

# Geostatistical indicators of nutrients concentrations in streams

C. Bernard-Michel<sup>1</sup>, C. de Fouquet<sup>1</sup>

<sup>1</sup>Ecole des Mines de Paris, centre de Géostatistique, 35 rue Saint Honoré, 77305  
Fontainebleau, France,  
E-mail: caroline.bernard-michel@ensmp.fr

## 1. Introduction

In order to assess water river quality, nutrients concentrations are measured and summarized by synthetic indicators such as the annual mean of concentrations or the 90% quantile per monitoring station. Today, these indicators are estimated using classical statistical inference, essentially based on hypotheses proved incorrect for many parameters: time correlation and systematic seasonal variations are neglected. This can have important consequences, in particular when sampling is preferential. For example, in France, nitrate concentrations are higher during winter because of the runoff and the vegetation cycle. If measurements are reinforced in winter, indicators are then falsely increased.

First, we show how to take into account time correlation using geostatistics, assigning kriging weights to measurements: yearly mean and quantiles of concentrations are better assessed.

Secondly, concentrations are generally linked to the land cover. These links are quantified using geographical information.

Finally, we examine spatial correlation of correlation and discharge along a network. Experimental variograms for discharges and nitrate concentrations are presented for three different distances: the usual 2D Euclidean distance and two distances along the hydrographic network.

The study focuses on the Rhin Meuse basin. A GIS was used to take into account the hydrographic network structure, land use information and to calculate streams distances, flows, drainage basins and upstream land characteristics.

## 2. Time correlation: how to improve current concentration indicators?

To describe nutrients concentrations, two indicators are commonly used by environmental agencies: the annual value of concentrations for micro pollutants and the 90% quantile for nutrients, calculated for each monitoring station. But, the statistical calculations rely on the hypothesis of independence between sampled values, which contradicts experimental variograms (Bernard-Michel et al., 2004). As a consequence, statistical indicators can be biased. For example, nitrate and nitrite measurements at the monitoring station 56000 on the Loire Bretagne River in 1995 are presented Fig. 1A. From November to April, sampling frequency has been doubled in order to control high values of nitrate concentrations. During this period, nitrite concentrations are low. Consequently, the annual mean is overestimated for nitrates and underestimated for nitrite when using usual statistics. Moreover, the confidence interval announced for the estimation is wrong because time correlations are not taken into account when calculating

the error variance. It is therefore necessary to take into account both time correlations and sampling dates, especially in case of preferential sampling, which can be done by kriging the annual average of concentrations (Chilès, 1999). In geostatistics, concentration is indeed interpreted as a realization  $z(t)$  of a correlated random function. Moreover, the quantity to estimate is not anymore the mean parameter of a distribution, but the temporal average  $Z_T = \frac{1}{T} \int_T Z(t) dt$ , still defined even if concentrations are not stationary during the year. This quantity is estimated using ordinary “block” kriging, with constant but unknown mean.

