2009

Spatial modelling and prediction on river networks: up model, down model or hybrid?

Vincent Garreta
CEREGE, UMR 6635, CNRS, Universit´e Aix-Marseille, Europˆole de l'Arbois, 13545 Aix-en-Provence, France

Pascal Monestiez
INRA, Unit´e de Biostatistique et Processus spatiaux, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France

Jay M. Ver Hoef
NOAA National Marine Mammal Lab, Alaska Fisheries Science Center

Spatial modelling and prediction on river networks: up model, down model or hybrid?

Vincent Garreta[1]*,[†], Pascal Monestiez[2] and Jay M. Ver Hoef[3]

[1]*CEREGE, UMR 6635, CNRS, Université Aix-Marseille, Europôle de l'Arbois, 13545 Aix-en-Provence, France*
[2]*INRA, Unité de Biostatistique et Processus spatiaux, Domaine St Paul, Site Agroparc, 84914 Avignon Cedex 9, France*
[3]*NOAA National Marine Mammal Lab, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA*

SUMMARY

Preservation of rivers and water resources is crucial in most environmental policies and many efforts are made to assess water quality. Environmental monitoring of large river networks are based on measurement stations. Compared to the total length of river networks, their number is often limited and there is a need to extend environmental variables that are measured locally to the whole river network. The objective of this paper is to propose several relevant geostatistical models for river modelling. These models use river distance and are based on two contrasting assumptions about dependency along a river network. Inference using maximum likelihood, model selection criterion and prediction by kriging are then developed. We illustrate our approach on two variables that differ by their distributional and spatial characteristics: summer water temperature and nitrate concentration. The data come from 141 to 187 monitoring stations in a network on a large river located in the Northeast of France that is more than 5000 km long and includes Meuse and Moselle basins. We first evaluated different spatial models and then gave prediction maps and error variance maps for the whole stream network. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: geostatistics; covariance function; stream network; kriging

1. INTRODUCTION

Preservation of aquatic biotopes and ecosystem management of river networks are major issues in public policies. For example, rivers and water resources may be subjected to intensive agriculture, causing pollution such as nitrates or pesticides. Environmental monitoring of large river networks is often based on measurement stations. Compared to the total length of river networks, the number of monitoring stations is often limited and there is a need to extend environmental variables that are measured locally to the whole river network.

---

*Correspondence to: Vincent Garreta, CEREGE, UMR 6635, CNRS, Université Aix-Marseille, Europôle de l'Arbois, 13545 Aix-en-Provence, France.
[†]E-mail: garreta@cerege.fr

Geostatistics can carry out this objective, and it seems natural to use a distance that is meaningful for a river, i.e. calculated along the network. This distance is called 'stream distance' in Ver Hoef *et al.* (2006) and 'river distance' in Cressie *et al.* (2006). The adoption of standard models and methods of geostatistics (e.g. Cressie, 1991; Chilés and Delfiner, 1999) cannot then be carried out in a simple way for several reasons. Firstly, stationarity (i.e. the invariance in distribution, mean and/or covariance by spatial translation) must be redefined according to the support topology of river networks and distances. Secondly, for any set of points belonging to a river network, it is necessary to have models that ensure positive-definiteness of the spatial covariance function. Finally, different spatial patterns may result from different variables and processes, so it is desirable to have enough valid spatial covariance models to fit the different processes.

Currently, two broad classes of models answer these conditions and were proposed independently. They involve different assumptions on the nature of the mechanisms that generate spatial variation in the data. Monestiez *et al.* (2005) proposed a first kind of geostatistical model based on conditional independence at branching points, while Ver Hoef *et al.* (2006) and Cressie *et al.* (2006) proposed a family of models with total independence between branches.

First, consider the models that are defined by conditional independence at branching points. Branches are modelled conditionally on what happens downstream. They are built from one-dimensional random functions (RFs) and easy to simulate using standard conditional simulation methods. Bailly *et al.* (2006) developed this kind of model to fit spatial variation of anthropogenic drainage networks. Their principal limitation is the absence of an explicit model of covariance for points that are not in an upstream–downstream relationship. The only way to calculate them is through simulation. Beyond the lack of a simple characterization of the whole covariance, kriging can become very difficult, even impossible, for important sets of data, except by removing most of the data for prediction (the same limitation appears in Cressie and Majure (1997)).

Second, consider the models that are defined by kernel convolution (Ver Hoef *et al.*, 2006) and are characterized by independence between branches. Built from a unilateral convolution kernel oriented upstream, they are not stationary by construction. Ver Hoef *et al.* (2006) provide a suite of models and showed how to add conditions on confluences to ensure stationarity in variance, and Cressie *et al.* (2006) used one set of confluence conditions and a particular model in a real application. De Fouquet and Bernard-Michel (2006) propose another point of view with another construction method of these models, extending them to the intrinsic hypothesis (variogram only exists).

In this paper, we show that the first models can be obtained in a similar way as the second by reversing the unilateral kernel downstream. Consequently, valid models of covariances can be given explicitly for all pairs of points, thus eliminating the main problem with those models. In the following, we call them *down* models in contrast to the models based on the upstream kernel which we call *up* models. This generalization and the resulting explicit expressions of spatial covariance allow us to implement classical methods in statistics, including maximum likelihood estimation, comparison and selection of models using Akaike information criterion (AIC) and kriging (interpolation at any network site).

These methods will be illustrated for a case study: the basin of the Meuse and the Moselle in the North-East of France. The dataset has two environmental variables sampled at 141 and 187 sites. These variables, summer water temperature and nitrate rates, have different characteristics. We model their spatial variation and map them everywhere on the river network by kriging.

2. MODELS

Considering the modelling of a RF on river networks, the main difficulty is the proper specification of the RF on a support that is both continuous, i.e. locally equivalent to a one-dimension line, and topologically a tree with branching points. The spatial covariance structure must be positive definite for any finite subset of points belonging to the network, whether they are located at or between branching points. Most spatial covariance models used in $\mathbb{R}$ or $\mathbb{R}^2$ cannot be simply generalized to branching supports without revisiting the stationarity hypothesis and what happens to the covariance function at branching points (see an example of what can go wrong in Ver Hoef *et al.* (2006)).

