On the Smoothness Properties of the Large Scale Random Process

S. N. Elogne and O. Leuangthong

The geostatistics scaling results as described in (Kupfersberger et al, 1998) are essentially derived from the point scale process using a regularization operator (Journel and Huijbrets, 1978, p. 171) and hold under the assumption that the nested covariance models involved in the finer scale spatial structure remain the same as those involved in the larger scale covariance structure. Because in practice discrete versions of the regularization operator are considered, applications of these scaling theories must be conducted with care. The main objective of this paper is to provide scaling results that include the changes in the smoothness property the auto-covariance function of the large scale process.

Introduction

An essential task in reservoir characterization relies on integrating various data sources to improve reservoir simulation models (Efendiev et al, 2005). In geostatistics, one considers that observed data sets from well cores and log traces as coming from a fine scale process. Others observed data sets such as seismic impedance are considered as large scale information. Geostatistical applications often rely on point support observed data and large scale simulated realizations are investigated. Prior knowledge about the spatial dependence structure of a process at a modeling volume cell (usually greater than the scale of the collected data) can be of interest when integrating data from different data sources of different lengths scales (Kupfersberger et al, 1998) (Frykman and Deutsch, 1999). The key point is that the correct modeling of the large scale spatial structure are crucial for integrating multiscale data using techniques such as cokriging (Xu et al, 1992), sequential Gaussian simulation with block kriging.

To our knowledge the geostatistics scaling laws (GSLs) as reviewed by (Kupfersberger et al, 1998) is the only approach to address that problem. The GSLs were derived in the early days of geostatistics and have been extensively used in modeling the core, log and seismic data (Frykman and Deutsch, 1999), (Frykman and Deutsch, 2002). The scaling laws are concerned with the changes to the spatial dependence structure, determined by nugget effect, the range and the sill.

Consider a second order stationary processes Z_v defined on a scale v (not necessarily the point scale) and let C_v be its auto-covariance function with finite variance σ_v^2. Moreover assume that Z_v is isotropic and write that

$$C_v(\mathbf{h}) = \sum_{k=0}^{L} a_k \rho_k(r) \tag{1}$$

where $r = \|\mathbf{h}\|$, $a_0 = \sigma_{\varepsilon_v}^2$ is the nugget component, ρ_k are positive definite correlation functions and a_k are real numbers for $k = 1, \ldots, L$. The anisotropic case can be studied in a similar way by selecting preferential directions. Let V be any scale usually large than the scale v and let Z_V be the associated process with covariance structure C_V. The scaling laws write the auto-covariance function C_V as

$$C_V(V_\alpha, V_\alpha + \mathbf{h}) = \sum_{k=0}^{L} b_k \rho_k(r) \tag{2}$$

with

$$\begin{aligned} &b_0 = \sigma_{\varepsilon_v}^2 = \eta_{v,V} a_0, \qquad r_V = r_v + (1 - \eta_{v,V})|V| \\ &b_k = \frac{\bar{\varphi}(V,V)}{\bar{\varphi}(v,v)} a_k \quad \text{for} \quad k = 1, \ldots, L \end{aligned} \tag{3}$$

where $\eta_{v,V} = \dfrac{|v|}{|V|}$ and $\bar{\varphi}(v,v)$ represents the average covariance given by

$$\overline{\varphi}(v,v)=\frac{1}{|v|^2}\int_v d\mathbf{u}\int_v d\mathbf{u}' c_\bullet(\mathbf{u},\mathbf{u}')$$

Note that the ratio $\eta_{v,V}$ appears only in the nugget b_0 and on the range r_V and is independent of the sill b_k of each nested correlation structure ρ_k for $k=1,\ldots,L$. They are at least two major drawbacks on using the scaling results.

(i) First, the theory behind the scaling starts from the point scale process $Z_\bullet(\mathbf{u})$ using the regularization operator

$$Z_v(v_\alpha)=\frac{1}{|v_\alpha|}\int_{v_\alpha} d\mathbf{u}\, Z_\bullet(\mathbf{u}) \quad \text{and} \quad Z_V(V_\beta)=\frac{1}{|V_\beta|}\int_{V_\beta} d\mathbf{u}\, Z_\bullet(\mathbf{u}) \tag{4}$$

But in practice the above characterization (4) is never used as one relies on a discretized version of (4) using a specific number of discretized nodes from the finer scale to the larger scale. As a consequence the nugget effect and the range of the coarse scale process given in (3) do not hold since the exact volumes of the scales are never called.