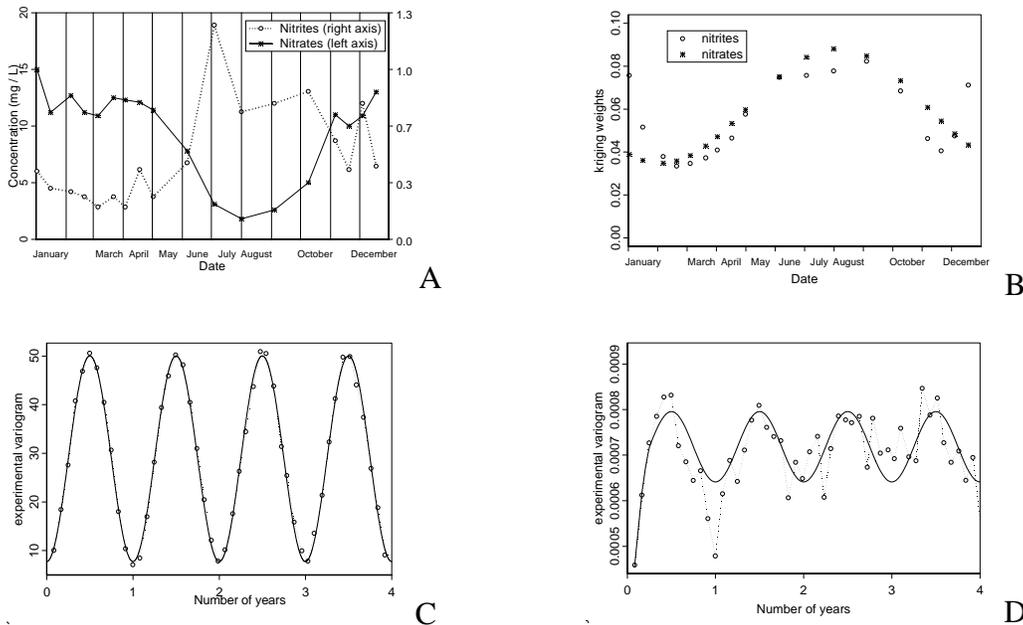


Fig. 1. A. Graph of nitrites and nitrate concentrations in 1995 station 56000 versus dates. B. kriging weights for nitrates and nitrites concentrations versus dates. C. Variogram of nitrate concentrations over 4 years versus years (lag = 30 days) and fitted model. D. Variogram of nitrite concentrations over 4 years versus years (lag = 30 days) and fitted model.

For the station 56000, variograms of nitrates and nitrites concentrations calculated over 10 years are presented Fig. 1C and 1D. They are fitted using a combination of a nugget effect, a spherical and a cosine model. Annual periodicity is predominant for nitrate concentrations. Fig. 1B presents kriging weights for the year 1995, lower when sampling dates are closer. Statistical and geostatistical results are compared in table 1. The bias induced by preferential sampling is then corrected by kriging. For both pollutants, the difference between the two estimations reaches 10%. The error standard deviation and the confidence interval, overestimated by classical statistics, are 25% reduced by kriging for nitrates. On the opposite, it is increased of 30% for nitrites.

The geometrical method of segments of influence can be use as a simplification of kriging. Theses methods are compared in (Bernard-Michel et al., 2004) for both annual mean and quantile estimations.

	Estimation of the annual mean		Error standard deviation	
	Statistics	Geostatistics	Statistics	Geostatistics
nitrate	9.69	8.59	0.9261	0.6893
nitrite	0.04927	0.05467	0.00691	0.00908

Table 1: Annual mean of nitrate and nitrites concentrations estimated by statistics and geostatistics. Station 56000 in 1995.

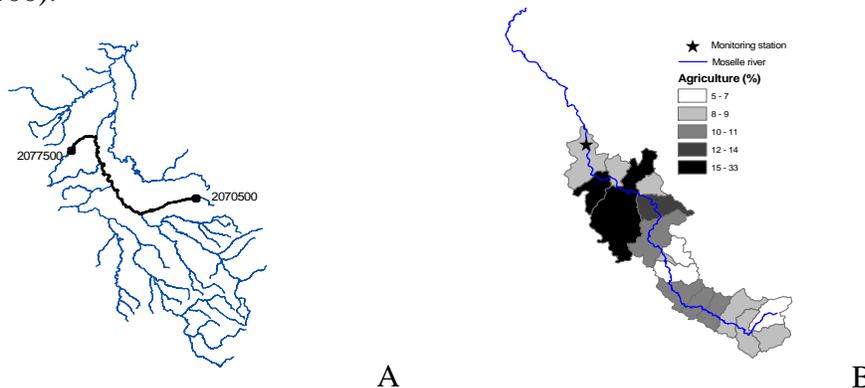
### 3. Relation between land use and concentrations

With a view to multivariate modeling, we examine now relations between concentrations and environment on the whole Rhin Meuse basin (more than 200 stations). Annual values of nitrate concentrations are plotted versus land use. Two possibilities were explored: the percentage of land use type (for example intensive agriculture) is calculated on the local drainage basin of each considered station, or on the cumulated drainage basin from the station to the source (Fig. 2B). This second choice let appear better correlation between concentrations and land use.

The correlation between nitrate concentrations and the percentage of intensive agriculture on the cumulated drainage basin is presented Fig. 2C. Means of concentrations are calculated over 4 years (1997-2000) for all the stations and presented by classes of 20 measurements. The experimental regression is quasi linear. The dispersion of the correlation scatter plot measured by the standard deviation in each class is higher in the last three stronger classes. Other correlations are presented in (de Fouquet, 2000). These types of relations have been exploited by (Cressie, 1997) to model nutrient concentrations.

### 4. Relation between discharge and drainage basin surface

To model nitrates flows on a hydrographic network, it is necessary to consider not only concentrations but also water discharges. Some authors have shown a relation of type  $D = \alpha S^\beta$  between the cumulated drainage basin surface  $S$  and the water discharge  $D$ . For the Rhin Meuse basin, the relation seems rather linear (Fig. 2D), with an increase of dispersion when the cumulated drainage surface increases. The relation between discharge and drainage basin shows that it could be interesting to work on specific discharge (discharge divided by drainage basin) in order to deal with non stationnarity (Sauquet, 2000).



