Construction of valid covariances along a hydrographic network. Application to specific water discharge on the Moselle Basin

C. Bernard-Michel[1], C. de Fouquet[1]
1 Centre de Géostatistique, Ecole des Mines de Paris, France
Corresponding author: chantal.de_fouquet@ensmp.fr

ABSTRACT: Estimating concentrations, flow rates or discharges along a stream network requires specific models of random functions because usual covariance models are no longer valid on such structures. Moreover, variables are generally highly non stationary because of the discontinuity at each confluence but also because of relationships with soil properties, that have to be taken into account.

We propose a global model of random functions along a tree graph introducing the concept of "elementary thin streams", defined by the whole set of paths between sources and outlet [de Fouquet and Bernard-Michel., 2006]. At each point of the network, the river is considered to be the linear combination of these streams on which one dimensional stationary random functions are defined. In practise, the coefficients of the linear combination are determined according to the conservation equation of discharge and flux at the forks.

An application to water discharge on the Moselle Basin (north-east of France) is presented. The hydrographic network of this Basin is made of about 100 important nodes and 100 monitoring stations are available for the last 10 years. Correlations with auxiliary information are examined, and theoretical models of covariances and variograms are presented.

KEYWORDS :hydrographic network, water discharge, geostatistics, covariance, streams.

1. Non stationarity of variables

1.1. Experimental results

The experimental study of water discharge on the Moselle Basin gives evidence of its non stationarity. The annual water discharge increases from the source to the outlet and experimental variograms (Fig. 1. left) show obvious non stationarity, according to the selected distance [Bernard-Michel, 2005 and 2006].

The non stationarity is due to the conservation of water discharge at confluences. If D_3 is the water discharge at a confluence, and D_1 and D_2 the water discharges immediately upstream on the two reaches forming the fork, then $D_3 = D_1 + D_2$, which involves an increase of the mean and of the variance of the associated random function on the network, asuming independence of water discharge on parallel rivers.

$$E(D_3) = E(D_1) + E(D_2) \text{ and } Var(D_3) = Var(D_1) + Var(D_2) \quad (1)$$

The non stationarity can also be "explained" by the strong correlation with the drainage basin surface (Fig. 1. right) that can be calculated locally by a SIG.

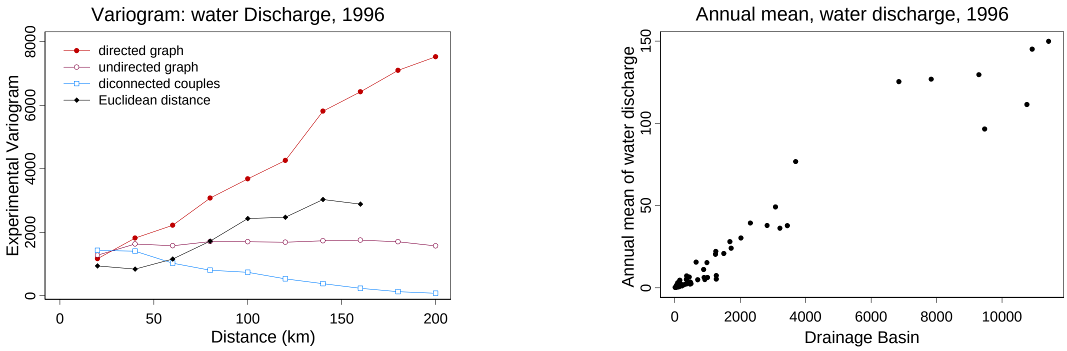

Fig. 1. On the left, experimental variograms of annual water discharge in 1996, calculated for different distances. On the right: annual mean of water discharge versus drainage basin surface in 1996.

1.2. Specific water discharge

In order to work with "more" stationary functions, the specific water discharge can be studied [Sauquet, 2000]. It is defined as the annual water discharge D divided by the drainage basin surface S: $D_S = \frac{D}{S}$. But as with water discharge, this variable is not stationary of order 2: at each confluence, the additive relation on water discharge $D_3 = D_1 + D_2$ implies the following relation for specific water discharge $D_{S3} = \frac{S_1 D_{S1} + S_2 D_{S2}}{S_1 + S_2}$. The mean on the whole network is then constant but the variance is still non stationary.