(ii) One also observes that the smoothness property of the auto-covariance function is preserved through the scaling process as the nested correlation functions ρ_k remain unchanged from (1) to (2). Since the auto-covariance function plays an important role in any geostatistical modeling, this assumption should be reviewed as the averaging operator may lead to a change in the spatial structure.

The main goal of this paper is to show that the auto-covariance of the coarser scale process is differentiable in the mean square sense regardless of the initial smoothness property of the finer scale process because of the averaging inherent in the data. As a result if C_v is modeled through a linear combination of non differentiable nested structures (spherical, exponential, etc…) then C_V will be modeled through differentiable structures (Gaussian, Cauchy, Matern etc...).

Review

Consider now a larger scale V such that any of its block V_α can be written as a partition of a finite set of elements $v_{1,\alpha},\ldots,v_{N_v^V,\alpha}$ of the fine scale v as $V_\alpha=\bigcup_{i=1}^{N_v^V} v_{i,\alpha}$ where the N_v^V denotes the number of blocks of the finer scale that discretize any block of the larger scale. Define the up-scaled process Z_V as

$$Z_V(V_\alpha)=\frac{1}{N_v^V}\sum_{i=1}^{N_v^V} Z_v(v_{i,\alpha}). \tag{5}$$

More generally for a deterministic, bounded and integrable function $\omega(v)$ over any volume finite volume v (the averaging kernel $\omega(v)$ allows variable weight within a specific block), we generalize Equation (5) as

$$Z_V(V_\alpha)=\frac{1}{\sum_{i=1}^{N_v^V}\omega(v_{i,\alpha})}\left(\sum_{i=1}^{N_v^V}\omega(v_{i,\alpha})Z_v(v_{i,\alpha})\right) \tag{6}$$

such that $\sum_{i=1}^{N_v^V}\omega(v_{i,\alpha})\neq 0$. Recall that Z_V is well defined if and only if (Christakos 1992)

$$\mathrm{E}\left[Z_V(V_\alpha)\right]^2<\infty \text{ that is } \quad \sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V}\omega(v_{i,\alpha})\omega(v_{j,\alpha})C_v(v_{i,\alpha},v_{j,\alpha})<\infty \tag{7}$$

for any V_α. If μ_v represents the stochastic mean of Z_v over the domain of interest, then the regularization operator (6) yields $\mu_V=\mu_v$, that is the stochastic mean is preserved.

Nugget effect

Any block $v_{i,\alpha}$ at the scale will be classically represented by its center $\mathbf{u}_{i,\alpha}^{v}$ and similarly any block V_α at the coarse scale V will be represented by its center $\mathbf{u}_\alpha^V$. For any block $v_{i,\alpha}$, the random variable $Z_v\left(v_{i,\alpha}\right)$ can decomposed (embedded) into a *smooth correlated fluctuation* X_v and a zero *random field noise* ε_v such that

$$Z_v\left(v_{i,\alpha}\right)=\mu_v+X_v\left(v_{i,\alpha}\right)+\varepsilon_v\left(v_{i,\alpha}\right). \tag{8}$$

The random field ε_v represents non-resolved inherent variability, purely random additive noise or random measurement errors (Cressie 1993, Section 2.3) (Christakos 1992, Section 7.4). For any α,α' and blocks v_α and $v_{\alpha'}$ it is assumed that

$$\text{(A1) } \operatorname{Cov}\left(X_v\left(v_\alpha\right),\varepsilon_v\left(v_{\alpha'}\right)\right)=0$$

$$\text{(A2) } \operatorname{Cov}\left(\varepsilon_v\left(v_\alpha\right),\varepsilon_v\left(v_{\alpha'}\right)\right)=\delta\left(v_\alpha,v_{\alpha'}\right)\operatorname{Var}\left[\varepsilon_v\left(v_\alpha\right)\right]$$

where the function $\delta\left(v_\alpha,v_{\alpha'}\right)$ takes the value one if $v_\alpha=v_{\alpha'}$ and zero otherwise. Using the regularization (5) and using the assumptions $(\mathrm{A1})$ and $(\mathrm{A2})$ one gets

$$\begin{aligned} C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_{\alpha'}^V\right)&=\left(\frac{1}{N_v^V}\right)^2\sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V}C_v\left(v_{i,\alpha},v_{j,\alpha'}\right)\\ &=\left(\frac{1}{N_v^V}\right)^2\sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V}C_v\left(\mathbf{u}_{i,\alpha}^{v},\mathbf{u}_{j,\alpha'}^{v}\right)\end{aligned} \tag{9}$$

