Bayesian spatio-temporal kriging with misspecified black-box

Papa-Abdulaye Faye, Pierre Druilhet, Nourddine Azzaoui, Anne-Françoise Yao

Bayesian spatio-temporal kriging with misspecified black-box

Papa Abdoulaye FAYE, Pierre DRUILHET*, Nourddine AZZAOUI, Anne-Françoise YAO

Laboratoire de Mathématiques, Université Blaise Pascal, UMR CNRS 6620, Clermont-Ferrand

Abstract

We propose a new algorithm for spatio-temporal prediction. At a given time t, we use a Bayesian kriging model for spatial prediction. The temporal evolution from t to $t+1$ is given by a deterministic black-box which can be a complex numerical code or a partial differential equation. As often in practice, the black-box is misspecified, in the sense that its parameters are imprecisely known or may be varying randomly over time. At time t, we use the black-box to obtain a rough prediction at time $t+1$. When new data are available, the black-box is used to estimate the hyperparameters of the Bayesian kriging at time $t+1$ by using Monte Carlo methods. Through a numerical application, we show that our method improves the values predicted by the black-box only.

Keywords: `spatio-temporal prediction`, `Monte Carlo methods`, `Bayesian kriging`, `misspecified black-box`

1. Introduction

Spatio-temporal modeling is a fundamental step to understand the mechanisms that govern the evolution of a natural phenomenon. It arises when data are collected across time as well as space. The spatio-temporal models take into account temporal correlations as well as spatial correlations of data. It allows to reconstruct a phenomenon over a domain from a set of observed values. Such a reconstruction problem occurs in many areas such as climatology, meteorology, geology, atmospheric sciences, hydrology, environment, geography etc. For example, it occurs when one aim to predict an atmospheric pollutant from a monitoring network which provides data that are collected at regular intervals.

There exists many spatio-temporal models in spatial statistic literature, see Banerjee et al. (2004) or Cressie and Wikle (2011) for a review. For example Cressie and Wikle (2006) consider spatio-temporal kriging by using Kalman Filter methods. Irwin et al. (2002) use spatio-temporal nonlinear filtering based on hierarchical statistical models. Hengl et al. (2012) propose a procedure to interpolate daily mean temperature over a whole year period by using time series of auxiliary predictors. Stein (2005) considers a number of properties of space-time covariance functions and how these relate to the spatial-temporal interactions

*Corresponding author

Email address: `Pierre.DRUILHET@univ-bpclermont.fr` (Pierre DRUILHET)

of the process. Ip and Li (2015) construct valid parametric covariance models which are computationally estimable for univariate and multivariate spatio-temporal random fields. Le Gratiet (2013) proposes Bayesian hierarchical multi-fidelity methods with covariates.

The research on spatio-temporal modelling is ongoing to deal with problems that are more and more complex. The temporal models are often given by a deterministic black-box corresponding to the modelization of the studied phenomenon. However, in practice, the black-box is often misspecified and gives a rough prediction. On the other hand, the new data give accurate information but only around the sites of observations. In the paper, we propose to combine these spatial and temporal information.

At given time t, we use Bayesian kriging to obtain the spatial prediction of phenomenon of interest. The black-box is used to obtain a first prediction at time $t+1$ and also to derive new prior distributions of the Bayesian kriging at time $t+1$. The latter is based on the new data available at time $t+1$.

The article is structured as follows: Section 2 provides a description of Bayesian kriging. In Section 3, we present our spatio-temporal prediction procedure. In Section 4, we present a numerical application.

2. Spatial Bayesian kriging

In this section, we give a presentation of Bayesian kriging. Let $\left(Z(s) \in \mathbb{R}^p, s \in \mathcal{D} \subset \mathbb{R}^d\right)$ be a random Gaussian spatial process. Most often $d = 1$, 2 or 3. The random process is observed at a few number of sites and the aim is to predict the values of $Z(.)$ at some unobserved locations.

We recall that the kriging involves the construction of a linear predictor who takes into account of the structure covariance of $Z(.)$. Classically, the stochastic model associated with kriging is defined by:

$$Z(s) = \mu(s) + \delta(s) + \epsilon(s), \quad s \in \mathcal{D}, \tag{1}$$

where:

- $Z(s)$ is called the regionalized random variable.
- $\mu(s) := \beta' (m(s))$ (β' being the transpose of β), is the deterministic component of $Z(s)$; with $m(s) \in \mathbb{R}^p$ is a vector of a known basis functions and $\beta \in \mathbb{R}^p$ is an unknown coefficients vector to be estimated.
- $\delta(.)$ a stationary Gaussian random field with zero mean and correlation function defined by $g(s, \tilde{s}) = \mathrm{Cov}(\delta(s), \delta(\tilde{s}))/\sigma^2$, for s, $\tilde{s} \in \mathcal{D}$ and $\sigma^2 = \mathrm{Var}(\delta(s))$.
- $\epsilon(.) = (\epsilon(s), s \in \mathcal{D})$ is a spatial white noise independent of $\delta(.)$. For $s \in \mathcal{D}$ the measurement error $\mathrm{Var}(\epsilon(s)) = \tau^2$ is called nugget effect.

