, then the systematic one is to be preferred as a more accurate surface representation and reflected on the Kriging prediction errors.

Isaaks and Srivastava (1989) demonstrated that additional sampling is most necessary in anomalous areas since it improves estimates where the proportional effect makes them least accurate. Even so, variograms with a proportional effect associated with clustered samples might not work well either, due to the apparent hole-effect, a dip variance at distances greater than the range. Certainly, a major variogram hard assumption is that only one single dominant correlation scale is assumed. In addition, heterogeneity on a larger scale is not apparent while heterogeneity on small scales is reflected in the nugget-effect. In this sense the variogram is a middle-aged man with difficulty seeing close objects and far away ones.

Groenigen *et al.* (1999) also report simulated annealing optimization of sand sampling at Ping, Thailand, river. By taking into account soft information and variogram shape, the maximum Kriging variance decreased from 146 to 69% with more observations along the boundaries and smaller spacing in the shortest-range direction.

THE VARIOGRAM

The variogram quantifies Tobler's Law on all scales by summarizing the degree of similarity between data values for all possible data pairs as a distance function. Most variograms show a behavior like the one presented in Fig. 5. The upper limit is the sill which implies no spatial dependence between data points because all variances are invariant with the sample separation distance. If the variogram has a sill then this behavior tends to support the assumption that samples come from a stationary distribution with a fixed mean and standard deviation (Clark and Harper, 2000). It is also worth noting that for stationary variables, as $h \rightarrow \infty$, $\gamma = \sigma^2 = C_0$

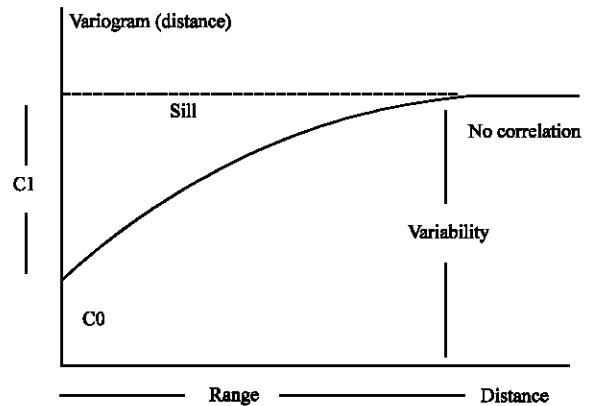


Fig. 5: Typical behavior of a transitive variogram

(Nugget-effect)+ C_1 (Partial Sill)=Total Sill. The percentage ratio between the nugget-effect and total sill is called the relative nugget-effect.

Sill fluctuations may also correspond to periodic structures. In addition, sill rescaling by 10 or 100 would not change prediction values (Kriging Description, 2000). Maps would still look the same although changes would be reflected in prediction variances: the higher the sill, the higher the prediction variances.

The separation distance at which samples are spatially auto-correlated is the range that gives the answer to the short scale stationary question posed in Fig. 2 (how large the search window should be, a local neighborhood stationarity issue). If the distance between the estimation and the sample is greater than the range, then those samples make no useful contribution to final interpolation: it is too far. In theory, its weights are zero. Stein and Zaadnoordijk (1999), samples deletion beyond the range is inappropriate even when those values suffer from the screen-effect, that is, the closest data tend to cover the influence of those further away. As expected, stronger spatial autocorrelation increases the range of the variogram and provides a fairly smooth map since Kriging weights become less powerful with respect to closer samples. On the contrary, as the range becomes shorter, more variation is expected. As Olivier (1996) reported that argues, bounded variograms imply transition features, patches or zones whereas unbounded ones suggest continuous change over a region (the use of the variogram as a spatial search pattern process).

The nugget-effect, C_0 , represents the measurement error variance and the spatial variation at distances much shorter than sample spacing, which cannot be resolved (Liebhold *et al.*, 1993). Nugget-effect measures the unreliability of the sample value and the lack of a micro-scale model. Confirming this belief, GASA

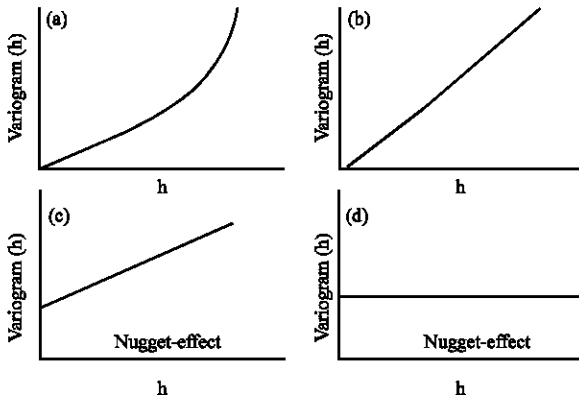


Fig. 6: (a-d) The four type of behavior near the origin

states that the nugget-effect is part of the variation that we cannot predict because it derives from spatial measurement error (a non-zero nugget-effect indicates repeated measurements at the same point with different outcomes, a precision issue) and purely random variation. Theoretically, the variogram is strictly zero for zero distance. If we do not accept zero at zero then what we are basically saying is that we do not believe in our sample values (Clark and Harper, 2000). Controversially, once a uranium sample has been drawn, it is not there anymore for another replica. As well, distinguishing an outlier cluster and a non-stationary pocket or a sampling variation from a true trend can lead to a problematical situation. If Universal Kriging (UK) handles true trend, Block Kriging (BK) applies to noisy circumstances because of lower variance values.

