Optimal weights for Location Dependent Moments

David F. Machuca-Mory and Clayton V. Deutsch

One approach for the estimation of location dependent moments is to weight the available samples according a function of their isotropic or anisotropic distance to a given location. These weights are then incorporated in the calculation of 1-point and 2-point moments. Several desirable properties must be fulfilled by them, such as smoothness, unbiasedness, positivity, global consistency and independency of units. The weights can be calibrated with declustering weights. The estimation of 2-point moments, such as the variogram, can be achieved from the weights assigned to sample pairs and an appropriate mixture rule of the sample weights involved (arithmetic average or geometric average). Several weighting functions can be used including inverse distance, Gaussian kernel and global ordinary kriging. The choice of the weighting function is more important than the choice of the mixture rule. Gaussian Kernel weighting proves to be very useful for location dependent moments estimation and yields weights that fulfill all the desirable properties. Two types of optimality criteria for weights are available: Prior criteria and posterior criteria. Prior criteria assess the suitability of weights for approximating the location dependent moments. Posterior criteria evaluate the performance of estimation and simulation with location dependent moments.

Introduction

In geostatistics, uncertainty at unsampled locations is modelled by the distribution of a Random Variable (RV). A group of RV's form a Random Function (RF) (Matheron, 1970). A fundamental assumption for spatial inference is that the moments describing the distributions and spatial interdependence of the RV's do not change by translation (Goovaerts, 1997). If this assumption is not deemed appropriate for an entire dataset, this can be subdivided within domains were the geological homogeneity may indicate the homogeneity of the attribute statistics (McLennan, 2007); however, local variation of some or all the moments may be still present and difficult to isolate within these domains. A widely studied case is when the mean presents a trend that can be deterministically modelled (Journel & Rossi, 1989). The a-priori modelling of the local variation of other moments is also important, especially when they may affect the results of the spatial prediction if considered.

These local moments can be obtained directly from available samples or from general knowledge. If the only information available is provided by the sample values, the moments at a particular location can be obtained by weighting the samples inversely proportional to their distance to such location. Thus, an important issue is how to determine the optimal sample weights for estimating the location dependent moments (LDM). These weights should fulfill some desirable properties, such as: unbiasedness, continuity, positivity, and they must account for declustering and anisotropy. There are several distance weighting functions that yield weights with these properties. The 1-point weights must be then converted in 2-point weights for their use in the calculation of location dependent measures of spatial variability by using one of the several mixture rule (Korvin G., 1982).

Several options are presented and discussed in this paper. The optimum choice of weights should yield the minimum error variance and fair reproduction of the uncertainty for estimation and simulation with location dependent moments. These criteria are discussed, but not tested here. Instead different choices of distance functions and mixture rules are tested for their capability of produce weights that better reproduce the location dependent variograms of synthetic examples.

Desirable Properties of weights for the estimation of location dependent moments

Location dependent moments are valid only locally. In a first instance, they can be calculated only in relation to given locations called "anchor points". If these moments are obtained only from available samples, we want that the contribution of sample values close to the anchor point be higher than those of farther samples. A straightforward way to attain this is to weight the samples inversely proportional to their distances to the anchor point, and then, include these weights in the calculation of the moments. Multiple

weighting functions can be used for such purpose; however it is desirable that the weights obtained from these functions share a number of desirable properties. The most important of these properties are smoothness, unbiasedness, correct filtering of local features, global consistency and independence of units. Besides these properties, they may incorporate declustering capabilities and may account for anisotropy.

Smoothness: Within a single domain and in the absence of noticeable discontinuities, weights should change smoothly between close locations. This is, if the separation distance between two anchor point locations $\mathbf{o}$ and $\mathbf{o}+d\varepsilon$ is very small, $d\varepsilon$, then the weights assigned to a sample $\mathbf{u}_i$ with respect to these anchor points are very similar:

$$\omega(\mathbf{o},\mathbf{u}_i) \approx \omega(\mathbf{o}+d\varepsilon,\mathbf{u}_i) \quad \forall i=1,...,n \tag{1}$$

Abrupt changes in the weights assigned to adjacent locations may be caused by a high steep in the shape of the weighting function close to the origin or if a discontinuous function is used. This can cause unwarranted instabilities in the local moments, such as outliers, when a sample is very close to an anchor point, as well as artifacts. Weights that fulfill this property are provided by smooth functions $f:\mathbb{R}^+ \rightarrow \mathbb{R}^+$ which are differentiable in $(0, \infty]$.