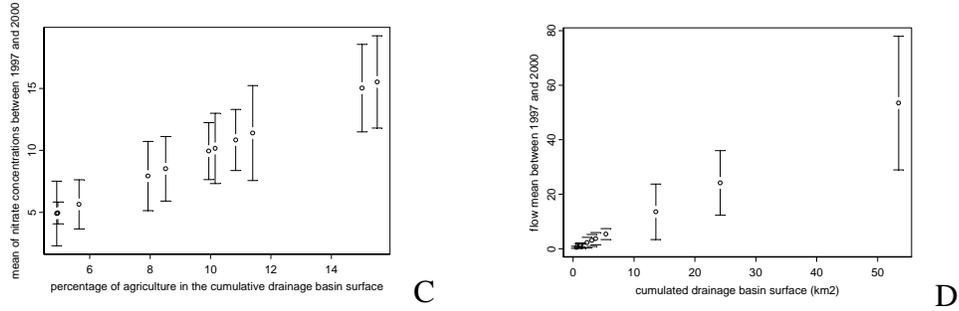


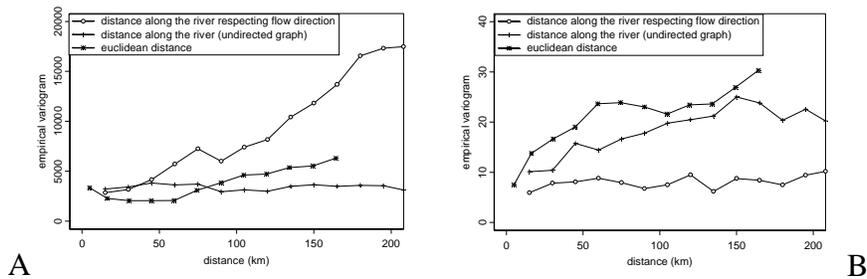
Fig. 2. A. Hydrographic network of the Moselle River. B. Cumulated drainage basin surface to the source. C. Mean of nitrates concentrations versus percentage of agriculture on the cumulated drainage basin surface and standard deviation. D. Mean of discharge grouped by classes versus cumulated drainage basin surface and standard deviation.

## 5. Spatial correlation along a hydrographic network

When modeling concentrations or discharges along a hydrographic network, the relevant distance to consider is not any more the Euclidean's one, but the “natural” distance along the hydrographic network. Moreover, usual covariances are not valid anymore on a tree support and appropriate models have been proposed (Monestiez et al., 2004; Ver Hoef et al., 2004; de Fouquet et al., 2005). Experimental variograms of annual values of water discharge and nitrate concentrations and flow along the hydrographic network in the Rhin Meuse basin (Fig. 3A, B, C) are calculated in 2001 on the whole network. Experimental variograms are presented according Eq. 3 for three different distances:

1. D1: Euclidean distance at 2D, all the couples of points being considered;
2. D2: Distance along the river according to discharge direction: only couples of points on the same stream line are considered (this correspond to a directed tree: the couple of stations 2077500 and 2070500 is not taken into account (Fig. 2A), ie the distance is infinite because there is no path between them)
3. D3: Distance along the river: all the couples of points are considered with the true distance respecting the river length and the network structure (non directed tree: the path between stations 2077500 and 2070500 is in grey in Fig. 3A)

$$\bar{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [Z(x_\alpha) - Z(x_\beta)]^2 \text{ with } N(h) = \{(\alpha, \beta), x_\alpha - x_\beta = h\} \quad (3)$$



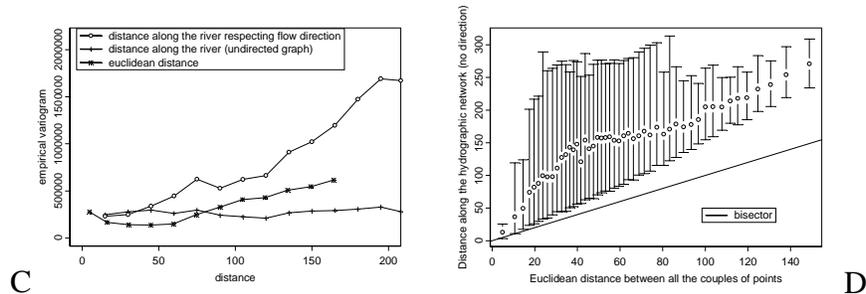


Fig. 3. On the Moselle's network for the 3 distances: A. Variograms of water discharge. B. Variograms of nitrates concentrations. C. Variograms of nitrates flows (nitrates concentrations multiplied by discharge). D. Comparison between distances calculated along the network (3) and Euclidean distances (1), means, 5% and 95% quantiles for groups of 100 couples. First bisector.