2.1. Definitions and notations for RF on river networks

We first need to introduce some notation to facilitate the writing of further expressions. Some of them are borrowed from Cressie *et al.* (2006). A site or a point on the river network—noted *s* below—is indexed on a continuous support whether it is located at a branching point or between branching points. A segment is the set of points between two branching points or confluences. A confluence is a node of order 3 or more on the river network. Referring to the stream flow, upstream endpoints are named sources and the downstream endpoint the outlet or the mouth of the stream network. To keep the network simple, particular cases such as lakes or deltas are not considered here but remain possible since the global topology remains a tree.

Reusing Cressie's notation, we can define at any point *s*, the upstream domain noted $\vee(s)$ and the downstream domain noted $\wedge(s)$. For two points *s* and *t*, a distance $\Delta = d(s, t)$ is defined as the stream distance along the river path if one of the two points belongs to the downstream domain of the other. We note $s \rightarrow t$ when *t* belongs to $\wedge(s)$ and the two points are said to be *flow-connected* (Figure 1(a)).

When two points are *flow-unconnected*, then neither *s* belongs to $\wedge(t)$ nor *t* belongs to $\wedge(s)$. In this case, let *u* be the upper common confluence in $\wedge(s) \cap \wedge(t)$. To be concrete, let *s* be at least as far, or farther, upstream than *t*, but on a different branch. Then the relationship between *s* and *t* is denoted $s \nrightarrow t$, and $s \rightarrow u$, $t \rightarrow u$ and $d(s, u) \geq d(t, u)$ (Figure 1(b)). Then the distance $d(s, t)$ depends on two quantities Δ_1 and Δ_2 where $\Delta_1 = d(s, u)$ and $\Delta_2 = d(t, u)$ with $\Delta_1 \geq \Delta_2$. The distance can be written as the sum $d(s, t) = \Delta_1 + \Delta_2$ but in most cases the covariance function will depend on both terms Δ_1 and Δ_2 and not only on the sum.
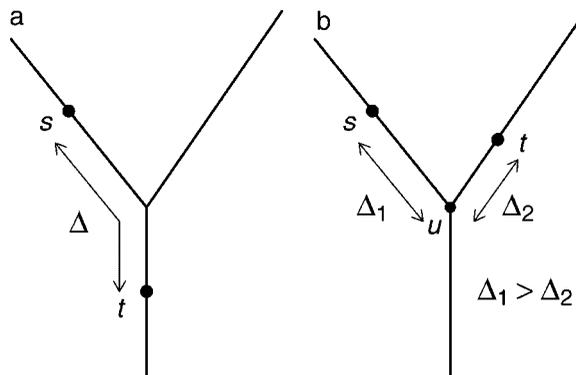

Figure 1. Notation for distance function at a confluent point in a river network

2.2. General RF building method

A general method for building a RF (1-D) or a random field (2-D or more) consists in smoothing a white noise by a kernel. The resulting RF is a Gaussian process with a spatial covariance function which is the self convolution of the smoothing kernel.

For example, in $\mathbb{R}^1$, a RF can be defined by

$$Y_s = \int_{\mathbb{R}} g(x - s) \mathrm{dB}(x)$$

where $\mathrm{dB}(x)$ is the derivative of a Brownian process, i.e. a white noise, and g is an integrable and square integrable kernel.

The covariance of $Y(\cdot)$ is thus

$$\mathrm{Cov}(Y_s, Y_t) = \int_{\mathbb{R}} g(x - s) g(x - t) \mathrm{d}x$$

For more details on the mathematical framework and direct applications, one can refer to Billingsley (1995), Barry and Ver Hoef (1996), Higdon (2001), Cressie and Pavlicova (2002) or Øksendal (2003).

2.3. Down models

In the general approach previously defined, we used a kernel defined for downstream points of s (Figure 2(a)). The integration is then performed on an unilateral 1-D kernel support:

$$Y_s = \mu(s) + \int_{\wedge_s} g(x - s) \mathrm{dB}(x)$$

where $\mu(s)$ is a mean term that may be assumed constant and $B(x)$ is a Brownian process starting from the outlet, progressing toward the sources, and separating themselves in independent Brownian processes at confluences.
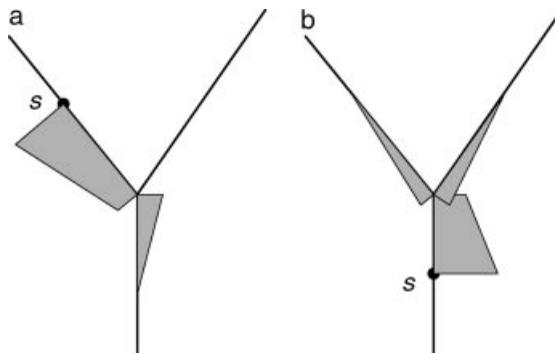

Figure 2.   Kernel shapes for (a) *down* model and (b) *up* model

It is then easy to deduce the spatial covariance structure; for all pair of points (Y_s, Y_t), with $\wedge_s$ and $\wedge_t$ their respective downstream domains, we express the covariance:

$$\text{cov}(Y_s, Y_t) = \int_{\wedge_s \cap \wedge_t} g(x-s)g(x-t)\mathrm{d}x$$

Considering the two different cases of Figure 1, the covariance becomes:

(i) If $s \to t$, then

$$\text{cov}(Y_s, Y_t) = \int_{\wedge_t} g(x)g(x-t+s)\mathrm{d}x \tag{1}$$

(ii) If $s \nrightarrow t$, then

$$\text{cov}(Y_s, Y_t) = \int_{\wedge_u} g(x-u+s)g(x-u+t)\mathrm{d}x = \int_{\wedge_u} g(x-\Delta_1)g(x-\Delta_2)\mathrm{d}x \tag{2}$$

2.3.1. Conditional independence between branches. In order to show the conditional independence property of *down* models let us define the conditional RF

$$Y_s|y_u = \int_{\wedge_s \backslash \wedge_u} g(x-s)\mathrm{d}B(x) + y_u \tag{3}$$

where $\wedge_s \backslash \wedge_u$ is the set $\wedge_s$ minus that of $\wedge_u$, i.e., the stream path between s and u, and where $y_u = \int_{\wedge_u} g(x-u)\mathrm{d}B(x)$ is the entirely known (considered as constant) RF dowstream of u, and note the probability as $P[Y_s|y_u] = P[\int_{\wedge_s \backslash \wedge_u} g(x-s)\mathrm{d}B(x)]$.