According to Fig. 6a-d, there are four types of behavior regarding C_0 : Quadratic (i) - The regionalized variable is highly continuous and differentiable with the presence of a drift. Linear (ii) - The regionalized variable is continuous but not differentiable and, thus, less regular. Discontinuous (iii) - The variable is highly irregular at short distances (the higher the nugget-effect, the higher the Kriging variance). Flat (iv) - Pure randomness or white noise, that is, the regionalized variables $z(x)$ and $z(x+h)$ are uncorrelated for all values no matter how close they are. This is the extreme case of a total lack of structure where the best estimate for an unknown site would be the population average.

Although insufficient data may create a deadlock situation for the variogram estimation, Aunón and Hernandez (2000) confirmed the creation of local variograms for specific areas according to their local mean. By using local models whenever possible, they demonstrate how stratification reduces interpolation errors based on prior information (qualitative maps, for instance). But should anyone risk an inappropriate but well defined global variogram in preference to several

poor local ones? The recognition of different populations might not also contribute to the study goal, either. For instance, if the fluid flow in a petroleum reservoir within several rock type environments is the main research aim then this last distinction will not contribute to the flow properties analysis. Further, sudden changes at the strata boundaries can be found because available data are based on two different sets (Goovaerts, 1997).

In Eq. 4 is depicted a weighted variogram. If stationary assumption is at risk, then relative variograms for each separate population with a constant local mean can be found by this formula where $\gamma_1(h), \dots, \gamma_n(h)$ are the local variograms from n regions, m_1, \dots, m_n are the local region mean, while $N_1(h), \dots, N_n(h)$ are the number of sample pairs from each region:

$$\gamma_w(h) = \frac{\sum_{i=1}^n N_i(h) \frac{\gamma_i(h)}{m_i^2}}{\sum_{i=1}^n N_i(h)} \tag{4}$$

Another option, presented by Isaaks and Srivastava (1989), is the pairwise relative variogram (Eq. 5), where the denominator serves to reduce the influence of very large differences. Note that if $z(x_i)$ and $z(x_i+h)$ are zero then dropping that calculation for the next pair is the solution. Geostatisticians may also use the covariance function ($\gamma(h) = C(0) - C(h) = \sigma^2 - C(h)$), the madogram and the correlogram model ($\gamma(h)/\sigma^2 = \rho(0) - \rho(h) = 1 - \rho(h) = 1 - C(h)/C(0)$).

$$\gamma_{PR}(h) = \frac{1}{2|N(h)|} \sum_{i=1}^{N(h)} \frac{[z(x_i) - z(x_i+h)]^2}{\frac{z(x_i) + z(x_i+h)}{2}} \tag{5}$$

Disregarding the variogram choice, the removal of the most erratic values can be a plus even if this might be a questionable solution. In mining excavation, an unusually large value may be a cause for celebration (Wang and Zhang, 1999). If outlier removal is crucial in generating negative spatial autocorrelation, then small high-grade deposits can equally make the difference between opening the mine and keeping it closed. Moreover, with high-skew distributions such as with gold grades, it is not possible to cut them in an arbitrary way.

Cressie (1993) proposed an OK resistant algorithm by using neighboring values to determine how much of the outlier should be down weighted. Data standardization is another option although this may create data artifacts. For instance, if a variable holds a zero value and a logarithm transformation is applied then a problem may emerge. A plausible solution may be to set 0 to 0.00000001. Even so, either decision may or may not change the variogram shape, depending on the variable range context.

If high-grade samples are close together in a specific sub-area, eliminating those samples may not be a good idea either. The model variogram is the spatial averaging result of obscuring regional variations and differences caused by local anomalies (Isaaks and Srivastava, 1989). Therefore, it is essential to examine spatial contiguity and short-range differences before modeling variograms with first-class spatial exploratory software, including spatial autocorrelation and association measures.

Another critical issue is anisotropy. That is, a different spatial autocorrelation behavior of the variogram according to several directions versus trend influences due to a physical process (e.g., wind, runoff or a geological formation). Anisotropy is a random process feature that shows higher autocorrelation in one direction than in another (ESRI, 2001). Knowledge of phenomena genesis can be helpful to detect this last constraint. If global trend affects all measurements due to a known physical process in a deterministic global polynomial framework, anisotropy differs from drifts since the directional influence process is not usually known. Thus, it is quantified as a random error process with directional autocorrelation variograms. This factor affects circle shape regarding the local search neighborhood window, as well.

Major literature recommends computation in four directions to divide the omni-directional search into four subsets: 0° (North-South), 45° (North East-South West), 90° (East-West) and 135° (North West-South East). The ranges of each of them can confirm or refute the assumption of the anisotropy. When plotting a rose diagram, for instance, circular contour lines are indicative of an isotropic surface, whereas an elliptic shape points to an anisotropy occurrence. Another option is to draw the contour map of the variogram surface. As expected, more weight will be given to samples that lie in the direction of maximum continuity and considerably less where minimum continuity is involved.

Concerning elliptic or geometric anisotropy (Fig. 7a, b), the variogram holds the same sill for major directions although their ranges are different. Then, the overall variogram becomes $\gamma(h) = \gamma(h_x, h_y) = \gamma_1(\sqrt{h_x^2 + r_{ani}^2 h_y^2})$ where, h_x and h_y are the cosine and sine of direction h , a_1 and a_2 equal the length of the major and minor axes of the resulting ellipse, while $r_{ani} = (a_2/a_1)$ represents the amount of survey space that has to be stretched in direction h so the variogram in direction h_2 is the same as that in direction h_1 .