If $\delta(.)$ is isotropic, the correlation function is reduced to the correlogram $g(h) = g(s, \tilde{s})$, where $h = \|s - \tilde{s}\|_2$ is the euclidean distance between s and $\tilde{s}$. One of the most popular family of correlogram is the Matérn family defined by:

$$g(h, \alpha, \kappa) = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1} \left(\frac{h}{\alpha}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\alpha}\right), \tag{2}$$

with $K_\kappa(.)$ denotes the modified Bessel function of order κ (see Gneiting et al. (2010) for further details). The corresponding variogram is

$$\gamma(h) = \mathrm{Var}(Z(s+h) - Z(s)) = \tau^2 + \sigma^2 \left\{1 - g(h;\phi)\right\}.$$

We denote by ϕ the correlation parameter, here $\phi = (\kappa, \alpha)$ in (2). In that follows, we assume that $Z(.)$ is observed at the sites $s_1, \dots s_n$ in $\mathcal{D}$. Let $Y = (Z(s_1), \dots, Z(s_n))$ represents the corresponding observations.

In the Bayesian framework, we put prior distributions on the parameters $(\phi, \tau^2, \sigma^2, \beta)$. One interest is to take into account prior knowledge and uncertainty on the parameters estimation.

2.1. Prior distribution specification

We first consider the situation where $\tau^2 = 0$, i.e we assume that there is no nugget effect and that the prior distribution is separable:

$$\pi(\beta, \sigma^2, \phi) = \pi(\phi)\pi(\sigma^2)\pi(\beta). \tag{3}$$

For practical reasons, we choose the following priors

$$\begin{array}{rcl} [\phi] & \sim & \Gamma(a, b) \\ [\sigma^2] & \sim & \mathcal{X}^2_{ScI}(\nu, S^2) \\ [\beta] & \sim & \mathcal{N}(e, V) \end{array} \tag{4}$$

where
- $\Gamma(a, b)$ is a gamma distribution with shape parameter a and scale parameter b.
- $\mathcal{X}^2_{ScI}(\nu, S^2)$ is a scale inverse chi-squared distribution with ν represents the number of chi-squared degrees of freedom and S^2 the scaling parameter.
- $\mathcal{N}(e, V)$ is a multivariate Gaussian distribution where e and V respectively represent the mean and the standardized covariance hyper-parameters for the prior distribution for the mean (vector) parameter β.

At starting time, if there is no prior information about the behavior of the phenomenon, one can use a flat prior $\pi(\beta) \propto 1$ for the parameter β, which corresponds to the limit case $V^{-1} = 0$ and the Jeffrey's prior $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$, for the parameter σ^2, which corresponds to $\nu = 0$ in (4), e.g. see Gelman et al. (2014); Diggle and Ribeiro (2002); Bernardo (1979); Bioche and Druilhet (2015).

2.2. Posterior distribution parameters and predictive distribution of regionalized variable

The posterior distributions for the prior (4) are given by, e.g. see Diggle and Ribeiro (2007):

$$\begin{array}{rcl} [\beta \mid \sigma^2, \phi, y] & \sim & \mathcal{N}(\tilde{e}, \sigma^2 \tilde{V}) \\ [\sigma^2 \mid \phi, y] & \sim & \mathcal{X}^2_{ScI}(\tilde{\nu}, \tilde{S}^2) \\ \pi(\phi \mid y) & \propto & \pi(\phi) |\widetilde{V}|^{\frac{1}{2}} |R(\phi)|^{-\frac{1}{2}} (S^2)^{-\frac{n+\nu}{2}}, \end{array} \tag{5}$$

with $\tilde{\nu} = \nu + n$, $\tilde{e} = \tilde{V}(V^{-1}e + M'(R(\phi))^{-1}y)$, $\tilde{V} = (V^{-1} + M'(R(\phi))M)^{-1}$ and

$$\tilde{S}^2 = \frac{\nu\tilde{V} + e'V^{-1}e + y'\left(R(\phi)\right)^{-1}y - \tilde{e}'\tilde{V}^{-1}\tilde{e}}{\nu + n},$$

where $M = (m(s1), \ldots, m(s_n))$, $R(\phi)$ is the correlation matrix of $(s_1, \ldots, s_n)$, and y is a realization of Y.
To accommodate a positive nugget variance $\tau^2 > 0$, in practice we use a discrete joint prior for ϕ and υ^2, where $\upsilon^2 = \dfrac{\tau^2}{\sigma^2}$. A prior distribution for υ^2 can be:

$$[\upsilon^2] \sim \mathcal{IG}(c, d), \tag{6}$$

where $\mathcal{IG}(c, d)$ is an inverse-gamma distribution with shape parameter c and scale parameter d.
In this case, we replace $R(\phi)$ in the equations above by:

$$V(\phi, \upsilon^2) = R(\phi) + \upsilon^2\mathbb{I}_n,$$

where $\mathbb{I}_n$ is $n \times n$ identity matrix, n is the number of observations.

