Concepts and Methods of Geostatistics

Jacques Rivoirard

Centre de Géostatistique, Ecole des Mines de Paris

1 Introduction

A mathematician but also an engineer, Professor Georges Matheron has shown an exceptional ability to formalize practical problems, edict relevant concepts and methods in order to find workable solutions. From 1954 to 1963, while working with the French Geological Survey in Algeria and in France, he discovered the pioneering work of the South African school on the gold deposits of the Witwatersrand, and formalized the major concepts of the theory that he named Geostatistics. The "classical" geostatistics (linear geostatistics based on stationary covariance and variogram) was fully developed in his thesis [11]. The regionalized variable under study, as Matheron called it, is then conveniently modelled as a Random Function. However one should not consider the model as the reality, and the analysis of the role to be assigned to the model led Matheron to write "Estimating and Choosing", an essay on probability in practice ([18] for the French version, [20] in English). In the mean time, the bases of both non-linear geostatistics and non-stationary geostatistics were laid out. The reader will find a detailed description of the different geostatistical methods, as well as a valuable bibliography, in the reference book by Chilès and Delfiner [1]. The present article describes the basic concepts of geostatistics, from today's perspective.

2 Linear geostatistics

2.1 The origin

Matheron discovered the work of the South African school during his first professional appointment with the French Geological Survey in Algeria in 1954. One year later, a paper written by R. Duval [4] in collaboration with R. Lévy and G. Matheron, presented the work by Krige [6] in a very concise way in the French journal "Annales des Mines".

The basic ingredients of geostatistics (currently referred to as the support, the dispersion variances, the conditional bias, see further on) were already present in the South-African work (see the paper by Krige and Kleingeld in this volume, page 5). The variances of gold sample values were observed to be higher when these samples were taken within a larger area, e.g. the variance of samples within a mine being higher than the variance of samples within a mining block or panel (here a "sample" is given its physical meaning or its corresponding ore value, which is different from the statistical meaning, i.e. the collection of such sample values). In addition, those variances (more exactly the logarithmic variances, see further on) are linked by Krige's additivity relationship:

variance of samples within mine
= variance of samples within panels + variance of panels within mine.

The fact that the variance of gold values within a given area depends on the "support" on which the variable is defined (sample, panel..., i.e. the 2D generic area in size and orientation, or the generic volume in 3D), has serious consequences when selecting panels from their sampling. As panels are less variant than samples, it follows that the mining panels, selected as being rich according to their samplings, are naturally less rich than these samplings. Similarly panels considered to be poor are richer than their samplings. No bias on samples is necessary to explain this overestimation of rich panels and underestimation of poor ones. It simply corresponds to a particularly dangerous case of conditional bias: conditionally to its samples, the expectation of the value of a panel is not equal to its sampled value, it is lower when the sampled value is high, higher when this value is low.

To avoid the overestimation of selected reserves caused by such a conditional bias, corrected estimators of panels from samples have been developed by the South-African school, based on the lognormal distribution. The lognormal distribution was observed to provide a good fit to gold content values. Let us recall that it is skewed positively, with a tail of large values, and a shape factor given by the logarithmic variance, i.e. the variance of the logarithm of the variable. The higher the logarithmic variance, the more dispersed the values relatively to their mean. Because of the skewness of the distribution, lognormal estimators of the mean of the distribution (computed from the arithmetic mean of log values, i.e. the geometric mean of values) from a limited number of samples, can be significantly better than the arithmetic mean of values. Krige's methodology assumes the permanence of the lognormality of the distributions, e.g. the distributions of sample values within a panel, or within a mine, are lognormal, as well as the distribution of panel values within the mine. The variations of the values of such distributions are then described by their logarithmic variances, these being linked by the additivity relationship mentioned previously.

In addition to the article by Duval, Lévy and Matheron on Krige's work, the same issue of Annales des Mines contained another remarkable article written by Matheron alone [8]. In particular, he derived the permanence of

lognormality from a principle of self-similarity when splitting blocks iteratively as considered by de Wijs [3]. This synthesis allowed him to distinguish two terms within the dispersion variance of the support v within the domain V: a term depending on the sole geometry of v and V; and a coefficient of "absolute dispersion" (as he called it) which is an intrinsic characteristic of the mineralization (this parameter will be later replaced by the more general structural tool given by the variogram). During these early years, Matheron produced several statistical reports and studies of deposits. It is worth noticing that the term "geostatistics" appeared explicitly in the title of a study of a lead deposit dated 1955 [9].

Next, Matheron stepped to the linear geostatistics, based on the additivity of variables and linear estimators, without any assumption on the type of statistical distribution. As a matter of fact, the gold content variable initially studied by Krige was the metal accumulation taken vertically across a reef extending in 2D. Supposing that a panel is partitioned into samples with same support v, the mean value over the panel is equal to the arithmetic average of sample values, whatever the statistical distribution of the values, lognormal or not. In other words, the accumulation is an additive variable (this is also the case for the thickness of the reef which gives the ore quantity, but not for the grade, equal to the ratio between the metal and the ore quantities - or equivalently between accumulation and thickness: the grade has to be weighted by the thickness when averaged). In the early South-African school, the estimation of a given block value was brought down to the estimation of the mean value of the statistical distribution made of all possible samples constituting the block, from the samples available at the periphery of the block, and the spatial aspect was not modelled. By contrast, Matheron rather addressed the estimation of the mean value over any block or domain, using located sample data and taking explicitly into account the geometrical configuration of all the elements. This enabled him to formalize the concepts at the basis of geostatistics in a very simple manner.

2.2 Additivity, support and dispersion variances

Let $z(x)$ be a "regionalized variable", as Matheron called it, depending on the location x, traditionally a point in 2D or 3D. In practice such a point often corresponds to the sample support. Assuming that the variable is additive, its "regularized" value over v (e.g. a block) is the arithmetic average of its points:

$$z(v) = \frac{1}{|v|} \int_v z(x)\, dx$$

Similarly we have:

$$z(V) = \frac{1}{N} \sum z(v_i)$$

when considering the regularized value over V, this being partitioned into N blocks v_i with same support v.