Occasionally, the variogram exhibits different ranges and sills, thus creating an overall variogram division into two linear components (Isaaks and Srivastava, 1989):

$$\gamma(h) = w_1\gamma_1(h_1) + w_2\gamma_1(h_2)$$

Where:

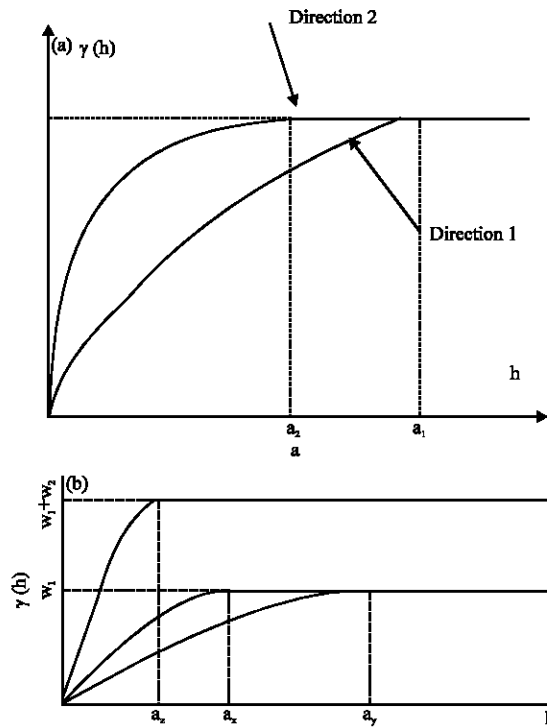


Fig. 7: (a) Geometric and (b) geometric with zonal anisotropy (Soares, 2000)

$$\gamma_1(h_1) = \gamma_1\left(\sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2 + \left(\frac{h_z}{a_z}\right)^2}\right)$$

whose sill equals w_1 with a range of 1.

$$\gamma_1(h_2) = \gamma_1\left(\frac{h_z}{a_z}\right)$$

is the equivalent model whose sill equals w_2 and only exists in z direction.

Modeling anisotropies is done by trial and error using a graphic terminal (Soares, 2000). For Armstrong (1998), it is not a good idea to apply least squares or other automatic methods to fit variogram models. First, the model must be a positive definite function to guarantee having one and only one stable solution. After all, the Kriging matrix should have an inverse. Polynomials obtained by least squares regression will rarely satisfy these conditions. Second, least squares assume that sample points are independent observations, which is clearly not true. Third, critical variogram behavior close to the origin is not known and least squares do not take this into account. Outlier adjustment can also be complicated. In addition, it does not take into account additional information that may justify the choice of one model instead of another.

For automatic fitting, Olivier (1996) supported weighted least square (WLS) using the Ross Maximum Likelihood Program whose weights are directly proportional to the number of paired comparisons in the estimates. Goovaerts (2000) used the minimally weighted sum of the squares of the differences between the experimental variogram and the theoretical one. Clark and Harper (2000) suggested Cressie's (1993) modification of the automatically fitting modeling by replacing the scaling division of the total number of pairs (Eq. 6), a situation implemented within EcoSSe®. Basically, this formula quantifies the squared difference between the theoretical, $\gamma(h)$ and experimental, $\gamma^*(h)$, variograms, but scaled by the model itself: the lower this index becomes, the better the model turns out to be. In addition, each point is weighted according to the number of pairs available: the more samples available, the smaller the error that is tolerated.

$$\text{Automatic_fitness_index} = \frac{1}{\sum |N(h)|} \sum N(h) \left[\frac{\gamma^*(h) - \gamma(h)}{\gamma(h)} \right]^2 \quad (6)$$

A last step concerns the creation of a mathematical model that fits the quantitative regionalized variation description so that autocorrelation can be characterized. The ultimate purpose is to obtain the covariance values to setup the A and C matrices within equation $A \times B = C$. Therefore, the model variogram formula is critical to the interpolation quality. It is imperative that this model fits experimental values well and avoids negative variances, the conditional negative definite property (Armstrong, 1998).

The next three equations present the common three admissible transitive variogram models. Other specialized literature mentions other models such the cardinal sine, wave-hole, intrinsic power function, nugget-effect, prismatic-magnetic, Cauchy, Bessel, cubic and prismatic-gravimetric model, but they are not reviewed here. Non-transition variogram models used to model intrinsic random functions are not enclosed either.

$$\text{Spherical model: } \gamma(h) = C_0 + \left[C_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \right] \text{ if } 0 < h \leq a \quad (7)$$

$$\text{Exponential model: } \gamma(h) = C_0 + C_1(1 - e^{-3h/a}) \text{ for } 0 < h \text{ (} a' = 3a \rightarrow \gamma(a') = 0.95(C_0 + C_1) \text{)} \quad (8)$$

$$\text{Gaussian model: } \gamma(h) = C_0 + C_1(1 - e^{-3h^2/a^2}) \text{ for } 0 < h \text{ (} a' = a\sqrt{3} \rightarrow \gamma(a') = 0.95(C_0 + C_1) \text{)} \quad (9)$$

As can be observed in Fig. 8-10, exponential model initially raises more rapidly, leading to fuzzier images (less