We used the posterior distribution of kriging parameters $\pi(\phi, \upsilon^2, \sigma^2, \beta \mid y)$ to obtain the predictive distribution at an unobserved site s:

$$p\left(z(s) \mid y\right) = \int p\left(z(s) \mid y, \phi, \upsilon^2, \sigma^2, \beta\right) \pi(\phi, \upsilon^2, \sigma^2, \beta \mid y)\, d(\phi, \upsilon^2, \sigma^2, \beta). \tag{7}$$

This Bayesian predictive distribution is an average of the predictive distributions for fixed value of $(\phi, \upsilon^2, \sigma^2, \beta)$, weighted with respect to the posterior distributions of $(\phi, \upsilon^2, \sigma^2, \beta)$. In practice, $[Z(s) \mid y]$ is not easy to obtain analytically, but it can be easily approximated by Monte Carlo methods. An estimate of $Z(s)$ is given by the mean, median or mode of $[Z(s) \mid y]$.

We now address the spatio-time prediction method based on the Bayesian kriging and the black-box. Actually, we deal with a process $(Z_t(s), s \in \mathcal{D})$ which is a stationary Gaussian process for each time $t \in \mathbb{R}^+$. As previously, we aim to predict the values of $Z_t(s)$ at unobserved sites s of $\mathcal{D}$.

3. Spatio-temporal modeling including misspecified black-box

A black-box is a tool which allows to follow the temporal evolution of complex dynamic systems. For example, in domains such as fluid mechanics, ecology or biology, the black-box may be an algorithm or a partial differential equation (PDE). For given inputs at time t, the black-box gives a prediction of the outputs at time $t + 1$. However, the black-box is often misspecified, i.e. its parameters are imprecisely known and are often considered as random variables.

We denote by $\mathcal{M}$ a mesh of $\mathcal{D}$, $\mathcal{M}_0 = (s_1, \ldots, s_n)$ a subset of $\mathcal{M}$, $Z_t^{\mathcal{M}_0} = (Z_t(s_1), \ldots, Z_t(s_n))$, the observations on $\mathcal{M}_0$ at time t and $\theta_t = (\beta_t, \sigma_t^2, \upsilon_t^2, \phi_t)$ the kriging parameter at time t. The idea is to predict the unobserved values of $Z_t(.)$ in the field $\mathcal{D}$ by combining Bayesian kriging with the information brought by the fuzzy temporal model. Now, we present our spatio-temporal procedure:

Step 1*: Bayesian kriging*

At time t, we obtain the predictive distribution of the regionalized variable $Z_t^{\mathcal{M}} = \{Z_t(s),\ s \in \mathcal{M}\}$ by Bayesian kriging where the prior distributions are obtained from time $t-1$:

$$p\left(z_t^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}\right) = \int p\left(z_t^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}, \theta_t\right) \pi(\theta_t \mid z_t^{\mathcal{M}_0})\ d(\theta_t), \tag{8}$$

Step 2*: temporal prediction maps*

From the black-box, denoted by f, and from distribution (8), the prediction distribution at time $t+1$ of $Z_{t+1}^{\mathcal{M}}$ is given by

$$\left[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}\right] = \left[f(Z_t^{\mathcal{M}}) \mid z_t^{\mathcal{M}_0}\right]. \tag{9}$$

In practice (9) is not accessible explicitly, but it can be easily approximated by using Monte Carlo methods as follows:

- First, do K simulations of (8) and get K spatial prediction maps $z_t^{\mathcal{M}[i]}$, $i = 1, \ldots, K$.
- Then, get K temporal prediction maps at time $t+1$ by $\widehat{z}_{t+1}^{\mathcal{M}[i]} = f\left(z_t^{\mathcal{M}[i]}\right)$, $i = 1, \ldots, K$.

Step 3*: prior distribution specification*

Using the K prevision maps, at time $t+1$, we can update the hyperparameters of the prior distribution for the Bayesian kriging at time $t+1$. For each map $\widehat{z}_{t+1}^{\mathcal{M}[i]}$, $i = 1, \ldots, K$, we compute the maximum likelihood estimate $\widehat{\theta}_{t+1}^{[i]} = \left(\widehat{\phi}_{t+1}^{[i]}, \widehat{\upsilon^2}_{t+1}^{[i]}, \widehat{\sigma^2}_{t+1}^{[i]}, \widehat{\beta}_{t+1}^{[i]}\right)$ of θ_{t+1} from the kriging model

$$\left[Z_{t+1}^{\mathcal{M}} \mid \theta_{t+1}\right] \sim \mathcal{N}\left(M_{t+1}\beta_{t+1}, \sigma_{t+1}^2\left(R(\phi_{t+1}) + \upsilon_{t+1}^2 \mathbb{I}_q\right)\right), \tag{10}$$

with $M_{t+1} = (m_{t+1}(s_1), \ldots, m_{t+1}(s_q))$, q is the number of points of $\mathcal{M}$, $R(\phi_{t+1})$ is the correlation matrix of $\mathcal{M}$ and $\mathbb{I}_q$ is $q \times q$ identity matrix.

From $\widehat{\theta}_{t+1}^{[i]}$, $i = 1, ..., K$ we can estimate the hyper-parameters a_{t+1}, b_{t+1}, c_{t+1}, d_{t+1}, e_{t+1}, V_{t+1}, ν_{t+1}, S_{t+1}^2 given in (4) and in (6) by moment methods.