The variation of the v_i values within V can be measured by the "dispersion variance" of support v within domain V:

$$s^2(v|V) = \frac{1}{N} \sum_i [z(v_i) - z(V)]^2$$

Similarly the dispersion variance of a point within V is:

$$s^2(O|V) = \frac{1}{|V|} \int_V [z(x) - z(V)]^2 \, dx$$

We also have:

$$s^2(O|v_i) = \frac{1}{|v|} \int_{v_i} [z(x) - z(v_i)]^2 \, dx$$

and the mean of these quantities over the different v_i's within V gives the dispersion variance $s^2(O|v)$ of a point within v. It is easy to demonstrate that such dispersion variances are linked through the additivity relationship:

$$s^2(O|V) = s^2(O|v) + s^2(v|V)$$

This formula is similar to the relationship obtained by Krige on logarithmic variances, but is satisfied for any statistical distribution of the variable, provided that this variable is additive.

In addition, such dispersion variances, which are of crucial importance when the sample support does not coincide with the support of interest (mining block, area to be depolluted...), are related to the "spatial structure" represented by the variogram. But this will be more conveniently developed later within the framework of Random Function.

2.3 Intrinsic approach versus transitive approach

When developing linear geostatistics, Matheron distinguished two approaches, the transitive approach (seldom used), and the intrinsic approach. In the transitive approach, the phenomenon to be studied (orebody, fish stock...) is supposed to be known, for instance, on a regular rectangular grid defined by its orientation, its mesh size, and its origin. The exact boundaries of the phenomenon are considered to be unknown, and the grid is supposed to extend beyond them. The transitive approach allows for instance to estimate the global abundance from such a grid, with an error variance derived from a structural tool called the transitive covariogram (similar to a noncentered covariance, where terms are summed instead of being averaged). The approach can be particularly useful when studying resources with diffuse limits such as some fish or larvae densities in the sea, or when estimating the extension area or the volume of a phenomenon only known at the nodes of a grid. In the case of a regular grid with a given orientation, the only element of randomness is

the origin of the grid, which is supposed uniformly distributed over the mesh size.

By contrast, in the traditional intrinsic approach, the domain is considered to be known, and the variable is supposed to have an intrinsic behaviour, independent of the geometry and of the boundaries of the domain. The domain can be considered as a window, allowing to see the variable, which is assumed to extend with the same behaviour beyond the window boundaries. To describe such an intrinsic behaviour, the methodology relies on a form of stationarity, i.e. invariance under translations. Although classical linear geostatistics was essentially completed by the early 60's [10], Matheron did not introduce the convenient framework of Random Function until he completed his thesis [11].

2.4 The intrinsic model and the variogram

The regionalized variable $z(x)$ is then considered as a realization, a possible outcome, of a Random Function (RF in short) $Z(x)$, whose specifications constitute the RF model. This framework allows a simple formulation of stationarity. Thanks to de Wijs's work [3], and to Krige's observation of variance increasing with the area, Matheron privileged the stationarity of the increments over the stationarity of the variable itself. In linear geostatistics the most common model is the intrinsic RF model, defined by the increments $Z(x+h) - Z(x)$ having a zero expectation and a stationary variance (depending on vector h, not on location x):

$$E[Z(x+h) - Z(x)] = 0$$

$$\mathrm{Var}[Z(x+h) - Z(x)] = E\{[Z(x+h) - Z(x)]^2\} = 2\gamma(h)$$

The intrinsic RF model is then completely specified by its structural tool $\gamma(h)$, the variogram, which expresses the mean variability between two points as a function of the vector between them, and which depicts the more or less regular behaviour of $Z(x)$. In practice, a set of distances is chosen, in accordance with the sampling grid. For each distance an experimental variogram value is computed from pairs of data separated by this distance, using implicitly the invariance of increments under translation. This experimental variogram is then interpreted (e.g. some variations being considered as insignificant fluctuations) and fitted by a mathematical function, the variogram model, assuming additional hypotheses on its regularity for instance. As the variogram directly measures the variability between pairs of points, Matheron called it the *intrinsic* dispersion function in early times. The intrinsic RF model, based on increments, allows to express the expectation and variance of linear combinations $\sum_\alpha \lambda_\alpha Z(x_\alpha)$, but only when $\sum_\alpha \lambda_\alpha = 0$ (as a matter of fact, such linear combinations can be seen as linear combinations of increments; the expectation and the variance of other linear combinations, in particular $Z(x)$, are

simply not defined in the model). These admissible linear combinations have a zero expectation, and a variance which depends on the variogram:

$$\text{Var} \sum_{\alpha} \lambda_\alpha Z(x_\alpha) = -\sum_{\alpha} \sum_{\beta} \lambda_\alpha \lambda_\beta \gamma(x_\beta - x_\alpha)$$

To ensure the positivity of such variances, the function $-\gamma(h)$ must be conditionally positive definite. The fitting task is usually simplified by considering a sum of known authorized model components (one of them being the white noise or "nugget effect", a simple discontinuity at the origin).

Such a measurement of variability is not specific to geostatistics. The novelty introduced by Matheron consisted in establishing it within a simple model while accounting for the important concepts of support, regularization, and not the least, the two different following types of variances, and to make all this available to the practitioner: the engineer and the mathematician are in perfect agreement.

If $Z(x)$ is an intrinsic RF, so is the regularized $Z(v)$ over support v, with a "regularized" variogram deduced from the variogram of $Z(x)$. The dispersion variance of the support v within the domain V also depends on the variogram of $Z(x)$:

$$D^2(v|V) = \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

and in particular for a point support:

$$D^2(O|V) = \bar{\gamma}(V, V)$$

where $\bar{\gamma}(V_1, V_2)$ stands for $\frac{1}{|V_1||V_2|} \int_{V_1} \int_{V_2} \gamma(x - y)\, dx\, dy$

Such dispersion variances are linked by the additivity relationship, e.g.:

$$D^2(O|V) = D^2(O|v) + D^2(v|V)$$

In practice this explains, or makes it possible to predict the drop of variance when changing the support, from a quasi point sample support to a larger interest support, for instance:

$$D^2(O|V) - D^2(v|V) = D^2(O|v) = \bar{\gamma}(v, v)$$

The other type of variance is the estimation variance, or variance of the error when estimating for instance the value $Z(V)$ over V by the value $Z(v)$ over v (or the contrary). The estimation error $Z(v) - Z(V)$ is an admissible linear combination. Therefore its mean is 0 (estimation is unbiased), and its variance is:

$$\text{Var}[Z(v) - Z(V)] = 2\,\bar{\gamma}(v, V) - \bar{\gamma}(v, v) - \bar{\gamma}(V, V)$$

Using a variation of this formula, the "global estimation variance" can be computed when estimating a whole domain with a regular sampling grid for

instance. Conversely, assuming a given variogram, this can help define a sampling grid in order to obtain a given estimation variance.