Now consider two flow-unconnected points $s \nrightarrow t$ given y_u, exhaustive knowledge of the RF $Y(\cdot)$ on the common downstream domain $\wedge_u = \wedge_s \cap \wedge_t$

$$P[Y_s, Y_t|y_u] = P\left[\int_{\wedge_s \backslash \wedge_u} \int_{\wedge_t \backslash \wedge_u} g(x-s)g(z-t)\mathrm{d}B(z) \cap \mathrm{d}B(x)\right]$$

but this factors as

$$P[Y_s, Y_t|y_u] = P[Y_s|y_u]P[Y_t|y_u] = P\left[\int_{\wedge_s \backslash \wedge_u} g(x-s)\mathrm{d}B(x)\right] P\left[\int_{\wedge_t \backslash \wedge_u} g(z-t)\mathrm{d}B(z)\right] \tag{4}$$

by noting that $\wedge_s \backslash \wedge_u$ and $\wedge_t \backslash \wedge_u$ do not overlap, and the independence of the Brownian process increments $\mathrm{d}B(x)$ and $\mathrm{d}B(z)$ among branch segments, and hence we have conditional independence.

Note that *down* models share the properties that define the models introduced in Monestiez *et al.* (2005)—stationarity along the stream and conditional independence among branches. By construction they are admissible and provide explicit covariance models for flow-unconnected points that was previously missing.

2.3.2. Exponential model and isotropy. To obtain the exponential model of covariance, the kernel on $\mathbb{R}$ is

$$g(x) = \sqrt{\frac{2C_1}{r}}\, e^{-x/r}\, I_{\{0 \leq x\}}$$

Equation (1) yields the exponential covariance function

$$C_1 e^{-(\Delta)/r} \quad \text{if } 0 \leq \Delta \tag{5}$$

and Equation (2) yields the following expression:

$$C_1 e^{-(\Delta_1 + \Delta_2)/r} \quad \text{if } 0 \leq \Delta_2 \leq \Delta_1 \tag{6}$$

By simply adding the two distances Δ_1 and Δ_2, Equation (6) does not differ from Equation (5) for flow-connected pairs and thus this model features isotropy—covariance depends only on the distance between points—in the network.

2.3.3. Explicit form of the spherical covariance model. The kernel associated with the spherical covariance model is

$$g(x) = \sqrt{\frac{3C_1}{r}}(1 - x/r)\, I_{\{0 \leq x < r\}}$$

For flow-connected points, using Equation (1) results in the well-known expression for the spherical model of covariance

$$C_1 \left(1 - \left(\frac{3\Delta}{2r}\right) + \frac{1}{2}\left(\frac{\Delta}{r}\right)^3\right) \quad \text{if } 0 \leq \Delta \leq r \text{ and } 0 \text{ when } r \leq \Delta \tag{7}$$

For flow-unconnected points Equation (2) gives a slightly more complex covariance model:

$$C_1 \frac{(r - \Delta_2)^2(2r + \Delta_2 - 3\Delta_1)}{2r^3} \quad \text{if } 0 \leq \Delta_1 \leq \Delta_2 \leq r \tag{8}$$

whose plot is given in Figure 3.
    Note that the actual range for flow-unconnected pairs is larger than the range for flow-connected pairs.

2.3.4. Other covariances models. Beyond the basic models such as exponential, spherical or pure nugget, other models of covariance can be developed for any down-sided integrable and square integrable kernel. For example, in the flow-connected case, Table 1 gives the kernel and the associated covariance function for three models: the one-dimensional linear-with-sill model, the Mariah model first given by Ver Hoef *et al.* (2006), and a new one based on the Epanechnikov kernel. Table 2 gives the covariances for flow-unconnected pairs using the same three kernels.
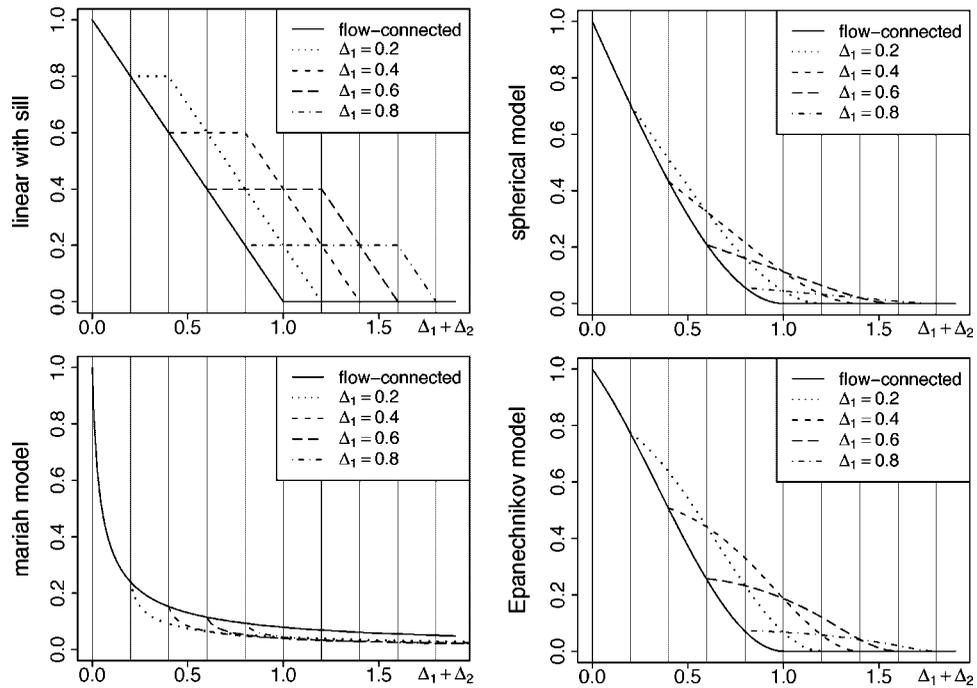
Figure 3. Theoretical covariance plots in the flow-unconnected context: linear-with-sill, spherical, Mariah model, Epanechnikov kernel model. Models are listed Tables 1 and 2 and Equations (7) and (8)

Table 1. Covariance functions resulting from different kernels for flow-connected pairs