- For the correlation parameter ϕ_{t+1} which follows a gamma prior distribution $\Gamma(a_{t+1}, b_{t+1})$:

$$\begin{cases} \widehat{a}_{t+1}\widehat{b}_{t+1} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K} \widehat{\phi}_{t+1}^{[i]} \\ \widehat{a}_{t+1}\widehat{b}_{t+1}^2 &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\phi}_{t+1}^{[i]} - \sum_{j=1}^{K} \widehat{\phi}_{t+1}^{[j]}\right)^2. \end{cases}$$

- For the scale parameter σ^2_{t+1} which follows a inverse-chi-squared prior distribution $\mathcal{X}^2_{ScI}(\nu_{t+1}, S^2_{t+1})$:

$$\begin{cases} \dfrac{\widehat{\nu}_{t+1}\widehat{S^2}_{t+1}}{(\widehat{\nu}_{t+1}-2)} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\sigma^2}^{[i]}_{t+1} \\ \dfrac{2\widehat{\nu}^2_{t+1}\widehat{S^2}^2_{t+1}}{(\widehat{\nu}_{t+1}-2)^2(\widehat{\nu}_{t+1}-4)} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\sigma^2}^{[i]}_{t+1} - \sum_{j=1}^{K}\widehat{\sigma^2}^{[j]}_{t+1}\right)^2. \end{cases}$$

- For the relative nugget parameter υ^2_{t+1} which follows a inverse-gamma prior distribution $\mathcal{IG}(c_{t+1}, d_{t+1})$:

$$\begin{cases} \dfrac{\widehat{d}_{t+1}}{(\widehat{c}_{t+1}-1)} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\upsilon^2}^{[i]}_{t+1} \\ \dfrac{\widehat{d}^2_{t+1}}{(\widehat{c}_{t+1}-1)^2(\widehat{c}_{t+1}-2)} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\upsilon^2}^{[i]}_{t+1} - \sum_{j=1}^{K}\widehat{\upsilon^2}^{[j]}_{t+1}\right)^2. \end{cases}$$

- For the mean parameter β_{t+1} which follows a normal prior distribution $\mathcal{N}(e_{t+1}, V_{t+1})$:

$$\begin{cases} \widehat{e}_{t+1} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\beta}^{[i]}_{t+1} \\ \widehat{V}_{t+1} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\beta}^{[i]}_{t+1}(\widehat{\beta}^{[i]}_{t+1})' - \widehat{e}_{t+1}(\widehat{e}_{t+1})'\right) \end{cases}$$

The procedure is summarized in Figure 1.

4. Numerical application

Here, we consider an application of our procedure in a convection-diffusion modeling problem which occurs in fluid mechanics such as groundwater pollution problem or flow of oil from a well's reservoir, see Dehghan (2005); Efendiev and Durlofsky (2003) or Song and Wu (2010).

Let $\mathcal{D}$ be a square domain $(a,b)^2$ in $\mathbb{R}^2$ and $(0,T)$ be a time interval. Given a function $\mathcal{S} = \mathcal{S}(x,y,t) \in L^2(\mathcal{D}\times(0,T))$, we consider the two-dimensional time-dependent convection-diffusion equation with homogeneous Dirichlet boundary condition.

$$\begin{cases} L(u) &:= \frac{\partial u}{\partial t} - \nu\Delta u + \varphi_1\frac{\partial u}{\partial x} + \varphi_2\frac{\partial u}{\partial y} = \mathcal{S}, \text{in } \mathcal{D}\times(0,T] \\ u(x,y,t) &= 0 \text{ on } \partial\mathcal{D}\times(0,T] \\ u(x,y,0) &= u_0(x,y) \text{ in } \mathcal{D} \end{cases} \tag{11}$$

where u is the state variable to be modelized, φ_1 and φ_2 are constants which represent the convection coefficients, and ν is the positive diffusion coefficient. As in Song and Wu (2010), we assume that the above problem admits a unique smooth solution.

The term $\mathcal{S}$, called the source term, represents a source which continually bring a quantity of pollution in the domain $(a,b)^2$ in the problem of interest of groundwater pollution. The