If we now consider the estimation of $Z(V)$ by a linear combination of samples $\sum_\alpha \lambda_\alpha Z(x_\alpha)$, the estimation error is $Z(V) - \sum_\alpha \lambda_\alpha Z(x_\alpha)$. It is an admissible linear combination iff $\sum_\alpha \lambda_\alpha = 1$ (if not, the error is simply not defined in the intrinsic RF model). Then its expectation is 0 and its variance can be developed:

$$\begin{aligned}\mathrm{Var}\left[Z(V) - \sum_\alpha \lambda_\alpha Z(x_\alpha)\right] \\ = 2\sum_\alpha \lambda_\alpha \bar{\gamma}(x_\alpha, V) - \sum_\alpha \sum_\beta \lambda_\alpha \lambda_\beta \bar{\gamma}(x_\alpha, x_\beta) - \bar{\gamma}(V, V)\end{aligned}$$

By minimizing this estimation variance, one gets an estimator, usually called Ordinary Kriging (OK). Kriging is a generic term for optimal – and generally linear – estimators in geostatistics, and the minimized variance is referred to as the kriging variance (not to be confused with the variance of the kriging estimator, which anyway is not defined in the intrinsic model). One will find an interesting historical study on the origins of kriging compared to other similar methods in a paper by Cressie [2]. The ordinary kriging weights, constrained by the above condition, are solution of a linear system. Such an Ordinary Kriging can be used to estimate the value at a point, the mean value over a block or a domain, or, in the case of mapping, a grid of points or a set of blocks.

2.5 The stationary case

The variogram $\gamma(h)$ may be unbounded and can increase infinitely. When it is bounded (for instance stabilizing on a "sill" level beyond a "range" distance), the IRF can be reduced to a 2^{nd} order stationary RF, characterized by its first two moments being stationary:

$$E[Z(x)] = m$$

$$Cov[Z(x), Z(x+h)] = C(h)$$

In particular the *a priori* variance $\mathrm{Var}\, Z(x)$ is equal to $C(0)$ and we have $\gamma(h) = C(0) - C(h)$. The dispersion variance $D^2(O|V) = \bar{\gamma}(V, V)$ tends towards the *a priori* variance $C(0)$ (absent from a purely intrinsic model) when V increases infinitely in an appropriate way. Then:

$$\mathrm{Var}\, Z(V) = E[(Z(V) - m)^2] = \bar{C}(V, V) = \frac{1}{|v|^2} \int_v \int_v C(x-y)\, dx\, dy$$

tends to zero and $Z(V)$ tends to the mean m (supposed to be ergodic).

The stationary model allows to develop the variance of any linear combination:

$$\text{Var} \sum_{\alpha} \lambda_\alpha Z(x_\alpha) = \sum_{\alpha} \sum_{\beta} \lambda_\alpha \lambda_\beta C(x_\beta - x_\alpha)$$

(the covariance being a positive definite function to ensure the positivity of such variances). The best linear estimator in this model is usually called Simple Kriging (SK). Being unbiased it can be written:

$$\sum_{\alpha} \lambda_\alpha Z(x_\alpha) + \left[1 - \sum_{\alpha} \lambda_\alpha\right] m$$

and the kriging weights minimizing the estimation variance are solution of a linear system.

2.6 Questioning the model

For practical applications, an interesting question is whether the mean parameter m is known, and even what meaning it has when we only have a single realization of the RF model, defined over a finite domain (think of the metal grade of an orebody for instance). If the mean is considered to be unknown, it can be removed from the above estimation by setting the condition $\sum \lambda_\alpha = 1$: we are then back to Ordinary Kriging. In practice, stationarity, as usually detected by the variogram stabilizing on a sill, is often not guaranteed for large distances, e.g. throughout the whole domain: stationarity is then only local. In that case Ordinary Kriging is preferred, using datapoints within a moving neighbourhood around the target (point or block of a grid), i.e. a neighbourhood that moves with the target (as opposed to a unique neighbourhood including all datapoints). In practice the choice of the neighbourhood is still a difficult and much debated question on which we will come back later: in theory the larger the neighbourhood, the better the estimation, but also the stronger the stationarity hypothesis. In particular, while Ordinary Kriging is a weighted average of data within the neighbourhood, Simple Kriging also makes use of the mean m, with a complementary weight. This compensates for a possible sparseness of data with respect to the spatial structure, by attracting the local estimate towards the mean in accordance to the stationarity hypothesis.

Now, assuming stationarity over the whole domain does not guarantee that m can be known, nor that it even exists outside the model. It is interesting to follow the evolution of Matheron's thoughts on models through time. Matheron introduced the RF framework for the intrinsic approach in 1965. Although the transitive approach does not use the randomization of the regionalized variable, Matheron ([13], p. 39) described the transitive methods as "being rich in implicit probabilistic contents", as he considered that the

operations used for fitting a transitive covariogram "constitute a disguised passage to expectations". Finally Matheron ([18, 20]) went back, stating that, if a methodology is performant, it should mean something on the regionalized variable (the reality !): this questions the objective contents of the RF model, by subordinating it to the regionalized variable. He proposed the operational indexrandom set of methods. The operations of fitting a variogram or a covariogram are then considered as resulting from the choice of anticipating hypotheses (which may or may not be correct, as could possibly be checked later), that allow to tell more than the data alone. Parameters of the models, that cannot have their counterpart in terms of the regionalized variable, are considered to be purely conventional and must disappear from the final estimation results. This is of course the case of the mean parameter m of the stationary model if the domain V is not large enough to identify the average Z(V), representing $z(V)$, with this mean. Such epistemological developments are worth being considered, if one aims at applying mathematical models.