Unbiasedness and strict positivity: The distance based weights assigned to each sample can be regarded as proportional to their probability contribution to the local CDF. In this sense, the sum of the contributions should sum 1, which is the unbiasedness condition, and all the contributions must be positive. Thus, the weights assigned to the n samples with relation to an anchor point, **o,** must fulfill the next constraints:

$$\begin{cases} \omega(\mathbf{o},\mathbf{u}_i) > 0 & \forall i,...,n \\ \sum_{i=1}^{n} \omega(\mathbf{o},\mathbf{u}_i) = 1 \end{cases} \tag{2}$$

The strict positivity property allows the inclusion of all available samples in the calculation of the location dependent moments, which is relevant when their number is limited. Besides, the absence of zero weights avoids computational problems, and the absence of negative weights avoids the eventual occurrence of negative values for strictly positive moments.

Inverse proportionality to distance: The underlying assumption for obtaining the locally weighted moments is that closer samples are more representative than farther samples for a given location. In this sense the distance weighting function should yield continuously decreasing weights as the distance to the anchor point increases.

Correct filtering: Smoothly changing weights produce smooth changes in the location dependent moments. However, at the same time that they vary smoothly, the moments must be able to represent the local features without overfitting. In particular, if the local mean is obtained everywhere within a domain the correlation between the residuals and the local mean must be close to zero, thus:

$$Cov[m^*(\mathbf{u}),(Z(\mathbf{u})-m^*(\mathbf{u})] \approx 0 \qquad \forall \mathbf{u} \in \mathbf{D} \tag{3}$$

Global consistency: The same set of weights calculated at each location for all the neighbouring individual samples should be used for all the one-point and two-point local dependent statistical parameters. This would assure the mutual consistency of these parameters. Besides that, this would result in an increased efficiency at the estimation of the different location dependent statistical parameters.

Independence of Units: the same weights should be obtained regardless the distance units used.

Account for declustering and anisotropy: If data clusters are present, the total contribution of these samples in the calculation of a location dependent moment will be greater of the corresponding to isolated samples. In order to avoid this and correct the available information by the local sample density, weights may include declustering properties.

If a strong anisotropic orientation is known a priori, weights may be adjusted by the orientation and degree of this anisotropy. Thus, the contribution of those samples in the direction of the major anisotropy radio will be increased. The modifications imposed to weights in order to account for declustering and anisotropy are discussed next.

Incorporation of declustering and anisotropy

If samples are highly clustered the estimation of the one-point location dependent parameters is biased by the predominance of information at densely sampled regions, which usually are coincident with high grade regions. This bias is translated to the two-point moments, where the estimates at short lag separations reflect everywhere the two-point properties of the highly clustered areas.

A set of weights that fulfill the properties indicated above may be still yield to biased results caused by samples clustering. The idea is to introduce a correction or calibration of distance based weights by declustering weights.

Self Calibration to account for declustering: Several declustering methods are available, among them: cell declustering (Deutsch, 1989), polygonal declustering (Deutsch, 2002), and kriging weights declustering (Brandon & Deutsch, 2007). Whatever method is used to obtain the declustering weights, if the distance based weights incorporate declustering capability, the next property must be fulfilled:

$$\bar{\omega}(\mathbf{u}_i) = \frac{1}{D}\int_D \omega(\mathbf{o};\mathbf{u}_i)d\mathbf{o} = \varsigma(\mathbf{u}_i) \qquad \forall\ i = 1,...,n \tag{4}$$

$$\bar{\omega}(\mathbf{u}_i) = \frac{1}{P}\sum_{\alpha=1}^{p} \omega(\mathbf{o}_\alpha;\mathbf{u}_i) = \varsigma(\mathbf{u}_i) \qquad \forall\ i = 1,...,n \tag{5}$$

This is, the average of all the weights calculated for a sample at the location $\mathbf{u}_i$ with respect to all the points **o** over the domain *D* might be equivalent to its corresponding declustering weight, $\varsigma(\mathbf{u}_i)$. If the distance weighting method employed does not yield weights that fulfill this condition they can be modified using the self calibration method , as it follows.