Model	Kernel $g(x)$	Covariance function $C(\Delta)$
Linear-with-sill	$g(x) = \sqrt{\frac{C_1}{r}} I_{\{0 \leq x < r\}}$	$\begin{cases} C_1(1 - \Delta/r) & \text{if} \quad 0 \leq \Delta \leq r \\ 0 & \text{if} \quad r < \Delta \end{cases}$
Mariah model (Ver Hoef *et al.*, 2006)	$g(x) = \sqrt{\frac{C_1}{r}} \frac{1}{x/r+1} I_{\{0 \leq x\}}$	$\begin{cases} C_1 & \text{if} \quad \Delta = 0 \\ C_1 \cdot \dfrac{\ln\left(\frac{\Delta}{r}+1\right)}{\Delta/r} & \text{if} \quad 0 < \Delta \end{cases}$
Epanechnikov kernel model	$g(x) = \sqrt{\frac{15C_1}{8r}} \times \left(1 - \left(\frac{x}{r}\right)^2\right) I_{\{0 \leq x < r\}}$	$\begin{cases} \frac{C_1}{16r^5}(\Delta - r)^2 \cdot (16r^3 + 17r^2\Delta - 2r\Delta^2 - \Delta^3) & \text{if} \quad 0 \leq \Delta \leq r \\ 0 & \text{if} \quad r < \Delta \end{cases}$

2.4. Up models

Ver Hoef *et al.* (2006) and Cressie *et al.* (2006) first proposed models built by kernel smoothing of white noise. They use a unilateral kernel that points upstream of s (Figure 2(b)). Following

Table 2. Covariance functions for flow-unconnected pairs in *Down* models

Model	$C(\Delta_1, \Delta_2)$ flow-unconnected pairs
Linear-with-sill	$\begin{cases} C_1(1 - \Delta_2/r) & \text{if} \quad 0 \le \Delta_1 \le \Delta_2 \le r \\ 0 & \text{if} \quad r < \Delta_2 \end{cases}$
Mariah model	$\begin{cases} C_1 & \text{if} \quad \Delta_1 = \Delta_2 = 0 \\ C_1 \left(\ln(\Delta_2 + r) - \ln(\Delta_1 + r)\right) \frac{r}{\Delta_2 - \Delta_1} & \text{if} \quad 0 < \Delta_2 \\ C_1 \left(\frac{1}{\Delta_2/r + 1}\right) & \text{if} \quad \Delta_1 = \Delta_2 > 0 \end{cases}$
Epanechnikov kernel model	$\begin{cases} \frac{C_1}{16r^5} (\Delta_2 - r)^2 \left(16r^3 + 17\Delta_2 r^2 - 15\Delta_1 r^2 - 20r\Delta_1^2 \right. \\ \left. -2r\Delta_2^2 + 10r\Delta_1\Delta_2 + 5\Delta_1\Delta_2^2 - \Delta_2^3 - 10\Delta_2\Delta_1^2\right) & \text{if} \quad 0 \le \Delta_1 \le \Delta_2 \le r \\ 0 & \text{if} \quad r < \Delta_2 \end{cases}$

Cressie *et al.* (2006), a weighted integration is performed on the upstream branching kernel support:

$$Y_s = \mu(s) + \int_{\vee_s} g(x - s) \sqrt{\frac{\Omega(x)}{\Omega(s)}} \mathrm{dB}(x)$$

where $\Omega(x)$ are weights that check an additivity constraint to ensure stationarity in variance.

The weights ensure stationary variance in the following way. Without weighting, the variance of a random variable will be $\int_{\vee_s} [g(x - s)]^2 \mathrm{d}x$. Note that there can be many possible configurations of branchings upstream of s. If we simply allow the function to go unweighted up all branches, then a location with many branches upstream will have a larger variance than a location with few branches upstream. Intuitively, if we weight the function $g(x)$ with proportions that sum to 1 at each branch, then we retain the same area under the function $g(x)$ as we move upstream. Ver Hoef *et al.* (2006) describe and construct their random variables this way, and note that weighting accumulates as we move upstream. The additive function described by Cressie *et al.* (2006) is an equivalent and more compact way to write the relationship (Money *et al.*, in press; Ver Hoef and Peterson, in press). When weights are arbitrarily defined for all source segments, the weight of any point s is defined as the sum of the weights of upstream segments. For example, in Figure 4, for all points x of a segment i, $\Omega(x) = w_i$. If one define all weigths for all source segments $w_1 = w_2 = w_3 = w_4 = w_5 = 1$, then we have $w_6 = w_7 = 2$, $w_8 = 4$ and $w_9 = 5$. This was the choice made by Cressie *et al.* (2006): the additive function value is simply the number of upstream source segments.

The spatial covariance structure for all pair of points (Y_s, Y_t), is then

$$\mathrm{cov}(Y_s, Y_t) = \int_{\vee_s \cap \vee_t} g(x - s)g(x - t) \frac{\Omega(x)}{\sqrt{\Omega(s)\Omega(t)}} \mathrm{d}x \tag{9}$$

leading to a null covariance for flow-unconnected $s \nrightarrow t$ because $\vee_s \cap \vee_t = \emptyset$. For the flow-connected case, $s \rightarrow t$, the integral is on one line:

$$\mathrm{cov}(Y_s, Y_t) = \sqrt{\frac{\Omega(s)}{\Omega(t)}} \int_{\mathbb{R}^+} g(x)g(x - t + s) \mathrm{d}x \tag{10}$$
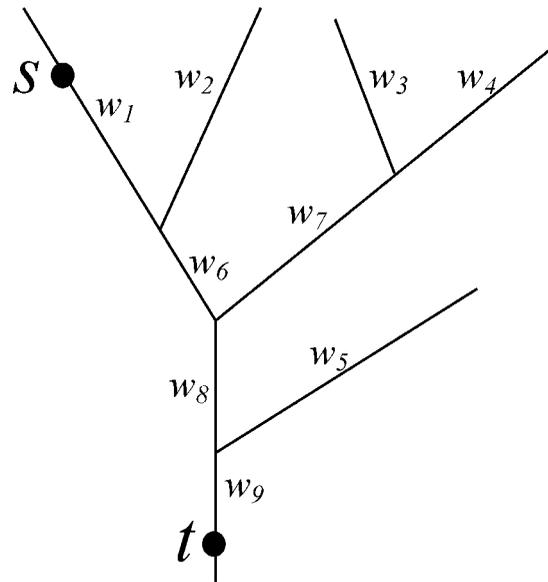
Figure 4.   Weighting example for *up* models

with the weighting due to the additivity of Ω (Cressie *et al.*, 2006). Other than the weights, model (10) yields the same covariance functions among flow-connected sites as the *down* models Equations (5) and (7), and those given in Table 1. However, the weighting clearly creates discontinuities at confluences.