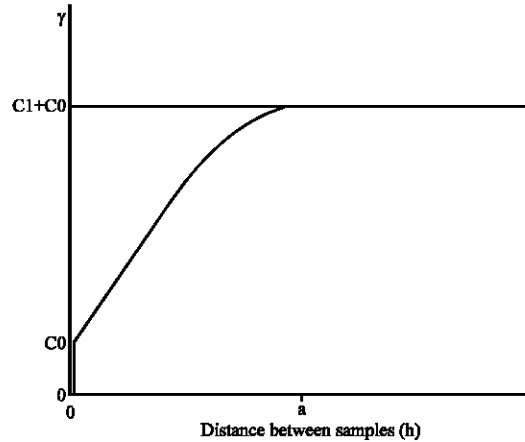


Fig. 8: The spherical model where C_0 is the nugget-effect, C_1+C_0 denotes the total sill of the considered variogram, $\gamma(h)$, while a represents the effective range

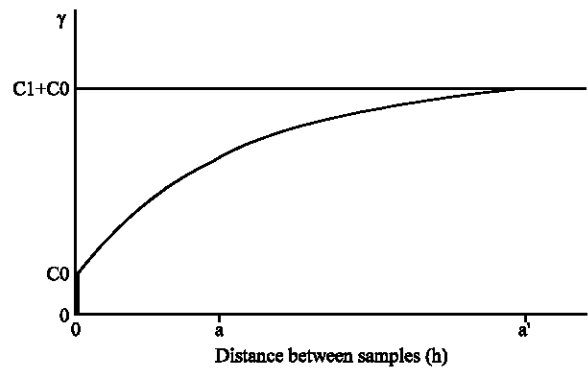


Fig. 9: The exponential model where C_0 is the nugget-effect, C_1+C_0 denotes the total sill of the considered variogram, $\gamma(h)$, while a represents the effective range

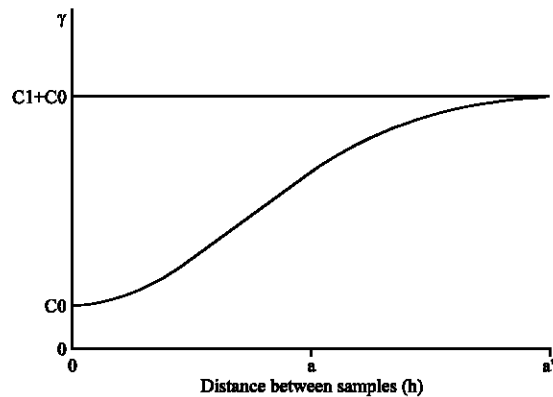


Fig. 10: The Gaussian model where C_0 is the nugget-effect, C_1+C_0 denotes the total sill of the considered variogram, $\gamma(h)$, while a represents the effective range

structured) than Gaussian one because it is linear near the origin. Gaussian variogram reaches about 66% of its height at the square-root of its range. Yet, according to Englund and Sparks (1991), this model does not work well, particularly when it has a low nugget-effect and/or closely spaced data.

Regarding computational issues, any experimental variogram must be aware of observations that layout in an irregular network. This is the reason for the associated tolerance angle. If the two major directions are perpendicular, then the angular tolerance should be less than 45° to avoid points overlapping. It is also crucial that samples with mistaken measurements should not be substituted by zero because it would distort the true variability. Olivier and Webster (1990) suggest between 150 and 200 observations to produce a reliable variogram while Englund and Sparks (1991) presented some other practical tips to improve a variogram: (a) 30 to 50 pairs at each lag distance (between 50 to 100 for Burrough and Mc Donnell (1998) and at least 30 for Journel and Huijbregts (1978)), (b) Similar lag space should be applied for different anisotropy axes, (c) Data transformation for skewed distributions, (d) An omni-directional variogram (angular tolerance of 90° degrees on either side) before reducing into directional ones. In fact, an omni-directional variogram contains more sample pairs than any directional one and is, therefore, more likely to show a clearly interpretable long-range structure, (e) Try several different lag intervals for plotting the experimental variograms to obtain the maximum detail at small distances without being misled by structural artifacts due to the particular interval used.

For Isaaks and Srivastava (1989), the search radius for irregular grid samples should be slightly larger than the average spacing between samples although additional sampling will always improve the estimation regardless of the distance from the point being estimated. According to Griffith and Layne (1999), the tolerance angle should be 23 degrees. Moreover, large lag tolerance leads to nugget-effect overestimation and masked short-range spatial autocorrelation. On the other hand, small lag tolerances may generate many empty bins.

Figure 11 depicts the experimental spherical variogram of previous Pb dataset, according to East-West direction ($\theta = 90^\circ$). No nugget-effect was found while its range equals 2500m. The minor direction ($\theta = 0^\circ$) presents a range of 1300 m.

If samples are located within a regular grid, the grid spacing is the lag itself. Otherwise, according to Englund and Sparks (1991), the maximum pair should be divided by two and then subdivided by ten equal distance classes. The lag tolerance should be half of the lag spacing

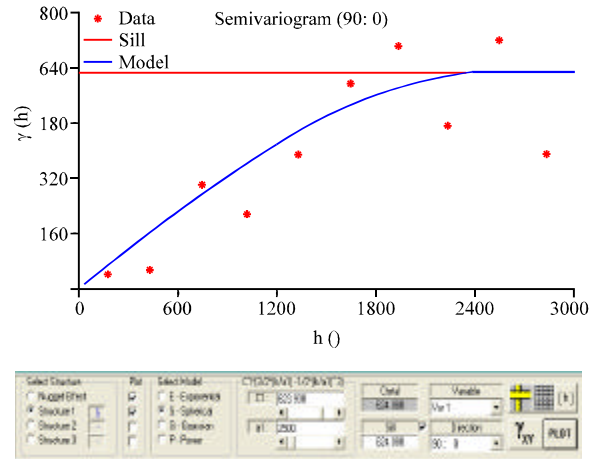


Fig. 11: Experimental variogram obtained for Pb dataset and estimates of some structural parameters of GeoMS®

(Geostatistics FAQ, 1997). For Armstrong (1998), this tolerance should not be too large, in order to avoid the anisotropy blurring and not too small, to avoid a diminutive number of pairs.