Given the one-point declustering weights $\varsigma(\mathbf{u}_i)$, which can be obtained by methods such as cell declustering, polygons, or global kriging, the weights for the calculation of local moments at an anchor point, **o**, can be calibrated in order to reproduce declustering by doing:

$$\omega_{updated}(\mathbf{o};\mathbf{u}_i) = \omega_{original}(\mathbf{o};\mathbf{u}_i)\frac{\varsigma(\mathbf{u}_i)}{\bar{\omega}(\mathbf{u}_i)} \qquad \forall\ i = 1,...,n \tag{6}$$

Therefore:

$$\frac{1}{P}\sum_{\alpha=1}^{p} \omega_{updated}(\mathbf{o}_\alpha;\mathbf{u}_i) = \varsigma(\mathbf{u}_i) \qquad \forall\ i = 1,...,n \tag{7}$$

Anisotropic weighting

Strong anisotropy can be present within a statistical homogeneous domain. This can be related to the presence of semi-parallel layering or a prevalent direction of deposition. If a dominant global direction of anisotropy can be identified within a domain, anisotropic distances may be considered for in the weighting of local moments.

The anisotropy used for distance weighting should be weaker than the true anisotropy of the variable under study. Proceeding in this way is necessary in order to allow enough flexibility in the conditioning of the location dependent parameters by the available data.

If the location of a sample *i* is given by its Cartesian coordinates $\mathbf{u}_i = (x_i, y_i, z_i)$ and those of the control point are $\mathbf{u}_0 = (x_\mathbf{o}, y_\mathbf{o}, z_\mathbf{o})$, the vector formed by their difference is:

$$\overline{\mathbf{o}\,\mathbf{u}_i} = (dx_{\mathbf{o}}x_i, y_{\mathbf{o}}y_i, z_{\mathbf{o}}z_i) = (x_i - x_{\mathbf{o}}, y_i - y_{\mathbf{o}}, z_i - z_{\mathbf{o}}) \tag{8}$$

Given the anisotropy ratios:

a_1=length of the major axis of anisotropy/ minimum horizontal range

a_2= maximum horizontal range / maximum vertical range

and the angles:

α = angle between the major axis of anisotropy and the E-W axis.

β = angle between the major axis of anisotropy and the E-W axis.

θ = angle of rotation of the minor axis about the major axis of anisotropy.

The anisotropic distance between the control point and the sample can be found using the next transformation matrix (Goovaerts, 1997):

$$\begin{bmatrix} d'x_{\mathbf{o}}x_i \\ d'y_{\mathbf{o}}y_i \\ d'z_{\mathbf{o}}z_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a_1 & 0 \\ 0 & 0 & a_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sin\theta & -\cos\theta \\ 0 & \cos\theta & \sin\theta \end{bmatrix} \cdot \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ \sin\beta & 0 & \cos\beta \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ \sin\alpha & -\cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} dx_{\mathbf{o}}x_i \\ dy_{\mathbf{o}}y_i \\ dz_{\mathbf{o}}z_i \end{bmatrix} \tag{9}$$

and the well known expression:

$$d'(\mathbf{u}_{\mathbf{o}},\mathbf{u}_i) = \sqrt{\left(d'x_{\mathbf{o}}x_i\right)^2 + \left(d'y_{\mathbf{o}}y_i\right)^2 + \left(d'y_{\mathbf{o}}y_i\right)^2} \tag{10}$$

This anisotropic distance can be incorporated in any distance weighting method.

Distance weighting functions

Three distance based weighting methods are considered here. Among many possible methods and functions, these were chosen because they fulfill all or most of the properties defined above and because their straightforward implementation. The three distance weighting methods analysed here are the Inverse Distance Weighting, the Gaussian Kernel and Global Kriging.

Inverse Distance

The inverse distance weighting method (Shepard, 1968) is very popular in spatial interpolation. In this method samples are weighted inversely proportional to their distance to the estimated point. This distance can be affected by a power, in order to control the degree of smoothing in the interpolation. The inverse distance weighting function is expressed as:

$$\omega(\mathbf{o},\mathbf{u}_i) = \frac{\dfrac{1}{\left(d(\mathbf{o},\mathbf{u}_i)+c\right)^p}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\left(d(\mathbf{o},\mathbf{u}_i)+c\right)^p}} \tag{11}$$

Where $d(\mathbf{o},\mathbf{u}_i)$ is the distance between the control point and the sample location, p is the power value and c is a constant that defines the maximum value of the function and avoids computation problems when the sample is very close to the control point. If the c value is comparable or bigger than the sample spacing, it also influences the smoothing.