2.5. Hybrid model

When building a hybrid RF model by summing independent RF, the resulting covariance model is the sum of covariance models. We are interested in a hybrid covariance model that balances *up* and *down* models

$$\text{cov}(Y_s, Y_t) = \sigma_1^2 \delta_{s=t} + \sigma_2^2 \left(p \sqrt{\frac{\Omega(s)}{\Omega(t)}} \rho_{\text{Up}}(s, t) + (1 - p)\rho_{\text{Down}}(s, t) \right) \tag{11}$$

where $\delta_{s=t}$ is the Dirac mass applied at $s = t$, σ_1^2 is the nugget effect, and σ_2^2 is the variance of the spatially autocorrelated part of the model. $\rho_{\text{Up}}(s, t)$ and $\rho_{\text{Down}}(s, t)$ are *up* and *down* autocorrelation models and p is a parameter, constrained in $[0, 1]$ to allow identifiability, representing *up* correlation weight in the spatially autocorrelated part of the model. Note that Equation (11) is not isotropic; i.e., depending only on the distance between two points. It keeps the two main characteristics of *up* and *down* models; a discontinuity at confluences like *up* models and conditional independence between branches given all common downstream segments as in *down* models.

3. METHODS FOR ESTIMATION AND PREDICTION

Classical geostatistics fits a parametric variogram (covariance) to an empirical variogram (covariogram), often obtaining variogram (covariance) parameter estimates by (weighted) least squares, mitigating any distributional assumptions on the RF. Note that simple and robust computation of the empirical variogram (all squared-differences among pairs of points plotted against distance) needs stationarity and isotropy of the RF. However, note that *down* models, except the exponential, are not isotropic. To fit this kind of model a partitioned variogram could be used for flow-connected and flow-unconnected pairs (Ver Hoef *et al*., 2006). *Up* models also have nonstationary covariances, so a corrected empirical covariogram would be necessary for inference (in the same way that weighting occurs in the models). Instead, we chose maximum likelihood estimation, assuming a Gaussian RF distribution. This allowed us to compare, in a consistent and robust way, different types of covariance models, especially through the use of AIC (Akaike (1974)).

3.1. Maximum Likelihood estimation

We introduce the following notation: $Y(s_\alpha) = Y_\alpha$, with $\alpha = 1, \ldots, n$, are measurements of the RF $Y(\cdot)$ at the s_α points, and $Y \equiv (Y_1, \ldots, Y_n)$ is the vector of measurements. Let $Y^*(s_0) = Y_0^*$ be the predicted value at point s_0. Also let $C(s_\alpha, s_\beta, \theta) = C_{\alpha\beta}(\theta)$ be the covariance between $Y(s_\alpha)$ and $Y(s_\beta)$ with parameter vector θ, also denoted $C_{\alpha\beta}$ when the model is chosen and parameters are fitted.

Then the likelihood function of the gaussian RF $Y(\cdot)$ with $C(\cdot, \cdot, \theta)$ covariance model and a constant mean μ is given by

$$L(Y, \mu, \theta) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma(\theta)|}} \exp\left(-\frac{1}{2}(Y - \mu\,\mathbb{1})^t \Sigma(\theta)^{-1}(Y - \mu\,\mathbb{1})\right)$$

where $\Sigma(\theta)$ is the covariance matrix with $C_{\alpha\beta}(\theta)$ as its α, β element and $\mathbb{1}$ the a vector of 1's of length n. Instead of maximizing the likelihood function, to reduce computing error we prefer to minimize the function:

$$l(Y, \mu, \theta) \propto -2\log(L(Y, \mu, \theta))$$

$$= \sum_{i=1}^{n} \log(\text{eigen}(\Sigma(\theta))) + (Y - \mu\,\mathbb{1})^t \Sigma(\theta)^{-1}(YY - \mu\,\mathbb{1})$$

This function is minimized for each covariance model using a Newton type algorithm (`nlm` function in R Development Core Team (2007)).

3.2. Model selection

The different RF models we proposed are intrinsically non-nested so the comparison of *up* and *down* models as well as the different covariance functions requires the use of a selection criterion or a measure of the model predictive quality.

3.2.1. AIC selection criterion. We choose to use the classic AIC selection criterion (Akaike, 1974; Burnham and Anderson, 2002) which estimates the expected relative distance of each model to a 'true'

unknown model. The AIC of a RF model is computed using

$$\text{AIC} = -2\log(L(\boldsymbol{Y}, \hat{\mu}, \hat{\boldsymbol{\theta}})) + 2k$$

where $L(\boldsymbol{Y}, \hat{\mu}, \hat{\boldsymbol{\theta}})$ is the likelihood of the RF model computed at $(\hat{\mu}, \hat{\boldsymbol{\theta}})$, the maximum likelihood estimates of $(\mu, \boldsymbol{\theta})$, and where k is the dimension of this parameter set.

The best model is the one with the smallest AIC, meaning that its likelihood is the highest, downweighted by its number of parameters. This criterion, based on logarithm, is used to compare models based on their difference. So we will look at ΔAIC, defined as the difference between AIC and minimum of AIC over candidate models. In the case of nested models, Burnham and Anderson (2002) suggest an empirical rule to discriminate or elect nested models: a ΔAIC of 0–2 means that both the minimal AIC model and the considered model are supported by data, a 4–7 difference means that the model with minimal AIC is considerably more accurate, and a difference of more than 10 discriminates totally for the model. This guideline may be somewhat larger for non-nested models. Note that, in the geostatistical context where we have dependent data, criterions such as Bayesian information criterion (BIC, Schwarz, 1978) and all corrected AIC, which use sample size for weighting, should be avoided.

3.2.2. Mean square prediction error measure. Mean-square-prediction-error (MSPE) is a measure of the predictive power of a model. It is estimated by leave-one-out cross-validation after estimating the covariance parameters. That is, we fit $\boldsymbol{\theta}$ covariance parameters by maximum likelihood using the whole dataset of size n, and then predict by kriging each point in turn using all the others and the previously estimated covariance function. Each prediction is compared to its associated 'left-out' real value and the MSPE is computed as

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i^*)^2$$

where Y_i^* is the ith point kriged using all the other $n-1$ points. This measure does not account for parsimony of the considered model (the task of AIC) but gives us an idea of its predictive ability.