ORDINARY KRIGING

The difference between Inverse Distance Weighted (IDW) and Kriging derives from the mathematical process that minimizes the minimum squared error variance of the estimation. The OK equations system can be obtained from the following two algebraic expressions (Soares, 2000) where w_j are the weights to be assigned to the i th observation, Ψ represents the Lagrange multiplier or slack value required for forcing total weights to 1 when the partial derivatives system, set to minimize the estimation variance is resolved. $\gamma(x_i, x_j)$ is the variogram value between sample i and j , while $\gamma(x_i, x_0)$ equals the variogram value between sample i and the estimation x_0 :

$$\sum_{j=1}^n w_j \gamma(x_i, x_j) - \Psi = \gamma(x_i, x_0), \quad i = 1..n \quad \text{and} \quad \sum_{j=1}^n w_j = 1 \quad (10)$$

Like Thiessen and B-Splines, Kriging is an exact interpolator in the sense that the sample and estimation are equal. This happens because the variogram value between the sample k and the estimation x_0 , $\gamma(x_k, x_0)$, equals zero since the estimated site is the sample itself. Therefore, its Kriging variance equals zero. By resolving the Kriging system presented in Fig. 12, weight w_k is one while all others weights and the Lagrangean dummy parameter becomes zero (row k of matrix B changes to zero if the nugget-effect is zero).

Once again, the goal is to find the weighted average estimator that provides an unbiased estimator with the smallest estimation error variance using the Lagrange multiplier and the weights sum to one constraint. Assuming an unknown constant mean and an intrinsic variable (the underlying idea is to work with increments rather than with the variable itself), the general arrangement to setup the Matheron equations are as follows:

$$\begin{aligned}
 w_1\gamma(x_1, x_1) + \dots + w_n\gamma(x_1, x_n) - \Psi &= \gamma(x_1, x_0) \\
 w_1\gamma(x_2, x_1) + \dots + w_n\gamma(x_2, x_n) - \Psi &= \gamma(x_2, x_0) \\
 \dots & \\
 w_1\gamma(x_m, x_1) + \dots + w_n\gamma(x_m, x_n) - \Psi &= \gamma(x_m, x_0) \\
 w_1 + \dots + w_n &= 1
 \end{aligned}
 \tag{11}$$

In an elegant matrix layout, each interpolated value is calculated as the sum of weighted known points whose weights are calculated from the (n+1) simultaneous linear equations set: $A \times W = B$ or $W = A^{-1} \times B$. The statistical distance (it incorporates anisotropic distance, direction, spatial autocorrelation and clustering information) between n samples and distances from each sample to the grid point are used to compute the model variance reproduced on matrices A (among samples) and B (between each sample and the estimated location). While A^{-1} underlies the declustering factor, B represents the structural distance between the estimation and all samples. In addition, the product of A^{-1} by B adjusts the

$$A = \begin{bmatrix} \text{Cov}(x_1, y_1) & \dots & \text{Cov}(x_1, y_n) & 1 \\ \dots & \dots & \dots & \dots \\ \text{Cov}(x_n, y_1) & \dots & \text{Cov}(x_n, y_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} \text{Cov}(x_1, y_0) \\ \dots \\ \text{Cov}(x_n, y_0) \\ 1 \end{bmatrix} \quad W = \begin{bmatrix} w_1 \\ \dots \\ w_n \\ \psi \end{bmatrix}$$

Fig. 12: The Kriging system in terms of matrices

raw inverse statistical distance weights of matrix B to account for possible redundancies between samples. As expected, if no spatial autocorrelation is found among the available samples, then the Kriging estimator equals the sample average. After a major geocomputation, an interpolated surface map can be achieved such as the one shown in Fig. 13 for our example corresponding to Pb distribution pattern.

With zero nugget-effect, the diagonal of A matrix equals zero while Kriging weights decrease as the dataset location gets farther from the estimated one. Hence, major software computation uses the covariogram model to setup A and C matrices in order to avoid the diagonal zeros of A. After weights are calculated, the estimation of the regular grid surface can be achieved with a great deal of calculation. In fact, the computation intensity to solve OK weights is proportional to the cube of the number of observations, although a reduction can be achieved by applying the sequential Kriging approach (Guzmán and Yeh, 1999).

The uncertainty layout of the conventional Kriging software is closely related to OK variance and given by the following equation where C_{00} is the variance of the estimated point value (as the variable becomes more erratic, this term increases in magnitude), C_{ij} is the covariance between the *i*th and *j*th sample (as samples get closer together, a clustering issue, the average covariance increases), w_i and w_j are the OK weights while C_{i0} represents the covariance between the *i*th sample and the unknown value being estimated (uncertainty decreases as samples are closer to the value being estimated):

$$\sigma_e^2 = C_{00} + \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} - 2 \sum_{i=1}^n w_i C_{i0} = -\gamma_{00} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} + 2 \sum_{i=1}^n w_i \gamma_{i0}
 \tag{12}$$