2.7 Additional remarks

Before going on the other parts of geostatistics, let us go back in more details to some difficult points that concern the choice or the use of models.

Back to conditional unbiasness

Let $Z(V)^\star$ be an estimator of the panel $Z(V)$. Moreover suppose that it is unbiased:

$$E[Z(V) - Z(V)^\star] = 0$$

Conditional unbiasness can be written:

$$E[Z(V) - Z(V)^\star | Z(V)^\star] = 0$$

or equivalently:

$$E[Z(V) | Z(V)^\star] = Z(V)^\star$$

i.e. the regression of $Z(V)$ on $Z(V)^\star$ is linear and coincides with the first bisector. This desirable property ensures that the results obtained when taking a decision on the basis of $Z(V)^\star$ are, on average, as expected (e.g. values selected for being rich, being as rich as expected). This is not the case of the estimator $Z(V)^\star$ taken as the sampled value of $Z(V)$. Having a variance higher than that of $Z(V)$ due to the support effect, it is conditionally biased, with $E[Z(V)|Z(V)^\star] < Z(V)^\star$ when $Z(V)^\star$ is high. This causes overestimation of panels considered as rich, e.g. above a cut-off z:

$$E[Z(V) | Z(V)^\star > z] < E[Z(V)^\star | Z(V)^\star > z]$$

Krige's lognormal estimator aims at correcting this conditional bias. Ideally, if we denote $data = \{Z(x_1), Z(x_2), ...\}$ the values of samples (possibly located outside the panel), the best estimator (unbiased with minimum error variance) is the Conditional Expectation (CE): $E[Z(V)|data]$. If the multivariate distribution $(Z(V), data)$ were normal (any linear combination of these being normally distributed), this regression would be linear, and would coincide with Simple Kriging in a stationary model. However, and particularly with skewed distributions, there is no reason for this regression to be linear. Assuming for instance multilognormality (multivariate distribution of logarithms being normal), the best regression is lognormal, hence based on the geometric mean of the sample values. Although linear kriging was proposed by Matheron as a means to correct in practice the above conditional bias [13], there is no reason why it should do it fully. A practical advantage of linear kriging comes from the reduced hypotheses: only the spatial structure, and possibly the value of the mean parameter are required. As a matter of fact, an expression such as $E[Z(V)|data]$ is simply not defined within the 2^{nd} order stationary RF model. On the other hand, Matheron has shown that the very shape of CE within a model may be used as a heuristic candidate for the estimation, but that it rapidly gets beyond a realistic meaning with a few conditioning datapoints ([18, 20]): CE is guaranteed to be the best estimator only in theory. In addition (Z being additive), if V is partitioned into congruent samples, the best estimator of $Z(V)$ when increasing the number of samples, converges towards the true value, which is the (linear) arithmetic mean. While strict conditional unbiasness seems an unreachable ideal, one can try to approach it. One way is to look for a *linear* regression of $Z(V)$ on $Z(V)^\star$ close to the first bisector (slope close to 1). In Simple Kriging, the estimation error is uncorrelated with data and therefore:

$$\operatorname{Var} Z(V) = \operatorname{Var} Z(V)^\star + \operatorname{Var}[Z(V) - Z(V)^\star]$$

So Simple Kriging reduces the variability, and the theoretical slope of the linear regression $Cov(Z(V), Z(V)^\star)/\operatorname{Var} Z(V)^\star$ equals 1. The mean compensates for the lack of information in the neighbouring data (with a pure nugget effect, Simple Kriging reduces to the mean m: no local estimation is really possible). In Ordinary Kriging, assuming 2^{nd} order stationarity with unknown mean, things are different. The theoretical slope can be notably less than 1, in particular when the spatial structure is poor and the data sparse. In many cases however, this slope takes low values when the neighbourhood is too small (Ordinary Kriging then has too high a variability, with an important conditional bias), but may have a value close to 1 if the neighbourhood is large enough. Hence this slope can be used as a criterion for choosing the neighbourhood.

Back to variances

A question considered by Krige was whether the log variance of samples within a panel could be considered a constant for panels in identical configurations

within a given mine. The same question formulated on the dispersion variance of samples within a given panel, in linear geostatistics, would generally receive a negative answer for skewed distributions. Rich panels would generally contain richer and more variant samples: this corresponds to a proportional effect (between the variance and the mean, or a function of the mean), or in statistical language, a heteroscedasticity. This does not imply non-stationarity as is often thought: exponentiating a 1D stationary autoregressive Gaussian process yields to a stationary lognormal process presenting such a proportional effect. When writing $D^2(O|V) = \bar{\gamma}(V, V)$ for instance, linear geostatistics considers an expectation, or average over all possible panels with same support V, then taking the average of different values.

Similarly the kriging variance (the variance of the error which is minimized by kriging) has the meaning of an average of the estimation variance when translating the geometrical configuration (respective locations of data and target to be estimated). It is not conditional on values of the variable observed locally and thus can appear as locally unrealistic: for instance, it is unrealistically small when the estimate is high, and conversely. To have a variance more realistic locally, a correction can be applied, based on the modelling of the proportional effect. On the other hand, non-linear geostatistics can give theoretical access to conditional variances.

Back to lognormal

Lognormality was central in the early development of geostatistics in the 50's. Permanence of lognormality was in particular advocated when working on different supports or within nested areas. In theory however, the average of independent variables with the same lognormal distribution, for instance, is not lognormal. Then, assuming that the variable is lognormal on a given support, it would not remain true for multiple or dividing supports. In other words, assuming that an additive variable is lognormal on some support, there is no chance that this very support coincides with any support available in practice. Since lognormality is not stable, Matheron [12] explored the family of stable distributions that would remain stable with a change of support. However this alternative has not led to practicable results. As one says, simple models are false, but complex models are useless... Lognormality, and its permanence, can be expected to hold, although deviations from supposed lognormality can have serious consequences (a third translation parameter in addition to mean and variance, is often considered to enlarge the model). The question of lognormality arises again from time to time. In a note with an eloquent title, Matheron [16] has proposed a variety of enlarged lognormal estimators, obtained notably by exponentiating a linear combination of logarithms with sum of weights equal to 1 such as OK, and assuming or not the permanence of lognormality when changing support. The sensitive estimator referred to as "lognormal kriging" can take many different forms.

