INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

SMR/402 - 24

COLLEGE ON SOIL PHYSICS
9 - 27 October 1989

"A Glossary of Terms in Spatial Variability "

M. VAUCLIN
Institut de Mécanique de Grenoble
C.N.R.S.
Saint Martin d'Hères
France

Please note: These are preliminary notes intended for internal distribution only.

A GLOSSARY OF TERMS IN SPATIAL VARIABILITY

Michel VAUCLIN

Random function or stochastic process $Z(\underset{\sim}{x})$:

For each fixed $\underset{\sim}{x}_i$ (e.g. a location in space) $Z(\underset{\sim}{x})$ is a random variable.

Probability density of $Z(\underset{\sim}{x})$: The function $f(z:\underset{\sim}{x})$ such that the probability $Z(\underset{\sim}{x})$ is, between a and b given by

$$P(a \leq Z(\underset{\sim}{x}) \leq b) = \int_a^b f(z:\underset{\sim}{x})\,dv$$

Expected value or mean of $Z(\underset{\sim}{x})$: The expected value ($E\{Z(\underset{\sim}{x})\}$) is the first moment of the probability density function:

$$E\{Z(\underset{\sim}{x})\} = \int_{-\infty}^{\infty} z\, f(z:\underset{\sim}{x})\,dz = m(\underset{\sim}{x})$$

Variance of $Z(\underset{\sim}{x})$:

$$\mathrm{Var}\{Z(\underset{\sim}{x})\} = E\left\{\left(Z(\underset{\sim}{x}) - E\{Z(\underset{\sim}{x})\}\right)^2\right\}$$

Covariance function for $Z(\underset{\sim}{x})$: it is defined by

$$\mathrm{Cov}(Z(\underset{\sim}{x}_1); Z(\underset{\sim}{x}_2)) = E\left\{(Z(\underset{\sim}{x}_1) - m(\underset{\sim}{x}_1))(Z(\underset{\sim}{x}_2) - m(\underset{\sim}{x}_2\right.$$

it measures the statistical relationship between field values at two locations $\underset{\sim}{x}_1$, $\underset{\sim}{x}_2$

Variogram function for $Z(\underset{\sim}{x})$: it is defined by:

$$2\gamma(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = \mathrm{Var}\{Z(\underset{\sim}{x}_1) - Z(\underset{\sim}{x}_2)\}$$

the function $\gamma(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ is then the "semi-variogram"

Statistical homogeneity: $Z(\underset{\sim}{x})$ is statistically homogeneous or second-order stationary if:

i) $E\{Z(\underset{\sim}{x})\} = m$, a constant

and ii) $\mathrm{Cov}(Z(\underset{\sim}{x}_1), Z(\underset{\sim}{x}_2)) \equiv C(\underset{\sim}{x}_1 - \underset{\sim}{x}_2)$ only depends on the separation (lag) vector $\underset{\sim}{x}_1 - \underset{\sim}{x}_2 = \underset{\sim}{h}$.

This leads to the following relations:

- $\mathrm{Var}\{Z(\underset{\sim}{x})\} = C(\underset{\sim}{0})$
- $2\gamma(\underset{\sim}{h}) = C(\underset{\sim}{0}) - C(\underset{\sim}{h})$

statistically isotropic process: A statistically homogeneous process where $C(\underset{\sim}{h}) = C(|\underset{\sim}{h}|) = C(h)$ only depends on the separation distance $h = |\underset{\sim}{h}|$. #2

Intrinsic random function: $Z(\underset{\sim}{x})$ is said to be intrinsic when:

i) $E\{Z(\underset{\sim}{x})\} = m$, a constant

and ii) $Var\{Z(\underset{\sim}{x} + \underset{\sim}{h}) - Z(\underset{\sim}{x})\} = 2\gamma(\underset{\sim}{h})$ only depends on lag $\underset{\sim}{h}$.

Ergodicity: it implies that the unique realization of $Z(\underset{\sim}{x})$ behaves in space with the same probability density function as the ensemble of all possible realizations.

Scale: An average distance over which points are significantly correlated.

sill: the limit reached by $\gamma(\underset{\sim}{h})$ as $|\underset{\sim}{h}|$ increases

range: the scale in a variogram for a statistically homogeneous process.

Unbiasedness: An unbiased estimator $Z^*(\underset{\sim}{x}_0)$ of $Z(\underset{\sim}{x}_0)$ at $\underset{\sim}{x}_0$ is such that:

$$Z^*(\underset{\sim}{x}_0) = E\{Z(\underset{\sim}{x}_0)\}$$

optimality condition: the estimation error of $Z(\underset{\sim}{x}_0)$ by $Z^*(\underset{\sim}{x}_0)$ is minimum:

$$E\{(Z^*(\underset{\sim}{x}_0) - Z(\underset{\sim}{x}_0))^2\} \text{ minimum}$$

linear estimator: An estimator of the form:

$$Z^*(\underset{\sim}{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\underset{\sim}{x}_i)$$

where the weights λ_i are constants.

kriging. The procedure that gives the Best (minimum variance) Linear Unbiased Estimator of $Z(\underset{\sim}{x}_0)$ based on n observations.