3.3. Kriging on river networks

After estimating covariance parameters and choosing a model, the covariance between $Y(s_\alpha)$ and $Y(s_\beta)$ is denoted $C_{\alpha\beta}$. We interpolate unknown Y_0^* at point s_0 of the RF $Y(\cdot)$ by ordinary kriging (Cressie, 1991) which gives us best linear unbiased predictor (BLUP).

Y_0^* is a linear predictor

$$Y_0^* = \sum_{\alpha=1}^{n} \lambda_\alpha Y_\alpha$$

Assuming the unbiasedness constraint $\sum_{\alpha=1}^{n} \lambda_\alpha = 1$ and noting μ the unknown constant mean of $Y(\cdot)$, we have

$$E\left[Y_0^*\right] = E\left[\sum_{\alpha=1}^{n} \lambda_\alpha Y_\alpha\right] = \sum_{\alpha=1}^{n} \lambda_\alpha E\left[Y_\alpha\right] = \mu\left(\sum_{\alpha=1}^{n} \lambda_\alpha\right) = \mu$$

The variance of prediction error, written as $\mathrm{Var}(Y_0^* - Y_0) = E[(Y_0^* - Y_0)^2]$, is

$$\mathrm{Var}(Y_o^* - Y_o) = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C_{\alpha\beta} + \sigma_o^2 - 2\sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o}$$

where σ_o^2 is the unknown variance of Y_o.

Minimizing $\mathrm{Var}(Y_o^* - Y_o)$, including the unbiasedness constraint, we obtain the λ_α kriging weights, using a Lagrange multiplier η to solve the linear system of $n + 1$ equations:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} - \eta = C_{\alpha o} & \text{for} \quad \alpha = 1, \ldots, n, \\ \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \end{cases} \tag{12}$$

involving an $(n + 1)$ dimensional matrix inversion and one matrix vector multiplication.

The variance of the ordinary kriging predictor is computed as

$$\mathrm{Var}(Y_0^* - Y_0) = \sigma^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha o} + \eta \tag{13}$$

where σ^2 is the variance of $Y(\cdot)$ already inferred and η is a solution from Equation (12).

4. RIVER NETWORK CASE STUDY

4.1. River network characteristics and monitored variables

The spatial river network analysis was applied to the Meuse and Moselle basins, which are limited to the northeast of France by the German border (East and North), the Luxembourg border (North) and the Belgium border (North). These borders define eight sub-river networks which are considered independent since we do not know their confluence. Altogether, they accumulate more than 5000 km of river network. Both summer temperature and nitrate rates were measured at 141 monitoring stations plotted in Figure 5. Summer temperature was also measured on additional stations giving a total number of 187 sites for temperature. The different models were fitted and compared on the whole data set for each variable. Summer temperature is approximatively normally distributed while asymmetry in the distribution of nitrate rates required a log transformation of the raw data before analysis. In the following, in order to show more readable maps, kriging is illustrated only on the Moselle subnetwork; however all analyses were computed on the whole network.
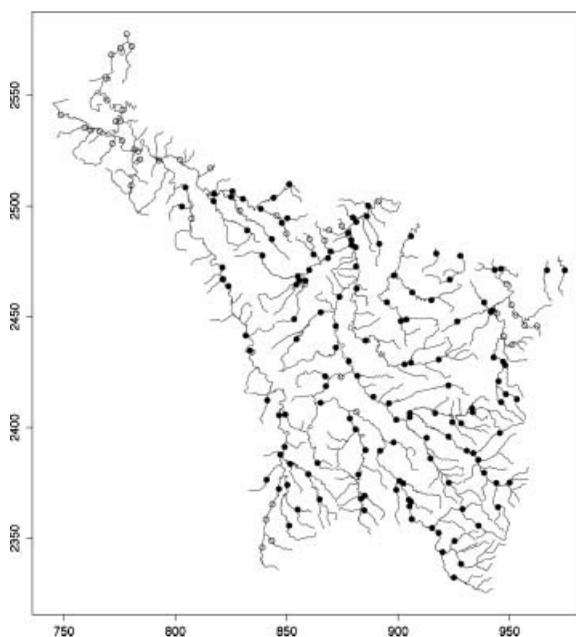
Figure 5.   Map of the Meuse and Moselle network. Filled symbols locates temperature and nitrate monitoring stations. Empty symbols denotes additional sites where only temperature were available. Coordinates for x and y axes are in Lambert II and unit in km

4.2. Model selection

Likelihood estimates and AIC criteria were computed for four different models; two *down* models with isotropic exponential covariance and spherical covariance respectively, one *up* model with exponential covariance using weights as described by Cressie *et al.* (2006), and finally a hybrid *up* and *down* model with exponential covariances. We limited our comparison to a small number of potential models, one for each category: *up*, *down* isotropic, *down* nonisotropic and hybrid. Results for AIC and MSPE are summarized in Table 3. Note that for nitrates, the MSPE is computed on back-transformed predictions and given in $(mg/l)^2$.

For summer temperature, both criteria give a clear advantage to the *up* model compared to the *down*. Meanwhile, it is the hybrid model that shows the best fit to spatial variation in summer temperature.

Table 3.   Relative differences on AIC criterion and computed MSPE by cross validation for temperature in $(°C)^2$ and nitrate rates in $(mg/L)^2$

	ΔAIC		MSPE	
	Temperature	Nitrate	Temperature	Nitrate
Down model exp. isotropic	74.3	32.2	2.21	10.67
Down model spherical	77.1	31.2	2.20	10.68
Up model, Cressie's weights	19.4	40.3	1.65	13.73
Hybrid model up and down	0	0	1.25	9.30

Table 4. Fitted parameters for the two selected hybrid models

Hybrid model up and down	Nugget	Var. up	Var. down	Range up	Range down
Summer temperature	0.311	1.122	1.533	∞	44.3
	10.5%	37.8%	51.7%		
Nitrate rates	0.014	0.085	0.273	∞	5.51
	3.7%	22.8%	73.5%		

Percentages of variance between nugget effect, *up* part and *down* part are given followed by respective scale parameters which are about one third of the effective range for these exponential covariance models.

The hybrid model is also preferred for nitrate rates with a significant ΔAIC, but differences between *down* and hybrid models remain low in term of MSPE.

Table 4 gives the estimated parameters for the two selected hybrid models. For both variables the estimated range for the *up* part of the model is infinity, corresponding to a RF that is simply a step function with discontinuities at confluences. Because more than one measurement rarely exists between two successive confluences, the range parameter is very poorly estimated. Essentially, the weights capture all spatial variation.