3 Multivariate geostatistics

Linear geostatistics can naturally be extended to several variables (such as different concentrations, or a variable and its gradient, etc.) Cokriging allows estimating linearly one variable from a set of variables, known at the same datapoints (isotopic case) or at different datapoints (heterotopic case). Cross structural tools (covariance, variogram) will measure the correlated parts between two variables, or between increments, as a function of the vector separating pairs of points. In the frequently used linear models of coregionalisation, the set of simple and cross covariances (or variograms) is modelled from a linear combination of structural basic components which usually represent different scales. Efficient filtering techniques (kriging analysis) make it possible to extract such a scale component by cokriging from the different variables. One advantage of these filtering techniques, and more generally of kriging and cokriging, is their capability to handle points in irregular geometrical configurations, and not only on regular grids.

Very often, the statistical features of the regionalized variables have a limited scope, since they change with the support on which the variables are considered (this is the case of lognormality, as we have seen before). In particular, the correlation coefficient between two given variables may be small on a small support, and much larger on a larger support – or the contrary. On the opposite, the correlation may be an intrinsic measure of the linear dependency between two variables. This is the case in the model of intrinsic correlation, where simple and cross variograms are proportional to each other, and where the correlation coefficient within a domain does not depend on the support, nor on the domain [11]. Then, in the isotopic case, cokriging is equivalent to kriging.

Similarly, the linear regression $aY(x) + b$ of one variable $Z(x)$ on another variable $Y(x)$ at the same location generally presents little interest, since it changes with the support. However, in the model with residual $Z(x) = aY(x) + b + R(x)$ where $R(x)$ is spatially not correlated with $Y(x)$ – a hierarchical model where $Z(x)$ is subordinated to the master variable $Y(x)$, the linear regression is independent from the support and then has an intrinsic signification. This is the case when the cross-structure between $Y(x)$ and $Z(x)$ is proportional to the structure of $Y(x)$. The decomposition of $Z(x)$ into $Y(x)$ and $R(x)$ has interesting properties in cokriging. When $Y(x)$ is available at every datapoint of $Z(x)$, and in particular when $Y(x)$ and $Z(x)$are available at the same datapoints, the cokrigings of Y and Z can be obtained by kriging Y and R separately.

An instructive example of coregionalisation (that makes the link with non-linear geostatistics) is the case of a concentration $Z(x)$ where the positive values correspond to a geometrical set A with unknown delimitation within the domain under study (ex: metal grade in a vein-type deposit). This set A can be represented by its indicator $1_{x \in A}$, equal to 1 if the point x belongs to A, and to 0 otherwise. If we consider this variable separately, its ideal

estimator at point x from datapoints consists in its Conditional Expectation, that is, its probability to belong to A conditionally to the datavalues. The mathematical shape of such a conditional probability is generally unknown. In the context of simulations, it is suggested to introduce such conditional probabilities under given configurations from training images, using multi-point statistics [22]. But most often a pseudo-probability is obtained by kriging the indicator, with a post-processing to eliminate values outside the interval [0, 1]. In addition a relationship may exist between variables, say grade and geometry: for instance the low grades may be preferentially located near the frontiers of the veins. Then the estimation of the indicator can be improved by considering values of Z in addition to values of the indicator. A straightforward exception corresponds to the positive values of $Z(x)$ being distributed within A, independently from its geometry. Then the estimation of the indicator, which has the meaning of a probability, can be performed separately, and complemented by an estimation of $Z(x)$ assuming it is positive (and using only the positive data values). Another exception is cokriging the indicator and Z within a model with residual $Z(x) = m1_{x\in A} + R(x)$ where we assume the absence of spatial correlation between residual and indicator. This corresponds to a reduced hypothesis of independence: internal independence between the variable $Z(x)$ and the set A [11] or absence of border effect, in the sense that the expected value of $Z(x)$ at a point x of A does not depend on whether a neighbouring point $x+h$ belongs to A or not. When data consist in Z values, the cokriging of Z can be obtained by kriging separately the indicator and the residual, and the ratio between the estimate of $Z(x)$ and the estimate of the indicator has the meaning of an estimation of $Z(x)$ in case it would be positive. We will come back to such considerations when dealing with non-linear geostatistics.

4 Non stationary geostatistics

4.1 Kriging with external drift

Consider the model with residual $Z(x) = aY(x) + b + R(x)$. In some circumstances the master variable Y is known everywhere and is used as an auxiliary variable for the estimation of a sparsely sampled target variable Z. Assuming that the coefficients a and b of the linear regression of $Z(x)$ on $Y(x)$ at the same point x are known, Simple Kriging of the residual gives the estimation of Z. If b is unknown, then an Ordinary Kriging of the residual is what is needed. When working with a single realization, the auxiliary variable can be considered deterministic, say $f(x)$ to be consistent with literature, and the model can be written $Z(x) = af(x) + b + R(x)$. This is helpful when a and b are not known (either globally when using a unique neighbourhood, or locally within a moving neighbourhood). A trick when estimating $Z(x)$ as a linear combination $\sum_\alpha \lambda_\alpha Z(x_\alpha)$ of data (borrowed from Universal Kriging described

further on), consists in imposing the following conditions:

$$\sum_{\alpha} \lambda_\alpha f(x_\alpha) = f(x)$$

$$\sum_{\alpha} \lambda_\alpha = 1$$

This ensures that the mean error is 0 whatever the coefficients a and b. In the model, the drift $E(Z(x)) = af(x) + b$ of $Z(x)$ depends upon $f(x)$, hence the name of kriging with external drift (KED). This can be naturally extended to drifts $\sum_l a_l f_l(x)$ where $f_l(x)$ are known shape functions, for instance (1, f(x), f^2(x)): this was actually the origin of KED when mapping a geological depth with f(x) representing the seismic travel time [1]. This KED method is very popular, probably because of its flexibility: kriging implicitly estimates the coefficients a_l that best fit the drift onto the shape functions. The difficult point may be the estimation of the residual structure to be used in kriging. Indeed, if the coefficients a and b are unknown, the value of the residual is unknown even at points where Y and Z are known. Coefficients and residuals can be estimated, but the structure of the estimated residual is not the true one. The residual structure is often determined indirectly e.g. through cross-validation.