For an intrinsic random function, this yields a set of linear equations for the weights:

$$\sum_{j=1}^{n} \lambda_j \gamma(\underset{\sim}{x}_i - \underset{\sim}{x}_j) + \mu = \gamma(\underset{\sim}{x}_i - \underset{\sim}{x}_0), \quad i = 1 \dots n$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

#5

<u>cross-validation</u> (jacknifing): procedure to check the validity of all the assumptions used in kriging (stationarity, good estimate of variogram, ...): all the data points are successively kriged by using all other ($n-1$) values. The resulting test is:

$$\frac{1}{n}\sum_{i=1}^{n}\left(Z(\underset{\sim}{x}_i)-Z^*(\underset{\sim}{x}_i)\right)\simeq 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Z(\underset{\sim}{x}_i)-Z^*(\underset{\sim}{x}_i)}{\sigma_K(\underset{\sim}{x}_i)}\right)^2\simeq 1$$

<u>drift</u>: non stationary expectation of $Z(\underset{\sim}{x})$:

$$E\{Z(\underset{\sim}{x})\}=m(\underset{\sim}{x})$$

The random function can then be expressed as:

$$Z(\underset{\sim}{x})=m(\underset{\sim}{x})+Y(\underset{\sim}{x})$$

where $Y(\underset{\sim}{x})$ is the residual term ($E\{Y(\underset{\sim}{x})\}=0$) which may or may not be stationary.

<u>anisotropy</u>: a phenomenon is said to be "anisotropic" when its variability is not the same in every direction. The structural function depends on the direction parameters in addition to $|\underset{\sim}{h}|$.

<u>kriging estimation variance</u>: the variance, σ_K^2, associated with the kriging estimator at $\underset{\sim}{x}_0$

$$\sigma_K^2(\underset{\sim}{x}_0)=\sum_{i=1}^{n}\lambda_i\gamma(\underset{\sim}{x}_i-\underset{\sim}{x}_0)+\mu\ .$$

<u>kriging of mean values</u>: instead of estimating the value $Z^*(\underset{\sim}{x}_0)$ at a point $\underset{\sim}{x}_0$ (ponctual kriging), it is possible to estimate directly any linear combination of such estimates, in particular, the average value over a given area S. The kriging system is as follows:

$$\sum_{j=1}^{n}\lambda_j\gamma(\underset{\sim}{x}_i-\underset{\sim}{x}_j)+\mu=\bar{\gamma}(\underset{\sim}{x}_i,S)\ ,\ i=1\ldots n$$

$$\sum_{j=1}^{n}\lambda_j=1$$

where $\bar{\gamma}(\underset{\sim}{x}_i,S)=\frac{1}{S}\int_S\gamma(\underset{\sim}{x}_i-\underset{\sim}{x})\,d\underset{\sim}{x}$

The associated estimation variance is given by:

$$\sigma_K^2(\bar{Z}^*)=\sum_{i=1}^{n}\lambda_i\bar{\gamma}(\underset{\sim}{x}_i,S)+\mu-\bar{\gamma}(S,S)$$

with $\bar{\gamma}(S,S)=\frac{1}{S^2}\iint_S\gamma(\underset{\sim}{x}-\underset{\sim}{x}')\,dx\,dx'$.

cross.variogram: it measures the spatial correlation between two random functions $Z_1(\underset{\sim}{x})$ and $Z_2(\underset{\sim}{x})$. Under the hypothesis of second-order stationarity, it is defined by:

$$2\gamma_{12}(\underset{\sim}{h}) = E\{(Z_1(\underset{\sim}{x}+\underset{\sim}{h})-Z_1(\underset{\sim}{x}))(Z_2(\underset{\sim}{x}+\underset{\sim}{h})-Z_2(\underset{\sim}{x}))\}$$

cross.covariance: Under the second-order stationarity assumption, it is defined by:

$$C_{12}(\underset{\sim}{h}) = E\{Z_1(\underset{\sim}{x}+\underset{\sim}{h}) \cdot Z_2(\underset{\sim}{x})\} - m_1 \cdot m_2$$

It is related to the cross.variogram by the expression:

$$2\gamma_{12}(\underset{\sim}{h}) = 2C_{12}(\underset{\sim}{0}) - C_{12}(\underset{\sim}{h}) - C_{21}(\underset{\sim}{h})$$

co-kriging: the extension of kriging to the case where $Z_1(\underset{\sim}{x})$ is estimated by using observations from two random functions $Z_1(\underset{\sim}{x})$ and $Z_2(\underset{\sim}{x})$.

autocorrelation function: it is defined by:

$$r(\underset{\sim}{h}) = \frac{\mathrm{Cov}(Z(\underset{\sim}{x}+\underset{\sim}{h}), Z(\underset{\sim}{x}))}{C(0)}$$

the plot of $r(\underset{\sim}{h})$ is called correlogram.

cross.correlation function: it is defined by:

$$r_c(\underset{\sim}{h}) = \frac{C_{12}(\underset{\sim}{h})}{[C_{11}(0)\, C_{22}(0)]^{1/2}}$$

spectral density function: it may be defined as the Fourier transform of the autocorrelation function:

$$S(f) = 2\int_0^\infty r(h)\cos(2\pi f h)\,dh .$$

where f is the frequency (cycle/m).