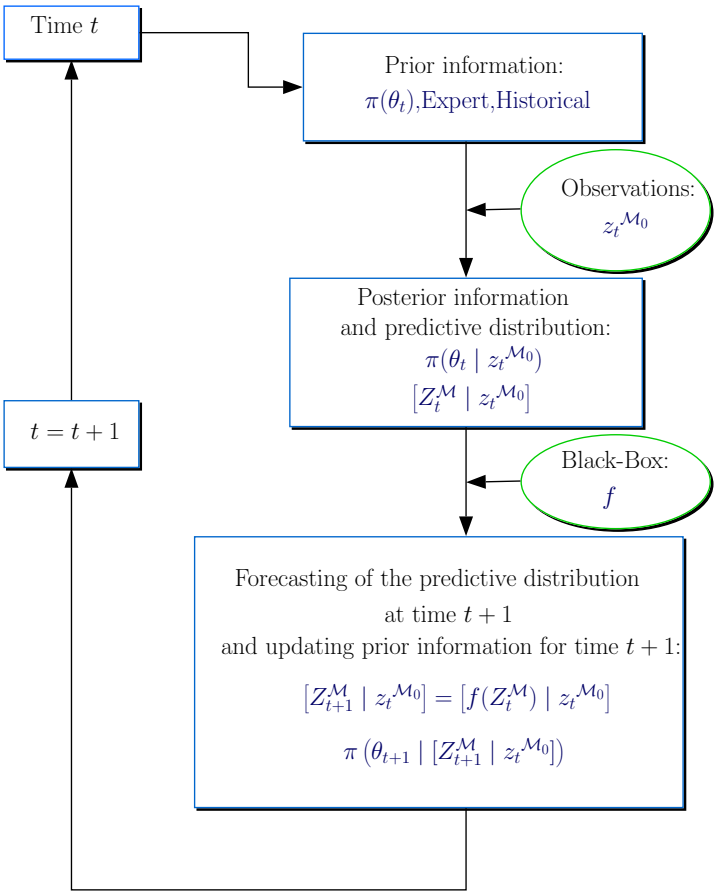

Figure 1: Spatio-temporal procedure

pollution's level due to $\mathcal{S}$ is not uniform in $(a,b)^2$ (higher in some parts of $(a,b)^2$ than in others). Once in the domain $(a,b)^2$, the quantity of pollution will change over time owing to the convection and diffusion phenomenon. Our aim is then to determine the state of the pollution, u, in $(a,b)^2$ at each time in view of Equation (11). The direction and velocity of pollution transport will depend on the values of φ_1 and φ_2, when the pollution's dispersion will depend on ν.

For our example, we assume that the pollution problem follows the PDE (11) with $(a,b)^2 = (0,1)^2$, $\varphi_1 = -0.09$, $\varphi_2 = -0.09$, $\nu = 0.001$, $\mathcal{S}(x,y,t) = (\frac{5}{4} - ((x-0.8)^2 + (y-\frac{1}{2})^2))$. We use a 24×24 mesh, say $\mathcal{M}$ to compute the values $u(x,y,t)$, $(x,y) \in \mathcal{M}$, and $t = 1, ..., 9$ which are displayed in Figure 3. We split the mesh points in two sets (see Figure 2): the set of observation sites, say $\mathcal{M}_0$, which correspond to the measurement points in real data applications and the set of sites of non-observed values.

Here, the black-box $f_{\varphi_1,\varphi_2,\nu}$ is defined by the same PDE (11) but the parameters $\varphi_1, \varphi_2, \nu$ are imprecisely known and considered as random with

$$\varphi_1, \varphi_2 \sim \mathcal{N}(-0.1, (0.01)^2) \text{ and } \nu \sim \mathcal{N}(0.006, (0.001)^2). \tag{12}$$

At the starting time, $t = 1$, we do a Bayesian kriging (Step 1) with the arbitrary values for the hyperparameters: $a_1 = 3$, $b_1 = 2$, $\nu_1 = 4$, $S_1^2 = 5$, $c_1 = 3$, $d_1 = 1$ and a flat prior for β_1 (see subsection 2.1). The Bayesian kriging was done using the package geoR of the R software, but an additional programming has been necessary to incorporate the priors for correlation and relative nugget parameters specified in (4) and in (6). At Step 2, we draw K maps $U_1^{[i]} = \{u^{[i]}(x, y, 1),\ (x, y) \in \mathcal{M}\}$ from the Bayesian kriging and K values for $\varphi_1^{[i]}, \varphi_2^{[i]}$ and $\nu^{[i]}$ according to (12), $i = 1, ..., K$, with $K = 10$. Then, we get 10 predictive maps $\widehat{U}_2^{[i]} = f_{\varphi_1^{[i]}, \varphi_2^{[i]}, \nu^{[i]}}(U_1^{[i]})$, $i = 1, ..., 10$. At Step 3, we use the maps $\widehat{U}_2^{[i]}$, $i = 1, .., 10$, to estimate the hyperparameters of the Bayesian kriging that will be done at time $t = 2$ when updated observations will be available.

Recursively, for $t = 2, \ldots, 9$, we follow steps 1 to 3, using the prior distribution obtained at the previous time.

At time t, after Step 2, we have a first prediction map of $u(x, y, t + 1)$ given by $\widehat{B}_{t+1} = \frac{1}{10} \sum_{i=1}^{10} \widehat{U}_{t+1}^{[i]}$ from the black-box. At time $t + 1$, after Step 1, we obtain a second prediction map $\widehat{u}_{t+1}$ obtained by the Bayesian kriging. The prediction errors of each method is evaluated resp. by $|\widehat{B}_{t+1}(x, y) - u(x, y, t + 1)|$ and $|\widehat{u}_{t+1}(x, y) - u(x, y, t + 1)|$. They are displayed in Figure 4 and 5, for $t = 2, \ldots, 9$.