4.2 Universal kriging

By considering increments, the model of intrinsic RF seen previously is far more general than the stationary one, which is constrained to vary around its mean. This freedom is visible when looking at 1D random walks such as a Brownian motion (which is intrinsic but not stationary). However the intrinsic model cannot force the function to show systematic variations (e.g. the sea floor depth increasing from coast line, or the dome-shape top of an oil reservoir). Then it can be helpful to consider a drift of the form $\sum_l a_l f_l(x)$, where the f_l's are known. For instance using the monomials 1, x^2, y^2, xy (in 2D space) corresponds to a quadratic drift centred on the origin. Most often, there is no point to be distinguished as the origin, and the family of functions f_l is taken invariant by translation: the drift can be written $\sum_l a_l f_l(x - x_0)$, whatever the choice of the origin x_0, e.g. 1, x, y, x^2, y^2, xy for a quadratic drift. More generally, monomials of coordinates with degree $\leq k$ allow representing a polynomial drift of degree k (sets of sine and cosine functions with a fixed period, and more generally sets of exponential polynomials, also have this property of invariance under translation). When making the assumption of a polynomial drift of order k, the coefficients a_l's of the drift should be considered unknown *a priori*. Imposing universality conditions $\sum_\alpha \lambda_\alpha f_l(x_\alpha) = f_l(x)$

to the linear estimation $\sum_{\alpha} \lambda_\alpha Z(x_\alpha)$ of $Z(x)$ will ensure the absence of bias, whatever the values of the coefficients: this gives the Universal Kriging.

Yet the remaining problem lies in the estimation of the residual structure when it is not small compared to the variations caused by the drift, since the drift coefficients, and therefore the drift, are unknown. Fitting the drift on the data tends to include some part of the true residual variability, and therefore leads to an underestimation of the residual structure. The variogram of the estimated residuals may in particular display a finite or a small range, whereas this is not the case for the true residual.

4.3 Intrinsic Random Functions of order k

A solution to this problem was offered by the theory of Intrinsic Random Functions of order k ("IRF-k ", [14]). By considering increments, the usual intrinsic RF model studied initially was more general than the stationary one: the constant mean, if any, disappeared, and the family of admissible structures was enlarged from stationary covariances to variograms. Similarly the IRF-k model is defined by considering increments of increments, ..., or generalized increments, that filter out any polynomial drift of order k, and the tools are enlarged to generalized covariances ($-\gamma(h)$ being a generalized covariance of order 0). An IRF-k constitutes an equivalence class of RF that have a given structure and differ by a polynomial of order k. Only those linear combinations that are generalized increments are defined in the model. Their variances can be developed using the generalized covariance and allow linear estimation by kriging, unbiased with minimum variance. In practice the difficult point is determining the structure (the generalized covariance), which is not direct as for the usual variogram, except with regularly spaced data. Indirect fitting of generalized covariances that are admissible linear combinations of odd monomials of $|h|$ is often used. Other generalized covariances are also admissible, in particular some for which kriging coincides with splines.

The interest of the model of Intrinsic Random Functions of order k is not limited to its theoretical clarity, in breaking the vicious circle between drift and residual. As integrating a stationary RF yields an IRF-0, similarly integrating an IRF-k yields an IRF-k+1, and IRF-k's appear as the perfect tool in some Partial Derivative Equations problems where the set of solutions exactly corresponds to the set of RF defining an IRF-k (for instance the set of Random Functions whose laplacian is a given stationary RF exactly constitutes a unique IRF-1).

5 Non-linear geostatistics

5.1 Estimation on point support

While requiring few hypotheses, linear geostatistics has a limited range correlatively. The need for non-linear geostatistics can be felt in different cir-

cumstances. Firstly the best estimator of a regionalized variable may be non linear (although any estimation of the regularized variable should be asymptotically linear, as seen above). Secondly, even when using a linear estimation such as kriging, the kriging variance is not conditional on the neighbouring data: with skewly distributed variables, local measures of uncertainties such as estimation variances should depend on the neighbouring data. Finally, and this was the very origin of the development of the non-linear geostatistics in the 70's (after the early lognormal framework of geostatistics), in some cases (selective mining, pollution...), the practical problem is not to estimate the variable itself, but a non-linear function of it: for example its exceeding a given threshold z, which is represented by the indicator $1_{Z(x)\geq z}$. Note that thresholding an estimated value of $Z(x)$ would be an incorrect answer to this problem, as this may very well exceed the threshold while the true unknown value does not.

When its assumptions are acceptable, a very convenient model is the (multi-)gaussian Random Functions one (denoted Y for reasons explained further on), where any linear combination on any finite set of points is normally distributed (which is far more demanding than to require the marginal or even the bivariate distributions to be normal). Firstly, the best estimator of $Y(x)$ from data $Y(x_\alpha)$ coincides with the linear regression, i.e. Simple Kriging (assuming stationarity). Secondly the conditional variance coincides with the unconditional one, i.e. the kriging variance. And finally, the conditional distribution is normal, which makes the best estimation of any function of Y possible through its Conditional Expectation. In particular the CE of the indicator $1_{Y(x)\geq y}$ is equal to the conditional probability of exceeding y, given the data.