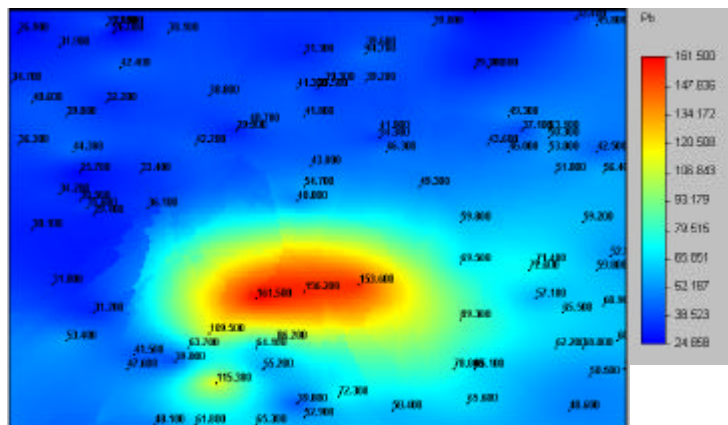


Fig. 13: Ordinary Kriging estimation of the Pb dataset of GeoMS®

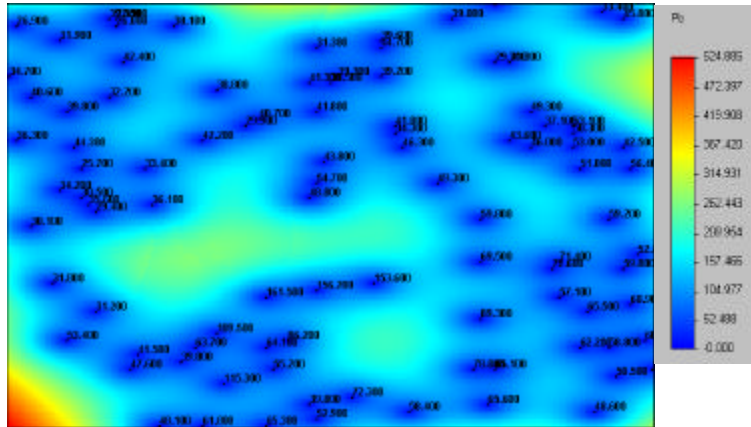


Fig. 14: Ordinary Kriging variance of the Pb dataset of GeoMS®

According to Soares (2000), if the sum of the weights is one and the average estimation error equals zero then the Kriging error variance (minimum estimation variance) for a general estimated point x_0 becomes as shown in Eq. 13, being n the number of neighbors intervening in the interpolation. Notice that no measurement error is included here:

$$\sigma_{OK}^2 = \sum_{i=1}^n w_i \gamma(x_i, x_0) + \psi \quad (13)$$

If errors respect the Gaussian curve then real values for a point x_0 will fall within the interval expressed in Eq. 14 for a 95% confidence level. This implies symmetry of the local distribution of errors, as well. Critically, OK variance is not sensitive to local error for two major reasons: (1) It is based on the same global variogram; (2) Distances among locations are the only relevant factor. Hence, OK variance is mainly a geometry dependent measure heading the assumption that an OK true error map is a better substitute. OK variance is too much of a spatial operation. Figure 14 shows the OK variance surface for our example dataset.

As expected, its values are higher when no samples are founded. A very important and maybe the biggest advantage of Kriging over other interpolation methods is that the estimation variance can be calculated before the actual sampling is made. This is true because the Kriging error variance depend on the variogram and the spatial configuration of the data points in relation to x_0 and do not depend on the observed values (Vieira *et al.*, 1981):

$$z(x_0)_{estimated} = \sum_{i=1}^n w_i z(x_i) \pm 1.96 \sigma_{OKx_0} \quad (14)$$

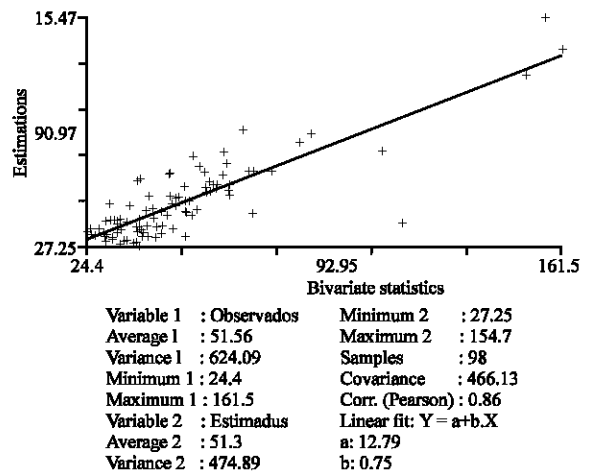


Fig. 15: The cross-validation layout of GeoMS®

THE CROSS-VALIDATION PROCEDURE

Before the final surface is produced, consistency and absence of biased estimations can be tested by temporarily dropping any sample from the dataset while its value is re-estimated using the remaining samples. Repeating this cross-validation procedure for all observations (Fig. 15), it is possible to obtain the experimental mean error:

$$\left(\frac{\sum (x_i^* - x_i)^2}{n} \right)$$

the root mean square prediction error:

$$\left(\sqrt{\frac{\sum (x_i^* - x_i)^2}{n-1}} \right)$$

and the average Kriging standard error:

$$\left(\frac{\sum \sigma_{ok}(x_i)}{n-1} \right)$$

If the average standard error is close to the root mean square prediction error then the user is assessing the prediction variability correctly. However, if the average standard error is greater or less than the root mean square prediction error then the user is over or under estimating estimation variability (ESRI, 2001).