These results show that our algorithm outperform the black-box prediction $\widehat{B}(x, y, t)$. Indeed, for any time $t = 2, \ldots, 9$, we see that the prediction errors of our algorithm are smaller or equal to those of the misspecified black-box in the domain $(0, 1)^2$. For example in the subdomain $(0, 0.6) \times (0, 0.2)$ the prediction errors of our algorithm are any time smaller than those of the misspecified black-box. The outperformance of our procedure means that for $t = 2, \ldots, 9$ the conditions (number of measurement sites, priors) required by Bayesian kriging are satisfactory to improve the prediction given by the misspecified black-box.

The code ran in approximately 72 minutes on a laptop computer (Processor: Intel(R) Core(TM) i5-4200 CPU @ 1.60GHz 2.30Ghz, RAM: 8Go). However, the duration can be largely reduced by parallelization of the code.

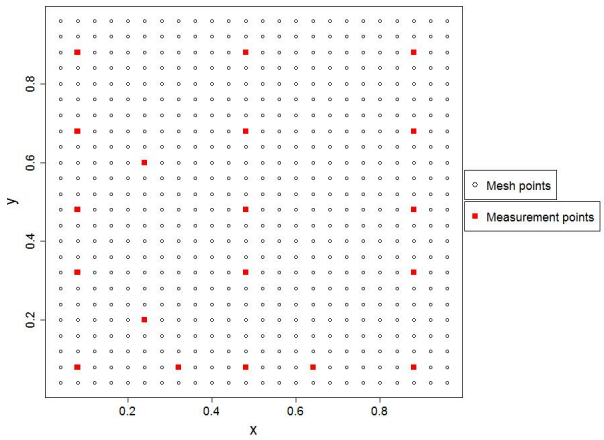

Figure 2: Mesh of domain $(0,1)^2$

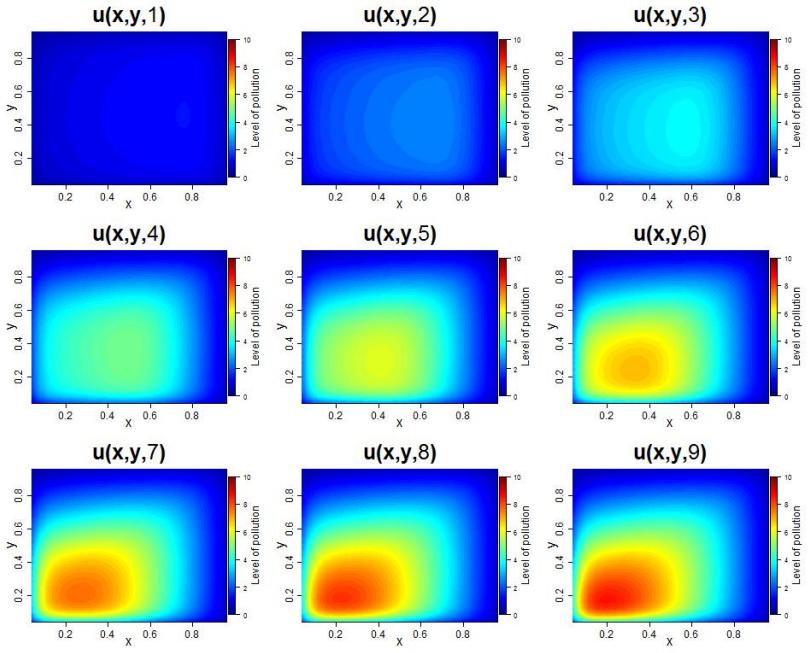

Figure 3: Evolution of real values of pollution to $t = 1$ at $t = 9$

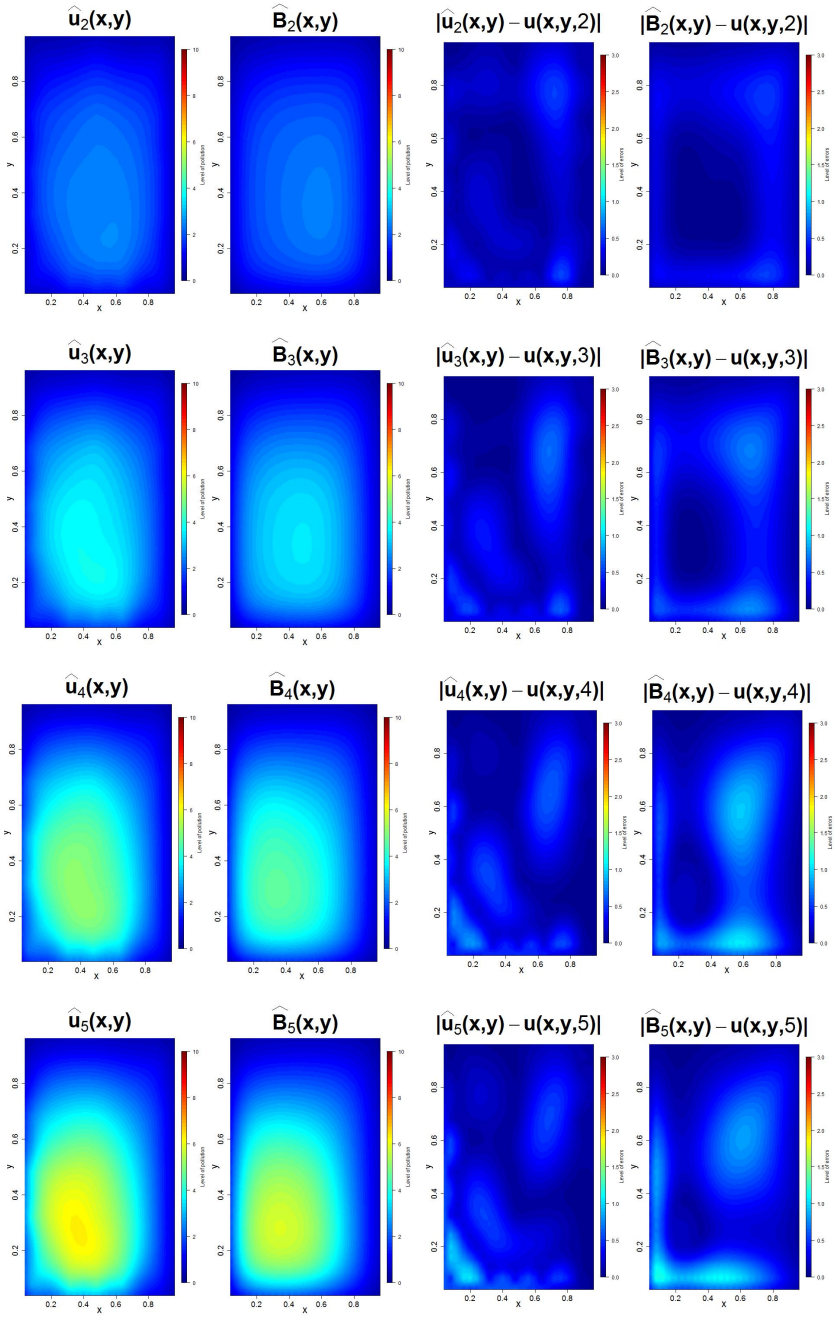

Figure 4: $\widehat{u}_t(x,y)$: Predicted values by our method, $\widehat{B}_t(x,y)$: Predicted values by misspecified black-box, $|\widehat{u}_t(x,y) - u(x,y,t)|$ predicted errors by our method, $|\widehat{B}_t(x,y) - u(x,y,t)|$ predicted errors by misspecified black-box, $t = 2, ..., 5$.