Assuming stationarity and a large domain, the marginal distribution corresponds to the histogram of the datavalues. In many cases it is quite different from the famous bell shape curve, and therefore the multi-gaussian model is most frequently used together with a Gaussian anamorphosis: the target variable $Z(x)$ is written as $Z(x) = \phi(Y(x))$, where ϕ is a non decreasing function and $Y(x)$ a stationary standard Gaussian RF, hence the notation. For instance, a lognormal distribution corresponds to an exponential anamorphosis. When the inversion from Z to Y is straightforward (ϕ strictly increasing or nearly so), data and estimation on Z can be expressed in terms of Y. This gives access to the conditional distribution of $Z(x)$ and therefore to the CE of $Z(x)$ or any function, e.g. $1_{Z(x)\geq z}$. The requirements for this theoretically best estimator are severe, as not only the marginal distribution but also the bivariate $(Z(x), Z(x+h))$ or the n-variate ones must be normal (after transformation). Another method for estimating an indicator above a given threshold consists in kriging directly this indicator. The method has been made popular by Journel [5] for its simplicity and the few hypotheses apparently required. Extension to several thresholds allows estimating variables obtained from linear combinations of the different indicators, such as the discretized variable under study or any function of it. On the other hand, kriging

an indicator ignores the extra information contained in the Z data or in the other indicators. As a matter of fact, indicators at different thresholds are not independent. The Disjunctive Kriging (DK), proposed early by Matheron [15], is the estimator based on the cokriging of all indicators. This requires the simple and cross structures of indicators, or equivalently the bivariate distributions $[Z(x), Z(x+h)]$. DK is simplified in isofactorial models, where it is obtained by Simple Kriging of factors common to all the bivariate distributions of the model. Through a number of reports, Matheron has undertaken a considerable work developing such models, including the change of support seen in next section. An extensive bibliography can be found in Chilès and Delfiner [1]. Most models correspond to diffusion processes, and refer to a specified statistical distribution (e.g. Gaussian, gamma, binomial negative, with Hermite, Laguerre, and Meixner polynomials as factors). Discrete diffusion models have been developed specifically for diamond mining (see Krige and Kleingeld in this volume). Another model is a hierarchical mosaic model, where the indicators correspond to sets that are nested without border effects and are factorized by indicator residuals [21].

One very particular model consists in a mosaic with independent valuation: each tile of a stationary partition of the space is given an i.i.d. value. The variographic structure of indicators or other functions is entirely based on the probability for two points to belong to the same tile or not. In particular there is no destructuring phenomenon when the cut-off increases or decreases, contrary to what is currently observed on real data. In this particular model, cokriging of indicators equals to their kriging, and DK of any function equals its kriging, with the same kriging weights.

5.2 The change of support

Very often, the question of whether a threshold value is exceeded or not, concerns areas much larger than the quasi point sample support: this may be the whole domain itself, subdomains with specific geometries, or blocks with same support v partitioning the domain. A general solution can be obtained through conditional simulations mentioned further on. However simulations are very demanding in terms of hypotheses.

The case of blocks with same support v partitioning the domain has been studied in details by Matheron. Due to the support effect, the global distribution of values on such blocks is different from the distribution with a point support. They both have the same mean, and their variances are related through the additivity relationship seen above. In addition other relations link the two distributions, which can be derived from the key Cartier's relationship [19, 7]:

$$E\left[Z(x) \mid Z(v)\right] = Z(v)$$

This stipulates that, conditionally to a block value, the expected value of a random point uniformly distributed within this block, is equal to the block

value. This relation, or equivalent ones, makes it possible to build *change of support models* under various hypotheses (Gaussian, mosaic, etc.) which aim at predicting the distribution of values on blocks when the point support distribution is known. They can be used in particular to predict global quantities like the proportion of blocks above a cut-off.

The local estimation, i.e. predicting whether a given block $Z(v)$ of the domain, or the N blocks $Z(v_i)$ within a panel, exceeds a cut-off value or not, is a difficult problem. In addition to the simulation approach, different methods have been proposed. A direct approach consists in assuming a given type of distribution (e.g. normal or lognormal) for the block value, only depending upon an estimated value of the block with an appropriate estimation variance. This is in particular justified within a Gaussian model, as the distribution of a block, conditionally to the data, only depends on its simple kriging estimate and variance, and provides the conditional probability of exceeding a cut-off, or more generally the CE estimator of any function of the block. In the usual case where an anamorphosis must be used, things are more complex, but CE can be developed when coupled with a Gaussian change of support model. Similarly, disjunctive kriging can be used under different models equipped with a change of support. EC and DK provide a local estimation consistent with their corresponding global change of support.

The previous non-linear methods are conveniently developed in a stationary framework. In an area with sparse data, the estimation is then strongly attracted by the global characteristics of the domain, just like Simple Kriging is attracted by the mean m. This is in accordance with the stationarity hypothesis, but may be considered too strong. For estimations on point support, the stationary hypothesis made in kriging or cokriging indicators can be reduced if an Ordinary Kriging is used instead of a Simple Kriging for indicators or factors. However, things are more complex when a change of support is used. Alternative methods, more flexible in this respect, have been proposed. One of them consists in firstly estimating the value of the indicator of each block that contains a sample data, and secondly performing the Ordinary Kriging or Cokriging of such a *service* variable. Another method considers the distribution of blocks within a given panel as being uniformly conditioned by the sole estimate of this panel, then using a change of support model up to the panel support. Still another approach is to derive the block distribution within a panel from the more easily estimated distribution of points.

5.3 Information effect and geometrical constraints

Geostatistics is more than the development of mathematical - generally probabilistic - models and methods, and their application. It also includes analyzing the practical problems to be solved, and formalizing them in terms of concepts, which Matheron considered very important. In the mining domain that he studied in details, Matheron proposed an advanced formalization of reserves, in particular for selective mining. As an illustration, consider the case

of an orebody divided into equal blocks, each one being possibly selected as ore independently (free selection). It is necessary to account for the support effect in order to predict the number of blocks above a cutoff, from the distribution of point sample values: selectivity on block support is less than it would be on points. Moreover, when a block is to be mined, the decision of selecting it as ore or waste is not made on its true value which is unknown. It is made on an estimate based on the data that are available at this stage (estimation by kriging for instance). The result of such a selection is necessarily worse than it would have been in the ideal case of a selection on the true values: this is the information effect [19]. In practice, additional data will be available at the pre-exploitation stage. When predicting the reserves months ahead, one has to estimate the chances for the future estimated block values to exceed the cutoff, and the correlative true quantity of metal that they contain.