It is hard to judge the influence of variogram parameters on cross-validation errors (Soares, 2000). In fact, Isaaks and Srivastava (1989) proved that if a relative nugget-effect of 4% lead to a lower cross-validation residuals for their 470 samples dataset, the original model performed better than the one whose nugget-effect was automatically adjusted on the basis of cross-validate residuals. As well, if the original dataset is spatially clustered then the cross validation pursues the same trend leading to a situation whose residuals become more representative of certain regions. Hence, a good statistical model does not signify that a good spatial genuine model is good. The QQ plot comparison layout between true and estimation with the linear correlation index or the residual map in order to check location bias are other common graphic alternatives.

UNCOVERED CONSIDERATIONS

Geographers and major GIS users do not think in spatial terms and do not ask questions about spatial pattern and spatial autocorrelation. The goal of most users is to handle large volumes of spatial data and produce beautiful maps rather than carrying out sophisticated statistical spatial analysis. For instance, Longley *et al.* (2001) argued that within specialized software there is a lack of spatial analysis techniques in general and of spatial statistics in particular. This may be a major impediment to GIS, which is a ‘data rich, theory poor’ environment. These authors put an emphasis on creativity, comparing spatial analysis to art and not to science. Applied geostatistics is a talent science because its results may be widely diverse depending on several factors such as the variogram modeling, the Kriging approach, the sampling design or the scientist’s previous experience.

Clearly, two methods emerge regarding spatial interpolation: the deterministic and the stochastic view. If the former has no capability for assessing uncertainty, the latter must be viewed as a process instead of a single function with an increase in computer power demand, particularly with Kriging.

Another relevant issue regards the discontinuity of the variogram origin. This can be viewed as a non-sense situation if spatial reality is considered a continuous surface. Among all variogram factors, the nugget-effect is also the most unpredictable because of the lack of close samples (Armstrong, 1998).

Another close issue is the assumption of precise and accurate sample values by considering the zero nugget-effect. It is quite common to have a discontinuity at the variogram origin although major Kriging software forces it to be zero at zero distance. On the other hand, if the variogram for zero distance equals total nugget-effect then the estimation Kriging variance becomes lower, leading to a nonsense situation.

The definition of the variogram model is also time-consuming and somewhat subjective operation. The variogram estimation includes a set of parameters to be setup by the expert user for implementation. According to Kelly (1996), there is little to be gained by making Kriging and spatial autocorrelation a main tool of spatial analysis unless the GIS user sophistication is sufficient for this information to be preferred. Major Kriging software offers a wide range of options but its interpretation requires a thorough background. Certainly, an essential foundation of GIScience is the availability of spatially-aware professionals. Finding the best variogram fitness procedure (interactive or automatic) can also be a problem.

A last issue concerns the structural analysis that lies behind Kriging. The variogram assumes that a surface holds a constant mean, a situation almost never true (Nalder and Wein, 1998). Spatial autocorrelation exists and samples follow a Gaussian distribution. The more severely these assumptions are violated, the less accurate interpolation results become. Further, outlier negative influence on the variogram and nonsensitivity to data values of Kriging variance are also real.

CONCLUSIONS

It is known that most spatial variables are a result of a vast number of processes that cannot be described quantitatively. However, Earth holds some spatial continuity patterns leading Geography to a new effort to create its own general space rules instead of begging methodologies from other sciences such as history or economics. Geography is the science of the differentiation of space based on its own laws (Longley *et al.*, 2001) to detect urban hierarchy, future industrial localizations, military targets accessibility, a hospital’s regional influence or agriculture patterns, for instance. Certainly, this narrow view of Geography emphasizes its major feature of its uniqueness in relation to other research

fields. According to Ferreira and Simões (1988), the kernel of Geography is to think geographically, that is, to study spatial distribution phenomena and their correlations.

New applied mathematical space formulations are becoming, thus, crucial. In this study, the Kriging interpolator is reviewed here in detail since, stating Griffith (1993), diagnostics are playing an increasingly vital role in spatial analysis. Kriging describes the best linear unbiased estimator in the sense of least variance. Kriging is BLUE (Best Linear Unbiased Estimator) and BUE (Best Unbiased Estimator if data respects the bell curve). When Kriging is compared with deterministic interpolators, there are major differences, e.g., the former provides uncertainty assessment, anisotropy detection or methodology assumptions. It is only hoped that this article provides a better understand of this stochastic process.

Regarding OK geocomputation, most spatial analysis computer programs are available on the 'World Wide Web' leading this platform to a new GIS one, as it is the main goal of SAKWeb[®] project (Negreiros, 2004). Ultimately, this world expects fast answers because if it works then it is obsolete. Surface[®] III, Regard[®], Ecosse[®], GeoEAS[®], GeoMS[®], GS+[®], Kriging for Berkley Unix[®], Surfer[®], GeoPack[®], GMS[®], Uncert[®], VarioWin[®], GSLib[®], SpaceStat[®] (former GeoDa[®]), Spring[®], ESRI Geostatistical Analyst[®] and Kriging ArcScripts[®] written in Avenue[®] are some possibilities. Whatever is the user choice, the respectability of the academic research concerning implementation should be a reality with GIS free platforms.