The inverse distance weights decrease quickly for short distances but slowly for long distances (see figure 1). This can cause the instability of local moments when a sample is very close to an anchor point. Increasing the value of the c constant reduces this problem, as does decreasing the power p, but the price may be an increased smoothing and reduced resolution of the local dependent moments. The optimum parameters p and c can be obtaining by minimizing the average estimation variance (Babak & Deutsch, 2008), which is given by:

$$\sigma_{est}^2 = \sigma^2 - 2\sum_{i=1}^{n}\omega(\mathbf{o},\mathbf{u}_i)C(\mathbf{u}_i,\mathbf{o}) + \sum_{i=1}^{n}\sum_{j=1}^{n}\omega(\mathbf{o},\mathbf{u}_i)\omega(\mathbf{o},\mathbf{u}_j)C(\mathbf{u}_i,\mathbf{u}_j) \tag{12}$$

Where the covariance C is calculated and modelled from all available samples and assuming stationarity and isotropy.

Gaussian Kernel

Gaussian kernels are frequently used in non-parametric regression and spatial statistics (Schabenberger & Gotway, 2005). At difference of the inverse distance weighting function, which decreases quickly, the well known shape of the Gaussian bell assigns high and medium weights to a wider range of distances (see figure 1) This range is controlled by the amplitude of the curve, which is equivalent to the standard deviation of the Gaussian distribution, σ, also called *bandwidth in* the spatial statistics context. Weights are sensitive to the value of σ, this value must be chosen in relation to sample spacing and the non-stationary features in a domain. The Gaussian Kernel function is given by:

$$\omega(\mathbf{o},\mathbf{u}_i) = \varepsilon + \frac{\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(d(\mathbf{o},\mathbf{u}_i)\right)^2}{2\sigma^2}\right)}{\sum_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(d(\mathbf{o},\mathbf{u}_i)\right)^2}{2\sigma^2}\right)} \tag{13}$$

where ε is a very small constant that avoids computational problems when the weights are smaller than 1.0E-21. Weights thus obtained are strictly positive and vary smoothly. However they account for declustering only when the amplitude is very narrow, which can cause the instability of the location dependent moments. Therefore, the weights for the estimation of locally varying moments must be calibrated by the declustering weights using the methodology explained above. An advantage of using the Gaussian Kernel is that only one parameter, the bandwidth, must be adjusted. This can be obtaining by the minimization of the average estimation variance (12) criterion as in the case of inverse distance weighting.

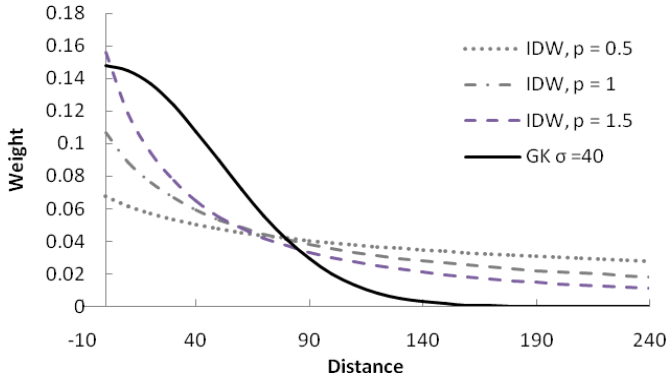

Figure 1: Inverse distance (IDW) and Gaussian kernel (GK) weighting functions.

Global Ordinary Kriging

Global Ordinary Kriging (GOK) is performed with all the available samples. It renders smooth estimates, and the obtained weights change smoothly and they account for declustering (Neufeld & Wilde, 2005). However they are not always strictly positive, and not necessarily decrease aas the distance increases. Farther samples and samples that are screened by closer samples to the estimated point usually get negative weights. These negative weights could be set to zero, but that means a considerable loss of information, sometimes related to not so farther, but screened, samples, as it can be observed in figure 2. Moreover, a stationary variogram must be assumed valid for the entire domain first, in order to obtain the weights for the estimation of the location dependent moments. The lower the nugget effect of this variogram model, the higher the number of negative weights that GOK yields. A primary stationary variogram can be useful for incorporating anisotropic distances in the calculation of the weights needed for the location dependent

moments estimation. But, the presence of sometimes abundant negative weights hinders the applicability of GOK for such task.

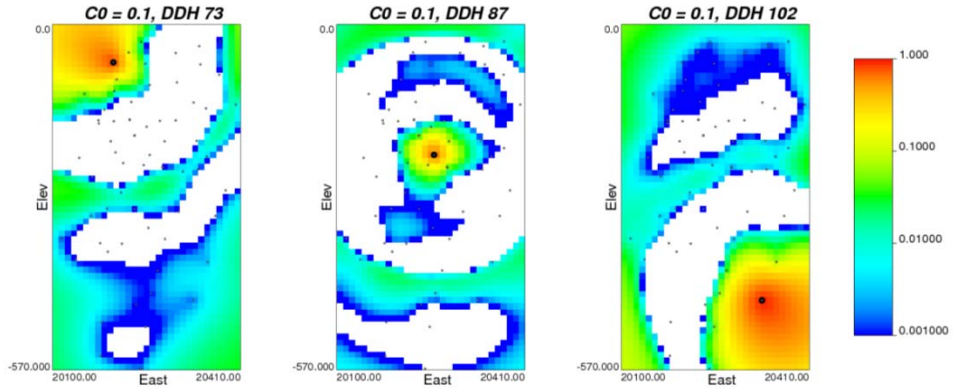

Figure 2: spatial distribution of kriging weights assigned to particular samples located in 3 different positions (black dots), in the white areas the sample gets a negative weight.

