Which relations between deterministic simulations and observations?

Chantal de Fouquet[a], Nicolas Flipo[a], Laurent Létinois[b], Laure Malherbe[b], Edwige Polus-Lefebvre[a,c], Michel Poulin[a], Antony Ung[b].

a: Mines ParisTech, Centre de Géosciences, Fontainebleau, France.

b: INERIS, Direction des Risques Chroniques, Verneuil-en-Halatte, France.

c: now at Ram-Ses, Les Isnes, Belgique.

ABSTRACT

The state of environment (air, rivers and groundwater, etc.) is described by observations on one hand and by deterministic simulations based on the physics and chemistry and/or on the biology of complex phenomenon on the other hand, with results that usually differ. Generally the deterministic simulation is less variable than observations, but the differences cannot be explained by the only differences of support between observations (spatial "point" values) and simulation results (representing rather averaged quantities on the grid).

In order to enhance the simulation to match the observations, a "simple" bivariate model consists in splitting the studied variable Y as the sum of the deterministic simulation S and a correction term R which is supposed to be not correlated (spatially or temporally) with S: $Y = S + R$. The observations Z differ from Y by a measurement error term. Within this model, the estimation of Y from the observations Z can be reduced to the kriging of the residual R from the « innovations » $Z - S$ at observation points.

Joint exploratory analysis of observations and results of deterministic simulations shows that this bivariate model does not always suit to the data. Innovations appear to be correlated with the simulation S. In order to take such correlations into account, Chilès, Séguret et al. (2008) proposed an intrinsic correlation model between the variable Y and the deterministic simulation S.

This intrinsic correlation model is generalized here to the linear model of co-regionalization. Examples are presented in air and river quality modeling. Consequences for the estimation are examined.

Key words: deterministic model, observations, estimation, simulation, air quality, water quality.

INTRODUCTION

The state of environment (air, rivers and groundwater...) is described with measurements (or « observations »), which are made continuously or episodically during campaigns, and with results of deterministic spatio-temporal simulations, which are based on the physics and chemistry or on the biology of complex phenomena. But observations differ from deterministic simulation results. Those differences are partly explained by the difference of support between observations, which can be considered as spatially quasi « point » values, and the deterministic simulation results, which represent rather averages on the discretization cells. But the difference of support is not sufficient to explain all the observed

divergences: mean deviations appear locally and the spatial, temporal or spatio-temporal variability is different for the simulations and the observations. Generally the simulation results appear to be less variable than observations, but the contrary can also occur. In addition to measurement errors, the imperfect description of a real phenomenon is then in question.

Figure 1a presents the superimposition of time series of hourly concentrations of dissolved Oxygen simulated by the deterministic model ProSe (Even, Poulin et al., 1998; Even, Mouchel et al., 2007; Flipo, Even et al., 2004; Flipo, Rabouille et al., 2007) for the year 2003, and the continuous automatic measurements made during a part of the year (summer) at a measurement station on the Seine River (France). A systematic deviation is visible, which is more or less important according to time. The two time series do not present the same time variability, specifically at small time intervals (Figure 1b).

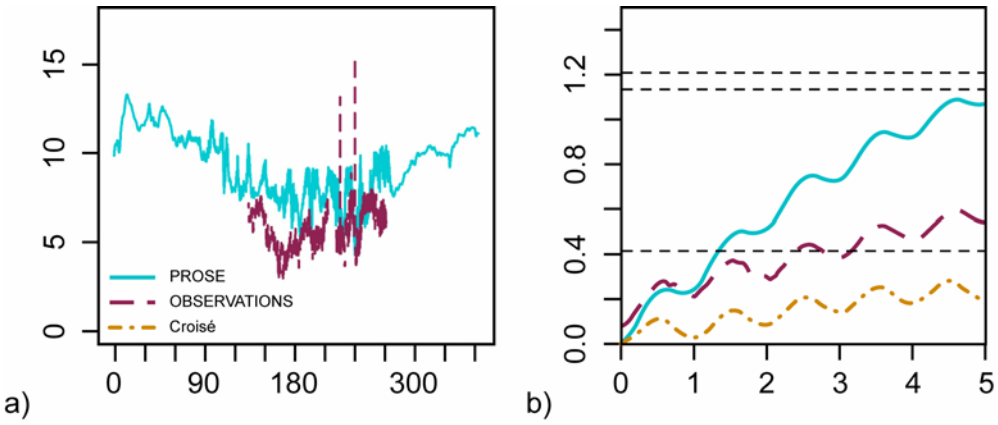

Figure 1. Time series of hourly concentration of dissolved Oxygen at the Sartrouville station on the Seine. a) Superposition of time series. ProSe model (all the 2003 year) and automatic continuous measurements in summer. In abscissa, time unit is the day number, with origin at 1er January 2003; in ordinate, measured concentrations expressed in mg O_2 per liter. From Polus-Lefèbvre (2010); b) sample simple and cross variograms, calculated on common time field (same data date), with visible day periodicity. The nugget component is visible on the observation variogram.