Generally non stationarity can be reduced by taking into account auxiliary information, either by a division (specific discharge) or by a regression. But in all cases, non stationarity will remain because of the conservation of flux and discharge at confluences. Maintaining stationarity by weighting random functions is then unconceivable.

2. Construction of valid models

2.1. Random functions on a hydrographic network

Usual geostatistical models as the spherical covariance, developed for Euclidean space, are not valid anymore on tree graphs. Generalizing recent models [Ver Hoef, 2006], we present a construction combining any model of one-dimension Random Functions defined on each path between sources and the outlet (Fig. 2., left). The principle is as follows: we consider the one-dimension random functions Y_i defined on each path linking one source to the outlet; when different paths join at a node, the resulting random function downstream is a linear combination of the corresponding Y_i using their respective weights. For example, in Fig. 2. (right), the specific discharge $T(x)$ is a linear combination of Y_1, Y_2 and Y_3. The weights depend on the surfaces S_1^1, S_1^2, S_2^2 and S_2^3 (equation (3)).

Let be

- ✓ R_G the hydrographic network
- ✓ $I(x)$: the set of elementary thin streams containing x
- ✓ $J(i,x)$: the set of confluences upstream x, occurred on the elementary thin stream *i*.
- ✓ Y_i: the random function defined on the elementary thin stream *i*. The functions Y_i are assumed to be stationary, independent, with an equal mean $E(Y_i(x)) = m \; \forall x \in R_G$, an equal variance $Var(Y_i(x)) = \sigma^2 \; \forall x \in R_G$ and a covariance $C_1(h)$.
- ✓ S_j^i: Drainage basin surface upstream confluence *j*, on the reach containing the elementary stream *i*.
- ✓ S_j: Drainage basin surface just downstream confluence *j*.

- ✓ $S_j \setminus S_j^i$: Drainage basin surface upstream confluence j, on the reach that doesn't contain the elementary stream *i*.

$w_j^i = \frac{S_j^i}{S_j}$: Weight assigned to the elementary stream i immediately downstream confluence *j*.

$W(i,x)$: Weight assigned to the elementary stream *i* at the position *x*.

Then, specific water discharge $T(x)$ at position x is equal to:

$$T(x) = \sum_{i \in I(x)} W(i,x) Y_i(x) \text{ avec } W(i,x) = \begin{cases} \prod_{j \in J(i,x)} w_j^i \\ 1 \text{ si } J(i,x) = \{\varnothing\} \end{cases} \quad (2)$$

In the example presented on Fig. 2. (right) :

$$T(x) = \frac{S_2^1 S_1^1}{S_2 S_1} Y_1(x) + \frac{S_2^2 S_1^2}{S_2 S_1} Y_2(x) + \frac{S_2^3}{S_2} Y_3(x) \quad (3)$$