Weighting schemas and mixing rules for 2-point weights

So far, only one-point distance weighting has been covered here. However, the estimation of one-point and two-point local dependent moments require an integrated approach for one-point and two-point distance weighting. Two basic weighting schemas for two-point distance weighting are proposed. The first one consists in weighting the sample pairs by the distance of their midpoint to a control point. The second schema considers the distances from the pair's end-points to the control point. Both schemas are illustrated in figure 3a and 3b.

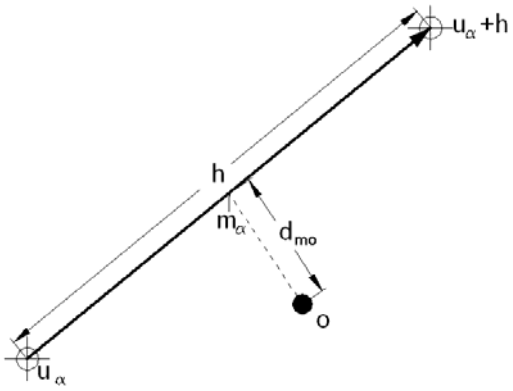

Figure 3a: Midpoint weighting schema

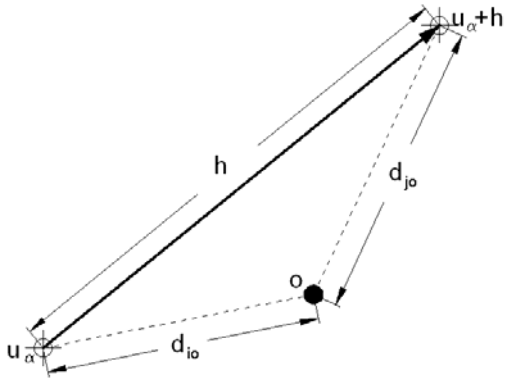

Figure 3b: Endpoints weighting schema

Any of both schemas converge to one-point weighting distance as the lag separation *h* approaches zero. A mixed schema would involve the joint distance weighting of the endpoints and the midpoint, this could be extended to the distance weighting of all the points in the segment of length *h* and by integrating the weighting function from the tail to the head of the segment formed by the sample pair. This later approach can be very demanding in computer resources and not necessarily better.

If the midpoint weighting schema is adopted, the pair distance weighting is reduced to a one-point distance weighting schema. However this schema is not recommendable since it may introduce a "diplopic effect" in the calculation of the local variogram sill. This is illustrated in figure 4, where the local variance and local sills are calculated for a simulated data set. This data set has an area of higher variance at the middle (top left), which is reproduced by the local variance calculated with individual sample weighting (top right). The local sill calculated with the midpoint weighting schema presents two areas of high values (bottom left), while the local sill calculated with endpoints weighting schema reproduce better the high variance area.