Let *Y* denote the studied variable (a concentration for example), and *S* the deterministic simulation on a uniform support (a regular grid). As the simulation *S* is generally less variable than the observations, a very simple and common model consists in adding a corrective term to the deterministic simulation (for example Blond, Bel et al., 2003):

$$Y = S + R \quad (Eq.\ 1)$$

The observations *Z* are supposed to contain a measurement error ε :

$$Z = Y + \varepsilon \quad (Eq.\ 2)$$

The corrective term R is then estimated from the differences or « innovations » *Z*-*S* at measurement sites or dates, and the concentration *Y* is estimated by

$$Y^* = S + R^* \quad (Eq.\ 3)$$

When the simulation is known everywhere (which is supposed throughout this article) this model is justified in the absence of (spatial or temporal) correlation between the simulation *S* and the « residual » *R*. Indeed, Eq. 3 then amounts to factorizing *Y* on *S* and *R* and the cokriging of *Y* by *Z* and *S* corresponds to separated kriging of each term. As *S* is supposed to be known everywhere, it is equal to its kriging *S**.

Chilès, Séguret et al. (2008) have shown that the hypothesis of absence of correlation between the innovation *Z*-*S* and the simulation *S* is not always verified. In the case they studied, the intrinsic correlation model better describes the relations between the observations and the simulation.

This intrinsic correlation model between observations and deterministic simulation is hereafter generalized to the linear co-regionalization model (for example Chilès and Delfiner, 2012). Examples are given and consequences on estimation are examined.

I. CORRELATION BETWEEN DETERMINISTIC SIMULATION AND OBSERVATIONS

I.1. Intrinsic Correlation

The presentation is made in the spatial context, but the application in temporal or spatio-temporal context is immediate. Chilès, Séguret et al. (2008) proposed the following model:

(i) the main variable *Y* and the simulation *S* are intrinsically correlated, i.e. their simple and cross variograms are proportional:

$$\gamma_Y = \omega^2 \gamma_S \text{ and } \gamma_{YS} = \omega \rho \gamma_S \quad \textit{(Eq. 4)}$$

(ii) the observations are affected by a measurement error ε, with an expectation equal to zero and with a variance σ_ε^2 supposed to be constant (as a simplification), cf. Eq. 2. These measurement errors are supposed to be spatially not correlated with *Y*. They induce a nugget component with amplitude σ_ε^2 on the sample variogram of the observations:

$$\forall |h| > 0,\ \gamma_Z(h) = \sigma_\varepsilon^2 + \gamma_Y(h) \quad \textit{(Eq. 5)}$$

This model describes a large variety of situations according to the value of ω and ρ:

- the simulation *S* is more or less variable than Y according to ω is lower or greater than 1 ;
- if *S* is (in the usual geostatistical sense) a non conditional simulation of *Y*, then ω =1 and ρ=0 ;
- the cross variogram γ_{YS} is located between the two simple variograms (or their opposite) when $|\rho| \geq \min\left(\omega, \frac{1}{\omega}\right)$: let's note that this situation does not necessarily indicate a strong link between *Y* and *S*, this inequality being verified as soon as ω is "large" (much greater than 1) or very small (much lower than 1).

As the measurement error is supposed to be not correlated with *Y*, the observations *Z* and the inaccessible variable *Y* have the same cross variogram with the simulation S:

$$\gamma_{ZS}(h) = \gamma_{YS}(h) \quad \textit{(Eq. 6)}$$

Interpretation

By analogy with Eq. 1, the simulation *S* is taken as the reference variable, because it is supposed to be known at each grid node. Then the intrinsic correlation model can be written (Chilès, Séguret et al., 2008):

$$Y(x) = \omega \rho S(x) + \omega \sqrt{1-\rho^2} R(x) \quad \textit{(Eq. 7)}$$

where *S* and *R* are spatially not correlated and with the same variogram.

The coefficient ω^2 is equal to the ratio γ_{Y}/γ_{S}. The simulation *S* proportionally mimics the amplitude of the fluctuations of the actual phenomenon Y with a ratio $1/\omega$.

The coefficient ρ represents the correlation between *Y(x)* and *S(x)* (at the same point). This correlation is « intrinsic » (Rivoirard, 2012), i.e. it remains unchanged for any regularized (or for any linear combination) provided the regularization support (resp. the linear combination) is identical for *S* and *Y;* ρ neither depends on the field.