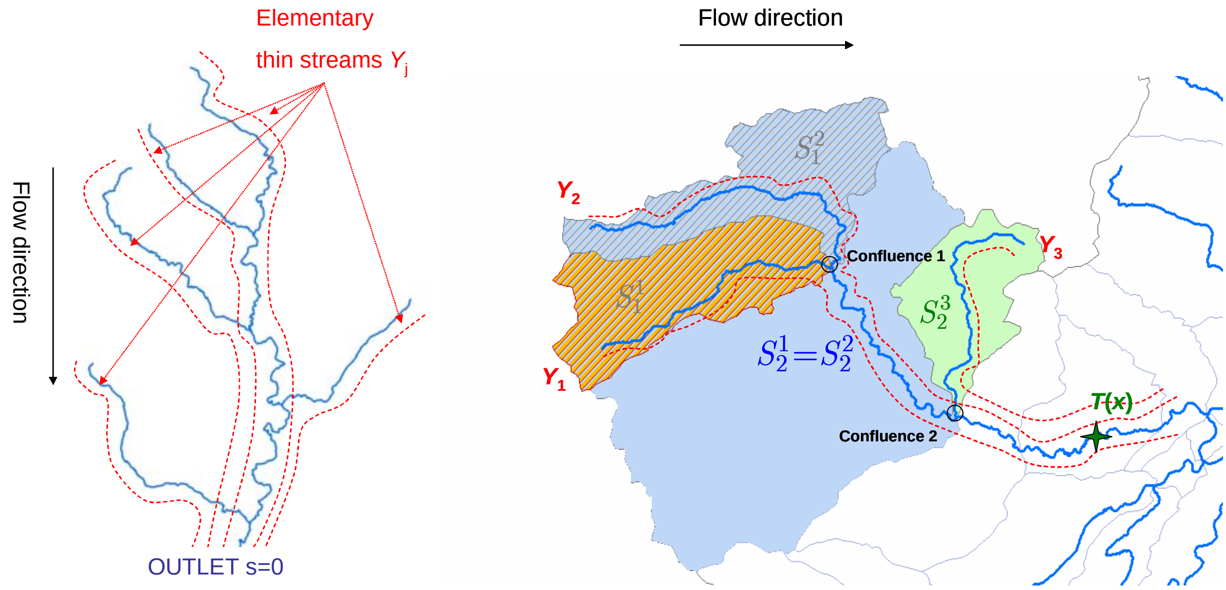

Fig. 2. Left: Construction of elementary thin streams. Right: Extract of the Moselle basin.

2.2. Covariance and variogram

The statistical properties, mean variance, covariance and variogram can then be deduced from equation (2).

<u>Mean and variance</u>

$$E(T(x)) = m \text{ and } Var(T(x)) = \sigma^2 \times \sum_{i \in I(x)} W(i,x)^2$$

<u>Covariance and variogram</u>

$$C(x,y) = \begin{cases} 0 \text{ if x and y are not flow connected} \\ C_1(0) \sum_{i \in I(x)} W(i,x)^2 \text{ si } x = y \\ C_1(|x-y|) \sum_{i \in I(x) \bigcap I(y)} W(i,x) W(i,y) \text{ if } x \text{ and } y \text{ are flw connected} \end{cases}$$

$$Var(T(y)-T(x)) = \begin{cases} 0 \text{ if } x=y \\ \sigma^2\left(\sum_{i\in I(y)} W(i,y)^2 + \sum_{i\in I(x)} W(i,x)^2\right) \text{ if } x \text{ and } y \text{ are not flow connected} \\ 2\gamma_1(y-x)\left[K(x,y)\sum_{i\in I(x)} W(i,x)^2\right] \\ \quad + \sigma^2\left[\sum_{i\in I(y)\setminus I(x)} W(i,y)^2 + (K(x,y)-1)^2 \sum_{i\in I(x)} W(i,x)^2\right] \\ \qquad \text{if } x \text{ and } y \text{ are flow connected, } y \text{ downstream } x \end{cases}$$

Before any model can be inferred, the validity of the underlying hypotheses must be established. But there are two major difficulties [Bernard-Michel, 2006]: first, there are very few data measurements (less than one per reach) and secondly the assumptions do not relate to the specific discharge itself, but to its value on an elementary stream, for which no observation is available except near the sources. The results show that hypotheses on independence and stationarity are sometimes verified depending on the year, but the inference of the model is not advised because the measurements are too few and much more auxiliary information is required.

3. Conclusion

This paper must be viewed as a first step in the application of geostatistics to river pollution problematics on a hydrographic network. It gives a general and easy construction of valid theoretical covariance and variogram models on hydrographic networks. The hypotheses underlying the model can be verified, but its inference remains difficult because of the few measurements available. In the future, measurements frequency has to be increased, and the model should be coupled with phenomenological models.

Acknowledgments : The authors would like to thank the French Ministry of environment and the Agences de l'eau Loire Bretagne and Rhin Meuse.

REFERENCES

Bernard-Michel C., de Fouquet, C. (2005). Geostatistical indicators of nutrients concentrations in streams. Proceedings of IAMG'05: GIS and Spatial analysis, (2):716-721. Quiming Cheng, Graeme Bonham Carter (eds).

Bernard-Michel C. (2006). *Indicateurs géostatistiques de la pollution dans les cours d'eau*. Centre de géosciences. Ecole des Mines de Paris. Thèse de doctorat.

de Fouquet C., Bernard-Michel C.(2006). Modèles géostatistiques de concentrations ou de débits le long des cours d'eau. *Comptes Rendus Géosciences* (338) 307-318.

Sauquet, E. (2000). *Une cartographie des écoulements annuels et mensuels d'un grand bassin versant structurée par la topologie du réseau hydrographique*. Institut national polytechnique de Grenoble, Unité de Recherche Hydrologie-Hydraulique, Cemagref. Thèse

Ver Hoef J. M., Peterson E., Theobald D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics (in Press)*