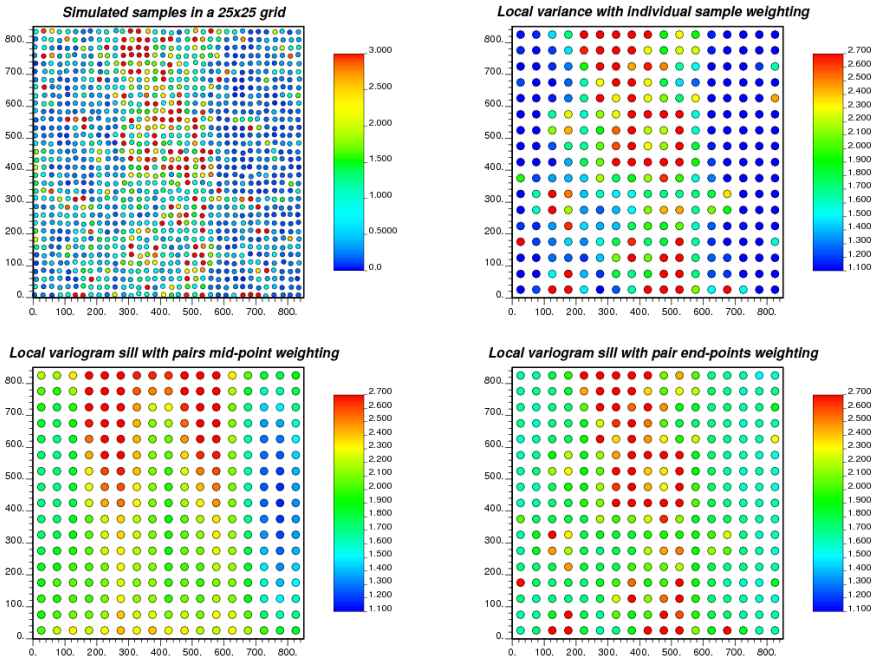

Figure 4: Simulated dataset with a high local variance in the central area (top left). Local variance calculated with individual sample weighting (Top right). Local variogram sill calculated with pair's mid-point weighting (bottom left), note double high variance area. Local variogram sill calculated with pair's mid-point weighting (bottom left),

In the case of the endpoints weighting schema a mixture rule must established in order to combine the weights assigned to each endpoint. Mixture rules have been defined by Korvin (1982) for calculating the value of a property for a composite material formed by two different components or phases. Given two phases with respective property values g_1 and g_2, and volume fractions ϕ and $1-\phi$, respectively, the general mixture rule is expressed as:

$$M(g_1, g_2, \phi, 1-\phi) = \left[\phi g_1^t + (1-\phi) g_2^t\right]^{1/t} \tag{11}$$

This mixture rule can be applied for calculating the composite weight of a sample pair. By doing $\phi = 0.5$ and $g_1 = \omega(\mathbf{o}, \mathbf{u}_i)$, $g_1 = \omega(\mathbf{o}, \mathbf{u}_i)$, the mixture rule for combining the weight of a sample at a location $\mathbf{u}_i$ and its pair separated by a vector distance **h**, is expressed as:

$$\omega(\mathbf{o}, \mathbf{u}_i, \mathbf{u}_h) = \frac{\left[\omega^t(\mathbf{o}, \mathbf{u}_i) + \omega^t(\mathbf{o}, \mathbf{u}_h)\right]^{1/t}}{2} \tag{12}$$

For t =1 the mixture rule is the arithmetic average of the weights:

$$\omega(\mathbf{o}, \mathbf{u}_i, \mathbf{u}_h) = \frac{\omega(\mathbf{o}, \mathbf{u}_i) + \omega(\mathbf{u}_0, \mathbf{u}_h)}{2} \tag{13}$$

When the parameter t approaches 0, the expression converge to the geometric average:

$$\omega(\mathbf{o}, \mathbf{u}_i, \mathbf{u}_h) = \sqrt{\omega(\mathbf{o}, \mathbf{u}_i) \cdot \omega(\mathbf{o}, \mathbf{u}_h)} \tag{14}$$

The arithmetic average is resistant to very low weights and tends to favour high weights, by the contrary, the geometric average is more sensible to low weights. Thus, the use of the arithmetic average favour the samples closely located to the control point in a greater measure than when the geometric average is used.

It can be easily demonstrated that if the one-point weights fulfill the properties above mentioned, the two-point weights obtained by any of the mixture rules will also fulfill most of the same properties, such as smoothness, positivity, independency of units and global consistency. The condition of unbiasedness of the

one-point weight is not translated to the unbiadsedness of the 2-point weights, but it can be achieved by doing:

$$\varpi(\mathbf{o},\mathbf{u}_i,\mathbf{u}_h)=\frac{\omega(\mathbf{o},\mathbf{u}_i,\mathbf{u}_\mathbf{h})}{\sum_{i=1}^{N_\mathbf{h}}\omega(\mathbf{o},\mathbf{u}_i,\mathbf{u}_\mathbf{h})} \tag{15}$$

If the two-point weights are constructed by the mixture of the pair endpoints weights, the former must be bounded by the values of the individual one-point weights that are used to calculate them. Thus:

$$\min\left[\omega(\mathbf{o},\mathbf{u}_i),\omega(\mathbf{o},\mathbf{u}_h)\right]<\omega(\mathbf{o},\mathbf{u}_i,\mathbf{u}_h)<\max\left[\omega(\mathbf{o},\mathbf{u}_i),\omega(\mathbf{o},\mathbf{u}_h)\right] \tag{16}$$

Testing the weighting criteria

The different tests for assessing the optimality of the weighting criteria can be grouped as prior and posterior tests. Prior tests evaluate the capability of weights to depict the local parameters. Posterior tests evaluate the performance of estimation and simulation with location dependents moments.