For temperature, the *up* and *down* components are quite well balanced even if the *up* type seems to be preferred at the selection step. The effective range of the down part is around 150 km, which is large compared to the largest distance computed between flow-connected points.

For nitrates, the *up* component explains considerably less variation than the *down* component, confirming the ranking obtained from AIC. The *down* part shows a limited effective range of 10–20 km, highlighting local spatial variation.
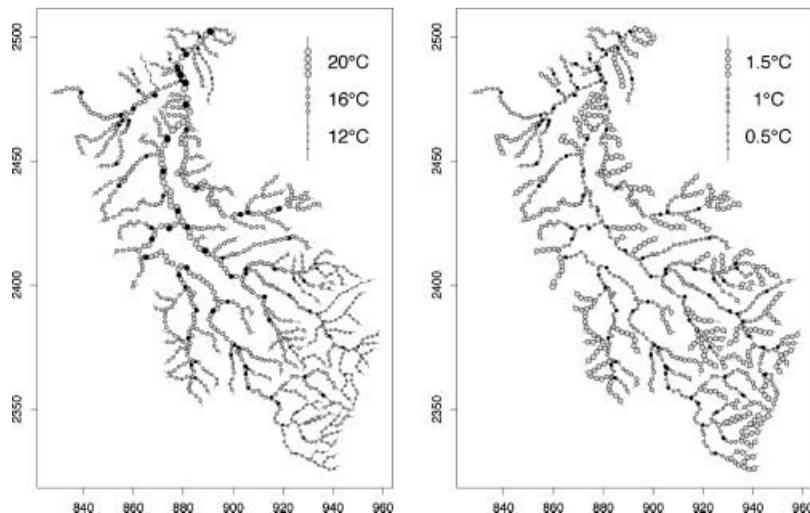

Figure 6. Map of kriged summer temperatures on the Moselle subnetwork (left) and map of standard error of prediction. Black dots are located at data sites and grey dots features interpolated points. Measured and predicted values of temperatures control the symbol sizes on the left and predicted errors on the right. Coordinates for x and y axes are in Lambert II and unit in kilometre

4.3. Kriged maps and error prediction

Maps were obtained by plotting kriged values (Figure 6, left) and predicted errors (Figure 6, right) at equidistant points laid along the network. Kriging values for Temperature were obtained for each point by solving the system (12) with the selected hybrid model. The standard errors of the predictions were computed as the square root of the corresponding kriging variance (13).

Temperature discontinuities at confluences, which are associated with the *up* model, are not highly visible except at a few confluences (Figure 6, left). These discontinuities are somewhat masked by the resolution of mapping, but also by a long range trend which is modelled here by the *down* range parameter. This could also be modelled as a non-stationarity mean trending from a mountainous area to plains.

The map of predicted error (Figure 6, right) shows several interesting patterns of the *up* model:

- segments with measurements have low prediction errors due to the infinity range in the *up* model part, the error resulting only from the *down* part of the model,
- the prediction error shows steps between segments with and without measurements because of the weighting system which creates discontinuities at confluences,
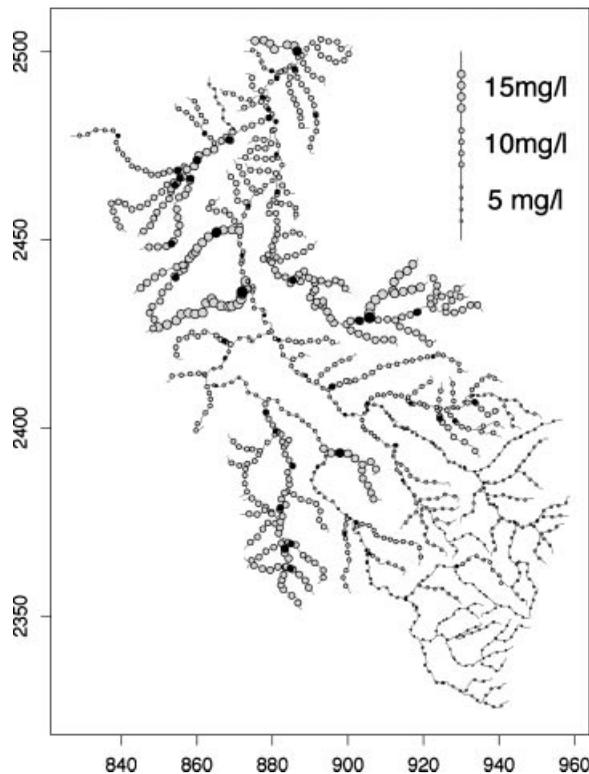

Figure 7. Map of kriged nitrate rates on the Moselle subnetwork. Black dots are located at data sites and grey dots features interpolated points. Measured and predicted values of nitrate control the symbol sizes. Coordinates for x and y axes are in Lambert II and unit in kilometre

- the step between a measured segment and a non-measured one is higher when the latter has little weight. This is apparent when a source segment has a confluence with stream of higher order.

Patterns induced by the *down* part of the model are not as visible due to the long effective range.

Kriging of nitrate rates (Figure 7) are performed on log-transformed values using the selected hybrid covariance model and then back-transformed by exponentiating before mapping. This is equivalent to kriging the median, which has been recommended over lognormal kriging (see, e.g. Chilés and Delfiner (1999)). Back-transforming the confidence interval end-points maintains the proper coverage as well. The corresponding 'error' maps are then the 95% confidence interval spans measured in mg/L (Figure 8, left), and we also present the confidence interval span as a per cent of the predicted value (Figure 8, right).

For kriged nitrates (Figure 7), the discontinuities associated with the *up* model part are small but visible at some confluences. The main variation from the *down* model part shows strong regionalization in small watersheds (15–20 km) which results from geographic structuring of pollution sources. It may be caused by agricultural practices whose spatial structuring are linked to stream segments, but in a complicated way throughout the river network.