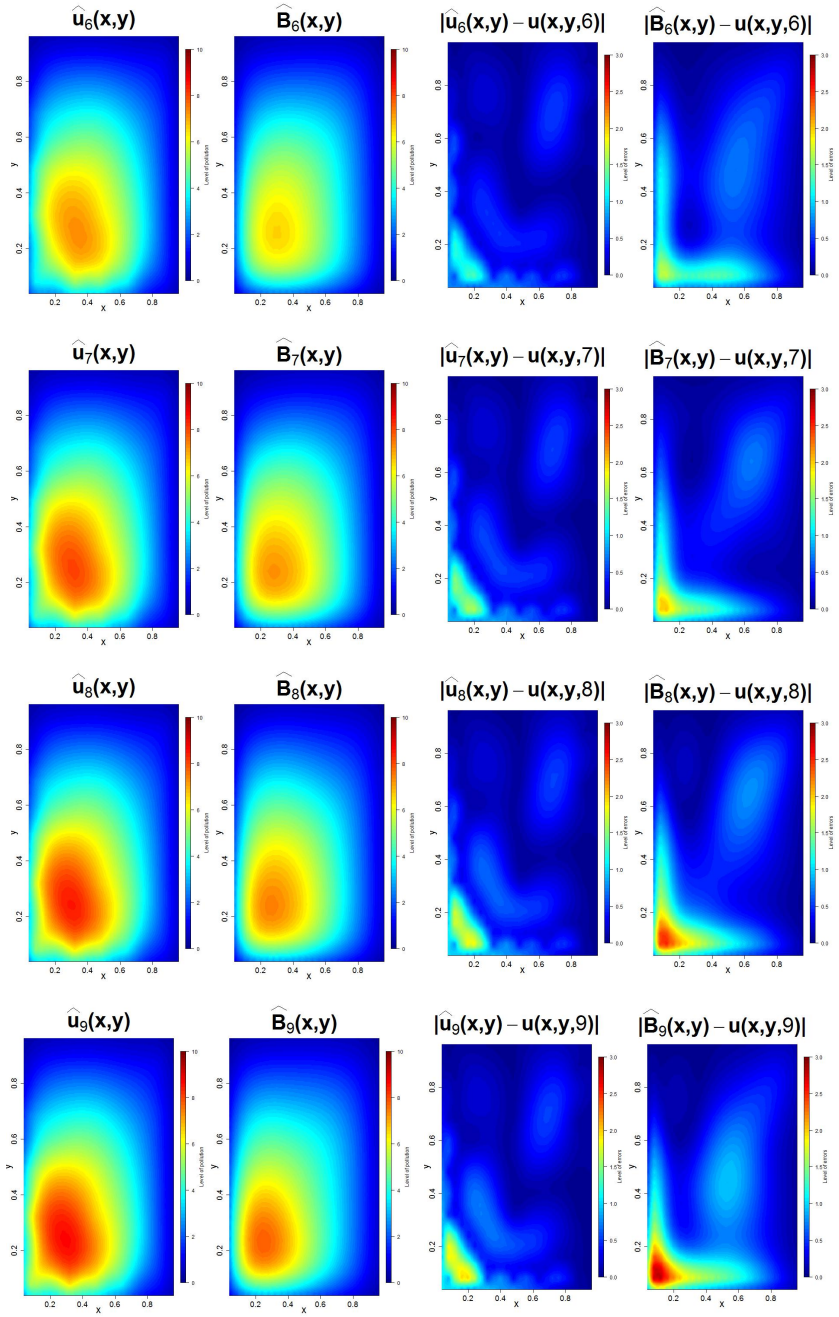

Figure 5: $\widehat{u}_t(x,y)$: Predicted values by our method, $\widehat{B}_t(x,y)$: Predicted values by misspecified black-box, $|\widehat{u}_t(x,y) - u(x,y,t)|$ predicted errors by our method, $|\widehat{B}_t(x,y) - u(x,y,t)|$ predicted errors by misspecified black-box, $t = 6, ..., 9$.

5. Conclusion

We have presented a spatio-temporal prediction algorithm which combines two kind of information: the information brought by the temporal model and the information brought

by the spatial kriging model. Both informations are fuzzy. The temporal model may be imperfect due to the imprecise knowledge of the inputs and/or because the black box represents a rough approximation of the underlying temporal model, inducing bias or extra-variability. On the other hand, the kriging model give accurate information locally around the observation sites but not on the whole region of interest. Our procedure use the temporal model to provide the prior information for the spatial model and therefore to combine two informations of different kind.

We think that this algorithm, or any variant, will find a wide scope of applicability in environmental science where time forward prediction and monitoring of spatio-temporal processes is a main activity, since it is easy to implement. The generality of the algorithm is due to the fact that the temporal evolution of the deterministic part of any physical phenomenon can be modelized by a black-box.

An application of our algorithm was done by using a convection-diffusion model which occurs in many domains of environmental sciences as problem of groundwater pollution. We used data simulated with a two-dimensional time-dependent convection-diffusion equation with homogeneous Dirichlet boundary condition. The black-box use the same equation but with misspecified coefficients. We have tested the performance of our method by comparing the prediction map obtained by the black-box at time t+1, before having the new data, with the one obtained by Bayesian kriging with updated prior once the data are observed.

Acknowledgments

This work has been funded in part by Auvergne Region and in by European Regional Development Fund.

References