Prior tests

These tests respond to several optimality criteria for 1-point and 2-point weights. For 1-point weights, the minimization of the average variance of estimation, presented in (12), can be used as optimality criteria. Additionally, the correct filtering property expressed in (3) should be also tested.

The optimum weights for local 2-point moments can be obtained using the mixture rules on 1-point weights tested by the criteria above. The assumption of optimality of 2-point weights based in the optimality of 1-point weights is supported on the global consistency criterion of weights. By other side, in the eventual case that the local true variograms are accessible, such in the case of having a representative training image, the next criterion can be tested:

$$E[\gamma^*(\mathbf{u},\mathbf{h})-\gamma(\mathbf{u},\mathbf{h})]^2\approx 0 \quad \forall\mathbf{u}\in\mathbf{D} \tag{17}$$

where $\gamma^*(\mathbf{o},\mathbf{h})$ and $\gamma(\mathbf{o},\mathbf{h})$ are the estimated and true local variograms. This criterion is applied next to evaluate the performance of inverse distance and Gaussian kernel weighting for approximating the local variograms of a synthetic data set. This data set was produced by Gaussian simulation in two separate domains with different direction of anisotropy. The variograms of both domains are spherical with a major range of 20 pixels and a minor range of 5 pixels, but the azimuth of the major range of anisotropy for the west domain is 0º, while for the east domain is 90º. The entire field was sampled in a quasi regular mesh of 4 x 4 pixels (see Figure 5, next).

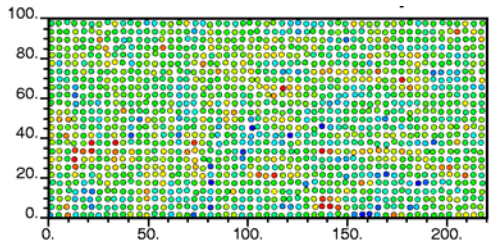

Figure 5: Location map of simulated samples

The 0-intercept of the inverse distance and Gaussian kernel functions chosen are close of each other (see figure 1). The inverse distance power is 1.5 and the offset is 40 distance units. For the Gaussian kernel the amplitude was set at 50 distance units. The experimental local variograms were calculated with respect to control points located in a 10 x 10 grid, and fitted semi-automatically.

The choice of weighting function becomes more important than the choice of weighting schema or of the mixture rule. This is demonstrated by comparing the square error of the difference between the estimated location dependent variograms and the "true" local variograms of the two different areas in the synthetic data in the figure 5. The location dependent variograms estimated with Gaussian kernel weighting are consistently closer to the "true" local variograms than those estimated using inverse distance weighting. In

this case, the improvement in the estimation of the local variogram can be higher than 50% when Gaussian kernel weighting is used instead of inverse distance weighting. By other side, the choice of mixture rule for the sample weights has also a minimal impact in comparison of the choice of weighting function (see figures 6a and 6b).

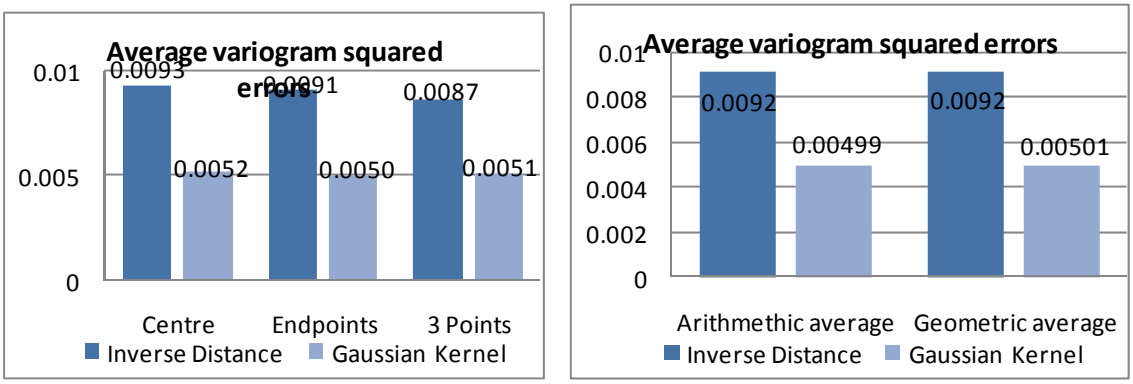

Figure 6: Average variogram squared error using (a) different weighting schemas and (b) different mixing rules.