The experimental results are presented below:

Relations between distances: Fig. 3D confirms shows the evolution of distances D3 as a function of Euclidean distance: from 0 to 35 km, the relation is linear then a sill is observed until 100 km and is followed by an increase. The 5% quantiles plot determined for each class is linear whereas the 95% quantile is nearly unchanging between 20 and 80 km. It is important to observe that with Euclidean distance and distance D3, much more couples of points are taken into account then with distance D2. Moreover, when two points are spatially close but on different rivers next to the sources, the Euclidean distance is very short whereas distance D3 is important and distance D2 is infinite.

Nitrate flux: the nitrate flux is calculated as the product of water discharge with nitrate concentration. Because discharge increases from source to outlet whereas concentrations are rather constant, the variogram of nitrate flux (Fig. 3C) is analogous to the discharge's one (Fig. 3A).

Water discharge: as water discharge increases from the source to the outlet, variogram calculated with distance D2 shows a non stationnarity with a quadratic behavior, because only couples on the same stream line are taken into account (Fig. 3A). In particular, great distances correspond to couples with one point next to the outlet and one next to one source and so to a strong difference of discharge. For distance D1, the omnidirectional calculation reduces the non stationnarity. Some couples of points on different rivers corresponding to weak discharge differences are now considered and the variogram is beneath the one calculated with D2. For distance D3, two stations spatially close but next to different sources correspond to a high distance with a weak difference of discharge. For high distances, the variogram calculated with D3 is beneath others.

Nitrate concentrations: For concentrations, variograms are different. Concentrations do not obligatory increase along the river and the variation between rivers is more important than along rivers. In this case the difference between concentrations of two stations near two different sources can be important. The variogram calculated with D2 is then beneath the others, because it takes into account only couples downstream. The variogram

calculated with distance D1 is upper the one calculated with distance D3, because the same quadratic increase of concentrations corresponds to a smaller distance.

## 6. Conclusion

When time series are irregular and more particularly when sampling is preferential, it is essential to avoid any bias and to take into account time correlation and also sampling strategy. A correction by kriging is essential for the estimation of the annual mean of concentrations and even for water discharge and nitrate fluxes.

The modeling of concentrations in streams has been approached from different points of view: relations with the land use to take into account, model along the river. The experimental analyze of data is a fundamental part before the construction of a model.

## 7. References

Agence de l'eau Loire Bretagne, 2002: Système d'évaluation de la qualité de l'eau des cours d'eau. Rapport de présentation SEQ EAU. <http://www.rnde.tm.fr/>

Bernard-Michel C., de Fouquet C., 2004: Estimating indicators of river quality by geostatistics. geoENV V – Geostatistics for environmental application.

Chilès J-P., Delfiner P., 1999: Modeling spatial uncertainty. Wiley series in probability and statistics.

Cressie N., Majure J., 1997: Spatio-temporal statistical modeling of Livestock Waste in Streams. Journal of Agricultural, Biological, and Environmental Statistics, Volume 2, Number 1, Pages 24-47.

de Fouquet C., Bernard-Michel C., 2005: Modèles Géostatistiques de concentrations ou de débits le long des cours d'eau. Submitted for publication to the Comptes rendus de l'académie des sciences.

de Fouquet C., 2000: Construction d'un réseau représentatif de qualité des cours d'eau. Phase 1. Analyse exploratoire des données. Rapport technique. ENSMP. Fontainebleau.

Monestiez P., Bailly P., Lagacherie P., Voltz M., 2004: Geostatistical modelling of spatial processes on directed trees: Application to fluvisol extent. Submitted and accepted for publication in Geoderma.

Sauquet E., 2000: Une cartographie des écoulements annuels et mensuels d'un grand bassin versant structurée par la topologie du réseau hydrographique. Thèse INPG.

Ver Hoef J.M., Peterson E., Theobald D., 2004: Spatial statistical models that use flow and stream distance. Submitted and accepted for publication in Environmental and Ecological statistics.

Thanks to the French ministry of environment and to the water Agencies of Loire-Bretagne and Rhin Meuse.