Following Eq. 4, the intrinsic correlation model *Y* and *S* can also be written similarly to Eq. 7, factorizing *S* on *Y* and a residual *Q* spatially not correlated with *Y* but with the same variogram:

$$S(x) = \xi \rho Y(x) + \xi \sqrt{1-\rho^2} Q(x) \text{ , with } \xi = \frac{1}{\omega} \qquad \textit{(Eq. 8)}$$

The variable *Y* is chosen as the reference in Chilès, Séguret et al. (2008) and de Fouquet, Malherbe et al. (2011). Eq. 8 (used in Polus, Flipo et al., 2011) is more convenient in order to write the estimation of *Y* from the observations and the deterministic simulation *S,* which is supposed to be known everywhere.

Determination of the model parameters

In the absence of nugget effect on the variogram of the simulation, and as the measurement errors are supposed to be spatially not correlated with *Y* and *S*, the model parameters are determined as follows:

- ω^2 (and then its positive root ω) is given by the ratio between the regular part of γ_Z and γ_S ;
- σ_ε^2 corresponds to the nugget component on the variogram of the observations and γ_Y to its regular part ;
- ρ is given by the ratio between the cross variogram γ_{ZS} and $\omega.\gamma_S$.

If a nugget effect is present on the variogram of the simulation (and thus also on γ_Y),

- ω^2 (and then its positive root ω) is given by the ratio between the regular part of γ_Z and γ_S ;
- the variance of the nugget component of *Y* is equal to the variance of the nugget component of *S* multiplicated by ω^2 ;
- the difference between the global nugget effect on γ_Z and the nugget effect of γ_Y calculated above, gives the variance of measurement errors σ_ε^2 ;
- the correlation coefficient ρ between *Y* and *S* is deduced from the following calculation :

$$\rho_{ZS} = {Cov(Z,S)}/{\sigma_Z.\sigma_S}$$
$$= {Cov(Y,S)}/{\sigma_Z\,\sigma_S}$$
$$= \omega\rho\frac{\sigma_S}{\sigma_Z}$$

Then, $$\rho = \rho_{ZS}\frac{\sigma_Z}{\omega.\sigma_S}$$

Verification of the pertinence of the intrinsic correlation model for the studied case

The pertinence of the intrinsic correlation model is verified on the difference *(*or « innovation »*) D=Z-S* between observation and simulation:

$$Z - S = \varepsilon + (\omega\rho - 1)S + \omega\sqrt{1-\rho^2}R$$

If the model is adequate, then the sample variogram of the « innovations » is fitted by its « theoretical variogram » (de Fouquet, Malherbe et al., 2011, with adapted notations)

$$\gamma_D(h) = \sigma_\varepsilon^2 + \left((1-\rho\omega)^2 + \omega^2(1-\rho^2)\right)\gamma_S(h) \qquad \textit{(Eq. 9)}$$

More simply, since ωS has the same spatial variability as *Y,* the variogram of the difference $H = Z - \omega S$ between observations and "rescaled" simulation is

$$\gamma_H = \sigma_\varepsilon^2 + 2(1-\rho)\gamma_Y \qquad \textit{(Eq. 10)}$$

and the cross variograms between *H* and observations *Z* or "rescaled simulation" are respectively:

$$\gamma_{HZ} = \sigma_\varepsilon^2 + (1-\rho)\gamma_Y \text{ and } \gamma_{H\ \omega S} = -(1-\rho)\gamma_Y \qquad \textit{(Eq. 11)}$$

Eq. 10 and 11 can be also used to check the validity of the model.

Logically, when the correlation ρ between *Y* and *S* is equal to 1, *H* corresponds to the measurement error (cf. Eq. 11).

I.2. Generalization: linear model of co-regionalization

Even if it is very flexible, the above model does not permit to describe complex relations between observations and deterministic simulation. Numerous phenomena appear rather as the sum of components corresponding to different spatial or temporal variability scales. In urban or industrial environment, time series of concentrations of diverse substances appear as the superposition of one or more components with stationary or linear variogram, and of periodic components with daily period, weekly period (reflecting namely the economical or scholar rhythms and the week-end) and annual period. All those different components are not necessarily present on the deterministic simulation and the ratio between the amplitude of the real fluctuations and the simulated fluctuations vary according to the components.

The linear co-regionalization model between *Y* and the deterministic simulation *S* describes wider situations than the intrinsic correlation model. Taking the deterministic

simulation *S* as the reference, because it is known everywhere on the grid, the linear co-regionalization model is written

$$\gamma_S = \sum_u \gamma_u + \sum_{u'} \gamma_{u'} \ , \ \gamma_Y = \sum_u \omega_u^2 \gamma_u + \sum_{u''} \gamma_{u''}, \quad \gamma_{YS} = \sum_u \omega_u \rho_u \gamma_u \qquad \textit{(Eq. 12)}$$

where the *u* denotes the structures common to *Y and S*, *u'* the structures possibly present on the simulation *S* but not on *Y,* and *u''* the structures possibly present on *Y* but not on *S*.