Obviously the random process Z_V is not isotropic as $C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_{\alpha'}^V\right)$ does not depend on the distance between $\mathbf{u}_\alpha^V$ and $\mathbf{u}_{\alpha'}^V$. Calculating the variance of Z_V is as

$$\begin{aligned}\operatorname{Var}\left(Z_V\left(V_\alpha\right)\right)&=\left(\frac{1}{N_v^V}\right)^2\operatorname{Var}\left(\sum_{i=1}^{N_v^V}\left[X_v\left(v_{i,\alpha}\right)+\varepsilon_v\left(v_{i,\alpha}\right)\right]\right)\\ &=\left(\frac{1}{N_v^V}\right)^2\sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V}C_v\left(v_{i,\alpha},v_{j,\alpha}\right)+\frac{1}{N_v^V}\sigma_{\varepsilon_v}^2\end{aligned} \tag{10}$$

It then follows from the equations (9) and (10) that the nugget effect of the coarse scale process Z_V is as

$$\sigma_{\varepsilon_V}^2=\frac{1}{N_v^V}\sigma_{\varepsilon_v}^2 \tag{11}$$

In others words the nugget effect from a finer scale v to any coarser scale V decreases proportionally to the inverse of N_v^V. This says that in case where the quantity N_v^V is to be chosen, one can reduce the effect of $\sigma_{\varepsilon_V}^2$ by taking large values of N_v^V. This can only be possible if the finer scale is assumed to be the point scale. In any case, the nugget component does not vanish as it depends of both quantities $\sigma_{\varepsilon_v}^2$ and N_v^V. Similarly by using the weighted linear average given in (6) we derive that

$$\sigma_{\varepsilon_V}^2=\sigma_{\varepsilon_v}^2\sum_{i=1}^{N_v^V}\left(\frac{\omega\left(v_{i,\alpha}\right)}{\sum_{k=1}^{N_v^V}\omega\left(v_{k,\alpha}\right)}\right)^2 \tag{12}$$

Since the non-uniform weight function ω is assumed to take nonnegative values, we easily find that

$$\sum_{v_j \subset V_\beta} \left(\frac{\omega(v_j)}{\sum_{v_k \subset V_\beta} \omega(v_k)} \right)^2 < 1$$

and derive that $\sigma^2_{\varepsilon_V} < \sigma^2_{\varepsilon_v}$, that is the nugget component decreases.

The main objective of this paper is to show that when N_v^V becomes increasingly large, the auto covariance function of the large scale is twice differentiable with respect to any direction.

Mean square differentiability

Recall that a second order stationary process $Z(\mathbf{s})$ defined on a support $D \subset \mathbb{R}^d$ with covariance function $c(\mathbf{s}, \mathbf{s}')$ is said to be mean square differentiable with respect to any direction $\boldsymbol{\varepsilon}_p$ (Christakos 1992, p. 44) , (Abrahamsen 2003) if and only if

$$\frac{\partial^2 c(\mathbf{s}, \mathbf{s}')}{\partial s_p^2} \tag{13}$$

exists and is finite at all diagonal $\mathbf{s} = \mathbf{s}'$ where $\mathbf{s} = (s_1, \ldots, s_p, \ldots, s_d)$. The mean square differentiability is equivalent to

$$\left. \frac{\partial c(\mathbf{s}, \mathbf{s}+\mathbf{h})}{\partial s_p} \right|_{\mathbf{h}=\mathbf{0}} = 0 \tag{14}$$

(Christakos 1992, p. 62). Thus if Equation (14) holds, then the quantity $\frac{\partial^2 c(\mathbf{s}, \mathbf{s}')}{\partial s_p^2}$ exists and is finite.

Consider the up-scaled process $Z_V(V_\alpha)$ as defined in (5) with covariance function

$$C_V(\mathbf{u}_\alpha^V, \mathbf{u}_{\alpha'}^V) = \left(\frac{1}{N_v^V} \right)^2 \sum_{i=1}^{N_v^V} \sum_{j=1}^{N_v^V} C_v(\mathbf{u}_{i,\alpha}^v, \mathbf{u}_{j,\alpha'}^v).$$

Using Equation (1) , one has

$$C_V(\mathbf{u}_\alpha^V, \mathbf{u}_{\alpha'}^V) = \left(\frac{1}{N_v^V} \right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i=1}^{N_v^V} \sum_{j=1}^{N_v^V} \rho_k \left(\left\| \mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha'}^v \right\| \right) \right]. \tag{15}$$

We have the following lemma.