Posterior tests

Posterior tests evaluate the spatial prediction performance of the location dependent moments obtained with a particular set of weights. These tests are based in criteria of accuracy and precision. In estimation with location dependent moments, accuracy is assessed by the average square error between true, $Z(\mathbf{u})$, and estimated, $Z^*(\mathbf{u})$, values:

$$E\left\{\left[Z^*(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\} \tag{18}$$

This should be low for improved performance. While the coefficient of correlation between true and estimated values should be high:

$$\frac{Cov\left[Z^*(\mathbf{u}), Z(\mathbf{u})\right]}{\sigma_Z^* \sigma_Z}$$

These statistics are obtained by cross-validation. By the other side, precision is increased when the variance of estimates is higher and when the average estimation variance is lower.

In simulation with location dependent moments the prior global distribution, $F_{prior}(z)$, is locally updated. The posterior local distributions, $F_{posterior}(z)$, must present a narrow uncertainty. This can be expressed by the minimization of the local distribution variance:

$$E\left\{Var[F_{post}(\mathbf{u}; z)]\right\}$$

However at the same time the local proportions must be fairly reproduced. Very narrow posterior narrow distributions would indicate an increased precision but at the expense of accuracy. Accuracy plots (Deutsch, 1996) can be used for evaluate the trade-off between accuracy and precision.

Conclusions

The mathematical properties of weights used for the estimation of location dependent moments are defined. These properties are unbiasedness, positivity, correct filtering, independency of units, and global consistency. Although these weights are not aimed for declustering, they can be calibrated in order to account for sample clustering. If a strong anisotropy is present, a directional correction in the distances used in the weighting functions can be incorporated in the distance weighting function.

Three weighting functions were analysed: inverse distance, Gaussian kernel, and Global Ordinary Kriging. The last one is discarded because it also yield a considerable number of non-positive weights. Moreover, the weights obtained by Global Ordinary Kriging heavily rely on the definition of a primary stationary variogram. On the contrary, inverse distance and Gaussian kernel weights require fewer parameters and satisfy most of the desirable properties for estimation of location dependent moments. These weights are calibrated by the weights obtained by cell, polygonal or global kriging declustering methods.

The pairs weighting schema can be based on its midpoint or on its endpoints. In the second case one-point weights of individual samples must be combined according mixture rules such as the geometric, arithmetic or harmonic averages. The improvement in the estimation of local dependent moments related to the choice of the pairs weighting schema and the mixture rule is very low when compared with the choice of the weighting function. The average square error in the estimation of the local variogram if the Gaussian Kernel weighting function is used can be more than 50% lower than when the inverse distance weighting function is applied

References

Babak, O., & Deutsch, C. V. (2008). Statistical approach to inverse distance interpolation. *Stochastic Environmental Research and Risk Assessment* .

Brandon, W., & Deutsch, C. V. (2007). *Wide array declustering for representative distributions (The ultimate DECLUS program).* University of Alberta. Edmonton: Centre for Computational Geostatistics (CCG).

Deutsch, C. V. (1989). Declus: A FORTRAN 77 program for determining optimul spatial declustering weights. *Computers & Geosciences , 15* (3), 325-332.

Deutsch, C. V. (1996). Directc assessment of local accuracy and precision. In E. Y. Baafi, & N. A. Schofield (Ed.), *Geostatistics Wollongong '96. 1*, pp. 115-125. Wollongong: Kluwer Academic.

Deutsch, C. V. (2002). *Geostatistical Reservoir Modeling.* New York: Oxford University Press.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

Journel, A. G., & Rossi, M. E. (1989). When do we need a Trend Model in Kriging? *Mathematical Geology , 21* (7), 715- 739.

Korvin, G. (1982). Axiomatic characterization of the general mixture rule. *Geoexploration* (19), 267-276.

Matheron, G. (1970). *La Théorie des variables régionalisées et ses applications.* . Les cahiers du CMM, Fasc. 5, Ecole des Mines de Paris, Fontainebleau.

McLennan, J. (2007). *The Desicion of Stationarity.* Civil & Environmental Engineering. Edmonton: University of Alberta.

Neufeld, C., & Wilde, B. (2005). *A global kriging program for artifact-free maps.* Edmonton: Centre for Computational Geostatistics.

Schabenberger, O., & Gotway, P. (2005). *Statistical mehotds for spatial data nalysis.* Boca Raton: Chapman & Hall/CRD.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 ACM National Conference* (pp. 517-524). Association for Computing Machinery ACM.