The predicted error confidence interval spans (Figure 8, left) are mapped as absolute. This representation mostly hides errors in regions of low nitrate rates. When mapping error interval span relative to the kriged value (Figure 8, right), we obtain a more informative map of the prediction quality. Errors increase dramatically for non-measured source segments due to the mix of the *down* type model with short range and the *up* type model.
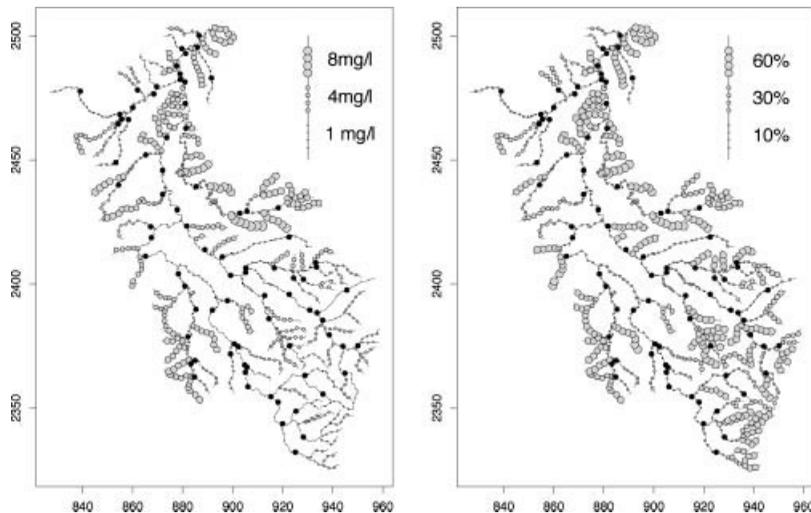

Figure 8. Map of 95% confidence interval span for nitrate prediction (left) and of relative confidence interval span (CI span/predicted value) (right). Black dots are located at data sites and grey dots features interpolated points

5. DISCUSSION

Both *up* and *down* models show a real efficiency in spatial predictions along the river network. When compared to a geostatistical approach solely based on geographical distance, the proposed models can be directly related to the river based on physical and chemical processes within a network. For example Ver Hoef *et al*. (2006) interpret their *up* covariance model as diffusion of a passive element following stream flow and mixed at each confluence. The *down* isotropic exponential model is intuitively interpreted as corresponding to a diffusion without taking the stream flow into account, etc.

In practice, it seems desirable to systematically test the mixture of the *up* and *down* models for real data. For actual environmental variables, a pure *up* or a pure *down* model is highly improbable due to the fact that these variables are generally influenced by several natural processes.

The way our covariances are built using unilateral kernels give only covariance models with strictly negative slope at the origin. Smoother models that are twice differentiable at the origin could be obtained by bilateral kernels differentiable at the origin.

The *up* model, which forms part of the hybrid model as well, introduces discontinuities in covariance and in the RF at each confluence. It is a rather atypical characteristic for covariance functions, which arises due to the desire to maintain stationarity in the variance. The choice made by Ver Hoef *et al*. (2006) and Cressie *et al.* (2006), maintaining stationarity in variance seems natural when modelling bounded processes such as mixing of a pollutant. When considering other processes showing an accumulation it may be useful to consider the opposite balance: a stationary covariance and non-stationnary variance. In theory, this could be handled by a different weighting scheme.

A problem with the *up* models is their identifiability when spatial samples are homogeneously distributed and rarely with more that one point between two confluences. This can lead to estimates of infinite range and the RF is modelled with constant values between confluences and discontinuities at confluence, giving a crucial role to the weighting system of Ω's.

This indicates that more work needs to be done on the spatial sample scheme. For exemple, several sampled sites between confluences should improve *up* part estimation, sampled sites surrounding a confluence will give more information on the system of Ω weights, and sampling the periphery of the network close to sources should give good overall error maps. Finding a good compromise with limited sample size on large network is a complex optimization problem.

Another perspective is to generalize the different models, *up*, *down* and hybrid, within the general framework of bilateral kernels. However, deriving analytical expressions of covariance becomes a challenging problem and interpretation of covariance structures by natural mechanisms becomes less straightforward.

A final extension of these models is to use them as latent variables in hierarchical spatial models. For example, we could consider modelling the spatial distribution of a disease potential along a river network, using these kinds of models as a latent variable for the mean (using the appropriate link function) of a Poisson distribution for count data, or the mean of a Bernoulli distribution for presence/absence data.

REFERENCES

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**: 716–723.
Bailly J-S, Monestiez P, Lagacherie P. 2006. Modelling spatial processes on directed trees with geostatistics: application to artificial network properties on rural catchment areas. *Mathematical Geology* **38**: 515–539.
Barry RP, Ver Hoef JM. 1996. Blackbox kriging: spatial prediction whithout specifying the variogram. *Journal of Agricultural, Biological and Environmental Statistics* **1**: 297–322.

Billingsley P. 1995. *Probability and Measure* (3rd edn). John Wiley and Sons: New York.

Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach* (2nd edn). Springer-Verlag: New York.

Chilés J-P, Delfiner P. 1999. *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. Wiley: New York.

Cressie N. 1991. *Statistics for Spatial Data*. Wiley: New York.

Cressie N, Frey J, Harch B, Smith M. 2006. Spatial prediction on a river network. *Journal of Agricultural, Biological and Environmental Statistics* **11**: 127–150.

Cressie N, Majure J. 1997. Spatio-temporal statistical modelling of livestock waste in streams. *Journal of Agricultura, Biological and Environmental Statistics* **2**: 24–47.

Cressie N, Pavlicova M. 2002. Calibrated spatial moving average simulations. *Statistical Modelling* **2**: 1–13.

De Fouquet C, Bernard-Michel C. 2006. Modèles géostatistiques de concentrations ou de débits le long des cours d'eau. *C. R. Géoscience* **338**: 307–318.

Higdon D. 2001. Space and space-time modeling using process convolutions. *Technical report*, Duke University, Institute of Statistics and Decision Science.

Monestiez P, Bailly J-S, Lagacherie P, Voltz M. 2005. Geostatistical modelling of spatial processes on directed trees: application to fluvisol extent. *Geoderma* **128**: 179–191.

Money E, Carter G, Serre ML. (in press). Using river distances in the space/time estimation of dissolved oxygen along two impaired river networks in new jersey. *Water Research*.

Øksendal B. 2003. *Stochastic Differential Equations: An Introduction with Applications* (6th edn). Springer-Verlag: Berlin Heidelberg.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0

Schwarz G. 1978. Estimating the dimension of a model. *The Annals of Statistics* **6**: 461–464.

Ver Hoef JM, Peterson E. (in press). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*.

Ver Hoef JM, Peterson E, Theobald D. 2006. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* **13**: 449–464.