Lemma 1 Assume that the auto-covariance structure $C_v(\mathbf{h})$ of the fine scale process is differentiable at the right of zero, that is

$$\sum_{k=1}^{L} a_k \left. \frac{d\rho_k(r)}{dr} \right|_{r=0_+} < \infty \tag{16}$$

Then the auto-covariance C_V is twice differentiable with respect to any direction $\boldsymbol{\varepsilon}_p$ at the diagonal $\mathbf{u}_\alpha^V = \mathbf{u}_{\alpha'}^V$.

Proof 1 : Observe that the covariance $C_V(\mathbf{u}_\alpha^V, \mathbf{u}_\alpha^V)$ can be written as

$$
\begin{aligned}
C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V\right) &= \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v\right\|\right)\right] \\
&= \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i=1}^{N_v^V} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{i,\alpha}^v\right\|\right)\right] + \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i\neq j} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v\right\|\right)\right] \\
&= \frac{1}{N_v^V}\sigma_v^2 + \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i\neq j} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v\right\|\right)\right]
\end{aligned}
\tag{17}
$$

since by definition $\sum_{k=1}^{L} a_k = \sigma_v^2$. Similarly one has

$$
\begin{aligned}
C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V + r\boldsymbol{\varepsilon}_p\right) &= \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i=1}^{N_v^V}\sum_{j=1}^{N_v^V} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v - r\boldsymbol{\varepsilon}_p\right\|\right)\right] \\
&= \frac{1}{N_v^V}\sum_{k=1}^{L} a_k \rho_k(r) + \left(\frac{1}{N_v^V}\right)^2 \sum_{k=1}^{L} a_k \left[\sum_{i\neq j} \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v - r\boldsymbol{\varepsilon}_p\right\|\right)\right]
\end{aligned}
\tag{18}
$$

since $\left\|\boldsymbol{\varepsilon}_p\right\| = 1$. Combining equations (17) and (18) yields

$$
\lim_{r\to 0} \frac{C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V + r\boldsymbol{\varepsilon}_p\right) - C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V\right)}{r} = \mathbf{I}_1 + \mathbf{I}_2
$$

where

$$
\mathbf{I}_1 = \sum_{k=1}^{L} a_k \left(\lim_{r\to 0} \frac{\rho_k(r) - 1}{r N_v^V}\right) \text{ and } \mathbf{I}_2 = \sum_{k=1}^{L} a_k \sum_{i\neq j}^{N_v^V} \left(\lim_{r\to 0} \frac{\rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v - r\boldsymbol{\varepsilon}_p\right\|\right) - \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v\right\|\right)}{r\left(N_v^V\right)^2}\right)
$$

Since by assumption $\rho_k^{'}(0)$ is finite for any k, we get $\mathbf{I}_1 = \frac{1}{N_v^V}\sum_{k=1}^{L} a_k \rho_k^{'}(0)$ and $\mathbf{I}_2 \propto \frac{1}{N_v^V}$. It follows that for large values of N_v^V, $\mathbf{I}_1 = \mathbf{I}_2 = 0$ for which completes the proof of the Lemma.

Note that the assumption (16) is classical in practice as most of the covariance models (spherical, exponential, Gaussian etc…) fulfill this property.

The next result investigates the case of finite values of N_v^V which occur when the finer scale has a relatively large volume. In such a case and based on the previous lemma, we may expect the covariance to be twice differentiable with respect to any direction. We have the following lemma.

Lemma 2 Assume that N_v^V is finite. In addition assume that the auto-covariance structure $C_v(\mathbf{h})$ of the fine scale process satisfy

$$
\sum_{k=1}^{L} a_k \left.\frac{d\rho_k(r)}{dr}\right|_{r=0_+} = 0 \tag{19}
$$

Then the auto-covariance C_V is twice differentiable with respect to any direction at the diagonal $\mathbf{u}_\alpha^V = \mathbf{u}_{\alpha'}^V$ when $N_v^V \to \infty$.