Banerjee, S., Gelfand, A. E., Carlin, B. P., 2004. Hierarchical modeling and analysis for spatial data. Crc Press.

Bernardo, J.-M., 1979. Reference posterior distributions for Bayesian inference. J. Roy. Statist. Soc. Ser. B 41 (2), 113–147, with discussion.

Bioche, C., Druilhet, P., 2015. Approximation of improper priors. Bernoulli To appear.

Cressie, N., Wikle, C. K., 2006. Space-Time Kalman Filter. John Wiley & Sons, Ltd.

Cressie, N., Wikle, C. K., 2011. Statistics for spatio-temporal data. John Wiley & Sons.

Dehghan, M., 2005. On the numerical solution of the one-dimensional convection-diffusion equation. Mathematical Problems in Engineering 2005 (1), 61–74.

Diggle, P. J., Ribeiro, P. J., 2002. Bayesian inference in gaussian model-based geostatistics. Geographical and Environmental Modelling 6 (2), 129–146.

Diggle, P. J., Ribeiro, P. J., 2007. Model-based geostatistics. Springer Series in Statistics. Springer, New York.

Efendiev, Y., Durlofsky, L., 2003. A generalized convection-diffusion model for subgrid transport in porous media. Multiscale Modeling & Simulation 1 (3), 504–526.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., 2014. Bayesian data analysis, 3rd Edition. Texts in Statistical Science Series. CRC Press, Boca Raton, FL.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. Journal of the American Statistical Association 105 (491).

Hengl, T., Heuvelink, G. B., Tadić, M. P., Pebesma, E. J., 2012. Spatio-temporal prediction of daily temperatures using time-series of modis lst images. Theoretical and applied climatology 107 (1-2), 265–277.

Ip, R. H., Li, W., 2015. Time varying spatio-temporal covariance models. Spatial Statistics To appear.

Irwin, M. E., Cressie, N., Johannesson, G., 2002. Spatial-temporal nonlinear filtering based on hierarchical statistical models. Test 11 (2), 249–280.

Le Gratiet, L., 2013. Bayesian analysis of hierarchical multifidelity codes. SIAM/ASA J. Uncertain. Quantif. 1 (1), 244–269.

Song, L.-J., Wu, Y.-J., 2010. A modified Crank-Nicolson scheme with incremental unknowns for convection dominated diffusion equations. Appl. Math. Comput. 215 (9), 3293–3301.

Stein, M. L., 2005. Space-time covariance functions. J. Amer. Statist. Assoc. 100 (469), 310–321.