REFERENCES

- Aguilar, F.J., F. Agüera and M.A. Aguilar, 2007. Accuracy assessment of digital elevation models using a non-parametric approach. *Int. J. Geograph. Inform. Sci.*, 21: 667-686.
- Aguilar, F.J., F. Agüera and M.A. Aguilar, 2007. A theoretical approach to modeling the accuracy assessment of digital elevation models. *Photogrammetric Eng. Remote Sens.*, 73: 1367-1379.
- Armstrong, M., 1998. *Basic Linear Geostatistics*. Springer-Verlag, USA., pp: 153.
- Aunón, J. and J. Hernandez, 2000. Dual kriging with local neighborhoods: Application to the representation of surfaces. *Math. Geol.*, 32: 69-85.
- Beckers, J., 1998. Modeling the oro moraine multi-aquifer system: Role of geology, numerical model, parameter estimation and uncertainty. Ph.D. Thesis, Department of Earth Sciences, University of Waterloo, Canada.
- Burrough, P.A. and R.A. Mc Donnell, 1998. *Principles of Geographical Information Systems*. Oxford University Press, New York, pp: 333.
- Clark, I. and W.V. Harper, 2000. *Practical Geostatistics*. Ecosse North America LLC, Columbus, OH., USA., ISBN-13: 978-0970331700.
- Clark, W.A.V. and P.L. Hosking, 1986. *Statistical Methods for Geographers*. 1st Edn., John Wily and Sons, UK.
- Cressie, N.A., 1993. *Statistics for Spatial Data*. John Wiley and Sons, Ontario, Canada.
- D'Agostino, V., E.A. Greene, G. Passarella and M. Vurro, 1998. Spatial and temporal study of nitrate concentration in groundwater by means of co-regionalization. *Environ. Geol.*, 36: 285-295.
- ESRI, 2001. *Using ArcGIS Geostatistical Analyst*. ESRI, USA., pp: 40.
- Ella, V., 2000. Spatial analysis and simulation of selected hydrogeologic and groundwater quality parameters in a glacial till aquifer. Ph.D. Thesis, Iowa State University.
- Englund, E. and A. Sparks, 1991. *GEO-EAS 1.2.1 users guide*. United States Environmental Protection Agency.
- Ferreira, C. and N. Simões, 1988. *Tratamento Estatístico e Gráfico em Geografia*. Gradiva Publicações, Gradiva.
- Gaudard, M., M. Karson, E. Linder and D. Sinha, 1999. Bayesian spatial prediction. *Environ. Ecol. Stat.*, 6: 147-171.
- Glick, B. and S. Hirsch, 2010. An integrated spatial statistics package for map data analysis. Landover Hills, Maryland.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford, pp: 483.
- Goovaerts, P., 1999. Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma*, 89: 1-45.
- Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.*, 228: 113-129.
- Griffith, D., 1993. Advanced spatial statistics for analyzing and visualizing geo-referenced data. *Int. J. GIS*, 7: 107-123.
- Griffith, D., 1996. Some Guidelines for Specifying the Geographical Weights Matrix Contained in Spatial Statistical Models. CRC Press, USA., pp: 307.
- Griffith, D. and J. Layne, 1999. *A Case Book for Spatial Statistical Data Analysis*. Oxford University Press, Oxford, pp: 506.
- Groenigen, J., W. Siderius and A. Stein, 1999. Constrained Optimization of Soil Sampling for Minimization of the Kriging Variance. *Geoderma*, 87: 239-259.
- Guzmán, J.A.V. and J.C.J. Yeh, 1999. Sequential kriging and cokriging: Two powerful geostatistical approaches. *Stochastic Environ. Res. Risk Assess.*, 13: 416-435.
- Isaaks, E.H. and R.M. Srivastava, 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, pp: 561.

- Journel, A.G. and C.J. Huijbregts, 1978. Mining Geostatistics. 1st Edn., Academic Press, London, pp: 600, ISBN-10: 0123910501.
- Journel, A.G., 1996. Modeling uncertainty and spatial dependence: Stochastic imaging. *Int. J. GIS*, 10: 517-522.
- Kelly, M., 1996. Spatial Analysis and GIS. Taylor and Francis, USA.
- Keranchenko, A.N. and D.G. Bullock, 1999. A comparative study of interpolation methods for mapping soil properties. *Agron. J.*, 91: 393-400.
- Liebhold, A., R.E. Rossi and W.P. Kemp, 1993. Geostatistics and geographical information systems in applied insect ecology. *Ann. Rev. Entomol.*, 38: 303-327.
- Longley, P., M. Goodchild, D. Maguire and D. Rhind, 2001. *Geographical Information Systems and Science*. John Wiley and Sons, New York.
- Matheron, G., 1965. *Les Variables Régionalisées et Leur Estimation*. Masson, Paris.
- Murteira, B., 1993. *Análise Exploratória Dos Dados: Estatística Descritiva*. McGraw Hill, USA.
- Nalder, I.A. and R.W. Wein, 1998. Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. *Agric. For. Meteorol.*, 92: 211-225.
- Olivier, M., 1996. Geostatistics, Rare Disease and the Environment. In: *Spatial Analytical Perspectives on GIS*, Fisher, M., H.J. Scholten and D. Unwin (Eds.). Taylor and Francis, London, pp: 87-99.
- Olivier, M.A. and R. Webster, 1990. Kriging: A method of interpolation for GIS. *Int. J. GIS*, 4: 313-332.
- Soares, A., 2000. *Geoestatística Para as Ciências da Terra e do Ambiente*. IST Press, USA., pp: 206.
- Stein, A. and W. Zaidnoordijk, 1999. Improved parameter estimation for hydrological models using weighted object functions. *Hydrological Proc.*, 13: 1315-1328.
- Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. *Economics Geography*, 46: 234-240.
- Vieira, S.R., D.R. Nielsen and J.W. Biggar, 1981. Spatial variability of field measured infiltration rate. *Soil Sci. Soc. Am. J.*, 45: 1040-1048.
- Wang, X. and Z. Zhang, 1999. A comparison of conditional simulation, Kriging and trend surface analysis for soil heavy metal pollution pattern analysis. *J. Environ. Sci. Health*, 34: 73-89.