Proof 2 : Using the same technique as in the proof of Lemma 1, on has

$$
\lim_{r\to 0} \frac{C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V + r\boldsymbol{\varepsilon}_p\right) - C_V\left(\mathbf{u}_\alpha^V,\mathbf{u}_\alpha^V\right)}{r} = \mathbf{I}_1 + \mathbf{I}_2
$$

where $\mathbf{I}_1 = \sum_{k=1}^{L} a_k \left(\lim_{r\to 0} \frac{\rho_k(r) - 1}{r N_v^V}\right)$ and $\mathbf{I}_2 = \sum_{k=1}^{L} a_k \sum_{i\neq j}^{N_v^V} \left(\lim_{r\to 0} \frac{\rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v - r\boldsymbol{\varepsilon}_p\right\|\right) - \rho_k\left(\left\|\mathbf{u}_{i,\alpha}^v - \mathbf{u}_{j,\alpha}^v\right\|\right)}{r\left(N_v^V\right)^2}\right)$

The proof follows from (19). □

Numerical examples

In this section, we evaluate the accuracy of the presented theories using simulated data. We consider a two-dimension example and simulate over a domain $D=[0,\ 100]\times[0,\ 100]\subset\mathbb{R}^2$ a zero mean second order stationary Gaussian random field $\{Z_\bullet(\mathbf{s}),\ \mathbf{s}\in D\}$ with non differentiable covariance function at the origin given by

$$c_\bullet(\mathbf{h})=0.3\ \delta(\mathbf{h})+0.7\ \mathrm{Exp}\left(\frac{\|\mathbf{h}\|}{12}\right) \tag{20}$$

with range $r_\bullet\approx 12$.

Then we consider a scale of interest v with volume blocks $[0,5]\times[0,5]$ where spatial analyses are requested. Then we define the up-scaled process Z_v by

$$Z_v(v_\alpha)=\frac{1}{N_\bullet^v}\sum_{i=1}^{N_\bullet^v}Z_\bullet(\mathbf{s}_i\in v_\alpha) \tag{21}$$

where the discretization number $N_\bullet^v=12\times 10=120$ is freely chosen large for accuracy. The number of simulated blocks at the scale v is clearly equal to $n=400$ which is sufficient for consistent statistics. The resulting standardized covariance model $c_v(\mathbf{h})$ that fits the sample covariance at the scale v is as

$$c_v(\mathbf{h})=0.0\ \delta(\mathbf{h})+1.0\mathrm{Gaus}\left(\frac{\|\mathbf{h}\|}{18}\right) \tag{22}$$

with range $r_\bullet\approx 20.75$. As proven in Lemma 1, the smoothness property of the covariance model of the large scale process defined in Equation (21) is governed by $\frac{1}{N_v^V}\sum_{k=1}^{L}a_k\rho_k^{'}(0)\propto\frac{1}{N_v^V}\sum_{k=1}^{L}\xi_k^{-1}\approx 0.002$.

This explains the smoothness property of $c_v(\mathbf{h})$ as the nugget $\xi_v=\frac{\xi_\bullet}{120}=0.0025$ is negligible.

Discussions and Conclusion

The geostatistics scaling laws are revisited and important notions such as the nugget effect component and the smoothness property of the underlying auto-covariance function are discussed. It is proved that when the number of discretization N_v^v used to define the regularized process is large, the auto-covariance function of the coarse scale is at least twice differentiable. The correct modeling of the spatial structure of the coarse scale process can be used in a multiscale data framework where the positive definiteness property of the coregionalization matrix can be obtained using the approach described by Elogne and Leuangthong (2009).

References

Abrahamsen A (2003) Review of Gaussian Fields and Correlation Functions. Norwegian Computing Center, Oslo, Tech. Rep. 917.

Christakos G (1992) Random field models in earth sciences: Academic Press, San Diego, 474 p.

Cressie NAC (1993) Statistics for Spatial data: Wiley, New York.

Efendiev Y, Datta-Gupta A, Osako I, Mallick B (2005) Multiscale data integration using coarse scale models. Advances in Water Resources (28), 303-314.

Elogne SN, Leuangthong O (2009) On a generalized linear model of coregionalization. Submitted to Mathematical Geosciences.

Frykman P, Deutsch CV (1999) Geostatistical scaling laws applied to core and log data. SPE 56822 paper presented at the SPE Annual Technical Conference and Exhibition, Houston, TX.

Frykman P, Deutsch CV (2002) Practical application of geostatistical scaling laws for data integration. Petrophysics 43:153-171.

Journel AG, Huijbregts J, (1978) Mining Geostatiscis, London: Academic Press, 600 p.

Journel AG (1999) Markov models for cross covariances. Math Geol, 31(8): 955-964.

Kupfersberger H, Deutsch CV, Journel AG (1998) Deriving constraints on small scale variograms due to variograms of large scale data. Mathematical Geology, vol. 38, No 7, p. 801-819.