In free selection, blocks are assumed to be possibly selected independently from one another and from their location. In the opposite case, another effect – geometrical constraints - must be considered. An example is given by vertical constraints in a stratiform exploitation, when at a given 2D location, all blocks between the lower and the upper selected blocks must be mined. Another common example is that of the open pit, where mining a given block implies that of all the blocks ahead in the cone defined by the stability slope have been previously mined. Matheron [17] has developed methods for the parameterization of reserves, which allow optimizing such contours of exploitation. They provide a family of technically optimal projects, among which it is possible to choose the best ones under present or hypothetical economical conditions. No probabilistic framework is used here.

6 Simulations

Non-linear methods are more powerful than linear ones. They are however more demanding in terms of hypotheses, reducing the robustness of models with respect to reality. Hence the economy principle of geostatistics, i.e. not using more than is necessary to solve a practical problem [20]. But in many cases, the problem is too complex to be solved directly, even by non-linear methods. Then stronger hypotheses are required, and a solution is to call for simulations, that is, realizations of a complete RF model, considered as plausible versions of the reality - in particular conditional simulations, which honour the values at data points (see the article by Chilès and Lantuéjoul in this volume). The increase in computational power and the possibilities of visualisation have boosted the use of simulations. The proper use of simulations in terms of the practical problem to be addressed, and the choice of the model, i.e. the hypotheses made on reality, are important points to be considered.

7 Perspectives

The need to model complex environments realistically calls for the development of new models, possibly process-based (e.g. diamonds deposited within traps of the sea floor), or defined in a non Euclidean system (stream lines) or in a higher dimensional space. In particular, spatio-temporal models must be considered for variables such as fish or contaminants which vary in space and time. Enlarging to new types of non-stationarity or to new multivariate models is also a promising challenge. A Bayesian approach for geostatistics is convenient to account for uncertainties in the model. It also constitutes a powerful framework for inverse problems in hydrogeology for instance. In this case and in many others, geostatistics offers a consistent approach to analyse several variables jointly and to deal with measurements on different supports. It also provides a flexible approach to combine, merge or filter different variables, measured at the same locations or not, whether on a regular sampling design (e.g. grids, images) or not. Finally geostatistics can also be used for the coupling between observation data and physical models.

References

1. Chilès, J.-P., and Delfiner, P. 1999. Geostatistics: Modeling spatial uncertainty, Wiley, New York, 695 p.
2. Cressie, N., 1990, The origins of kriging, *Math. Geol.,* Vol. 22, 3, p. 239-252.
3. De Wijs, H. J. 1951-53. Statistical ore distribution, Geologie en Mijnbouw, Part I nov. 1951, Part II Jan. 1953.
4. Duval, R., Lévy, R., Matheron, G., 1955. Travaux de D. G. Krige sur l'évaluation des gisements dans les mines d'or sud-africaines. *Annales des Mines*, Paris, vol. 12, p. 3-49.
5. Journel, A. G. 1982. The indicator approach to estimation of spatial distributions. In: Johnson, T. B., and Barnes, R. J., eds, Proceedings of the 17^{th} APCOM International Symposium, Society of Mining Engineers of the AIME, New York, p. 793-806.
6. Krige, D. G., 1951, A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, dec. 1951, p. 119-139.
7. Lantuéjoul, C., 1990. Cours de sélectivité. Lecture Notes, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau.
8. Matheron, G. 1955. Applications des méthodes statistiques à l'évaluation des gisements, *in Annales des Mines*, 12, Paris, p. 50-75.
9. Matheron, G. 1955. Etude géostatistique du gisement de plomb de Bou-Kiama, Technical report, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau.
10. Matheron, G. 1962-63. Traité de géostatistique appliquée. Mémoires du Bureau de Recherches Géologiques et Minières. Tome I, no 14, Editions Technip, Paris, Tome II: le krigeage, no 24, Editions BRGM, Paris.
11. Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris. 306 p.

12. Matheron G., 1968, Lois stables et loi lognormale. Technical report, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau, 17 p.
13. Matheron, G. 1971. The theory of regionalized variables and its applications, Les Cahiers du Centre de Morphologie Mathématique, no. 5, Ecole des Mines de Paris, 211 p.
14. Matheron G., 1973, The intrinsic random functions and their applications, Adv. Appl. Prob. 5, 439-468.
15. Matheron G., 1973, Le krigeage disjonctif. Technical report, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau, 40 p.
16. Matheron G., 1974. Effet proportionnel et lognormalité ou le retour du serpent de mer. Technical report, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau, 43 p.
17. Matheron G., 1975, Le paramétrage technique des réserves. Technical report, Ecole des Mines de Paris, Centre de Géostatistique, Fontainebleau, 26 p.
18. Matheron, G. 1978. Estimer et choisir. Les Cahiers du Centre de Morphologie Mathématique 7. Ecole des Mines de Paris, Fontainebleau. 175 p.
19. Matheron G., 1984, The selectivity of the distributions and "the second principle of geostatistics". In Verly, G. et al eds. Geostatistics for natural resources characterization, Reidel, Dordrecht, Part 1, p. 421-433.
20. Matheron G., 1989, Estimating and choosing, Springer-Verlag, Berlin, 141 p.
21. Rivoirard, J., 1989, Models with orthogonal indicator residuals, in M. Armstrong (ed.), Geostatistics, Kluwer, Dordrecht, Vol. 1, p. 91-107.
22. Strebelle, S., 2002. Conditional simulations of complex geological structures using multi-point statistics*: Math. Geology*, v.34, no. 1, p. 1-21.

Springer

http://www.springer.com/978-0-387-20331-7

Space, Structure and Randomness
Contributions in Honor of Georges Matheron in the Fields of Geostatistics, Random Sets and Mathematical Morphology

Bilodeau, M.; Meyer, F.; Schmitt, M. (Eds.)
2005, XVIII, 398 p. 121 illus., Softcover
ISBN: 978-0-387-20331-7