In this model the variables are decomposed in different structures. In the bivariate case, this decomposition can be written:

$$S = \sum_u S_u + \sum_{u'} S_{u'} \text{ and } Y = \sum_u \omega_u \left(\rho_u S_u + \sqrt{1-\rho_u^2} R_u \right) + \sum_{u''} Y_{u''} \qquad \textit{(Eq. 13)}$$

or (cf. Eq. 7 and 8)

$$Y = \sum_u Y_u + \sum_{u''} Y_{u''}, \ S(x) = \sum_u \xi_u \left(\rho_u Y_u(x) + \sqrt{1-\rho^2} Q_u(x) \right) + \sum_{u'} S_{u'} \text{ with } \xi_u = \frac{1}{\omega_u} \quad \textit{(Eq. 14)}$$

Eq. 12 means that *S* reproduces different components of the true system, but with different mistakes on their amplitude and with different correlation with real fluctuations. A component S_u for which the correlation coefficient ρ_u is zero behaves as a « non conditional simulation », in the geostatistical sense, of the true homologous component R_u. If the determination coefficient ρ_u^2 is strictly lower than 1, then the deterministic simulation contains a component Q_u corresponding (up to a factor) to a « geostatistical non conditional simulation » of the true component Y_u.

The difference *D=Z-S* can be used to control the pertinence of the model. As

$$Z - S = \varepsilon + \sum_u \left(\omega_u \rho_u - 1 \right) S_u + \sum_u \omega_u \sqrt{1-\rho_u^2} R_u - \sum_{u'} S_{u'} + \sum_{u''} Y_{u''} \qquad \textit{(Eq. 15)}$$

or

$$Z - S = \varepsilon + \sum_u \left(1 - \xi_u \rho_u \right) Y_u - \sum_u \xi_u \sqrt{1-\rho^2} Q_u - \sum_{u'} S_{u'} + \sum_{u''} Y_{u''} \qquad \textit{(Eq. 16)}$$

the theoretical variogram of the difference is

$$\gamma_D = \sigma_\varepsilon^2 + \sum_u \left(\left(1 - \rho_u \omega_u \right)^2 + \omega_u{}^2 (1 - \rho_u{}^2) \right) \gamma_u + \sum_{u'} \gamma_{u'} + \sum_{u''} \gamma_{u''} \qquad \textit{(Eq. 17)}$$

The linear co-regionalization model differs from the intrinsic correlation in the following cases:

- the ratio ω_u varies according to the structures (the ratio γ_Y / γ_S is not constant);

- the correlation coefficient ρ_u varies according to the structures (le ratio between the cross variogram γ_{YS} and at least one of the simple variograms γ_Y or γ_S is not constant);

- some structures are present only on *Y* or *S*.

II. EXAMPLES

The linear co-regionalization model between the studied variable *Y* and its deterministic simulation *S* is able to describe various situations. In a temporal context, an example for atmospheric pollution on national scale (France) is presented in de Fouquet, Malherbe et al. (2011); an example for ecological modeling of water quality on a part of the Seine hydrographical network is presented in Polus, Flipo et al. (2011). In the two cases, the measurement error on observations is taken into account.

Those publications are very briefly synthesized and new additional results are presented.

II.1. Atmospheric Pollution: CHIMERE model and observations

The links between the deterministic CHIMERE simulation and summer hourly ozone measured concentrations are very well described with the linear co-regionalization model for metropolitan France in 2005 and 2006 for most of the fixed measurement stations. For a temporal window of a few days, the temporal variograms show a daily periodic component and an exponential component with range between two and three days. The parameters of the bivariate model between observations and CHIMERE were automatically fitted for each station and each year. The pertinence of the bivariate model was controlled on the variogram of the innovations.

The Principal Component Analysis of fitted parameters ρ_u and ω_u indicates a high similarity between 2005 and 2006 for the periodic component. For the exponential component, the correlation coefficient and the amplitude correction of 2005 and 2006 are less similar. Lastly, the variogram analysis and the mapping of fitted parameters are very useful to identify which improvements are needed on the deterministic CHIMERE simulation.

Spatially, the CHIMERE simulation (on a grid with about 10km cells) appears to be systematically less variable than the observations. In the example of Figure 2 the cross variogram between the observations and CHIMERE indicates a poor spatial correlation. The variogram of the innovations is here greater than the variogram of the observations. Indeed, it includes the Q_u components (Eq. 16) corresponding to geostatistical « non conditional simulations » (Eq. 14). In addition, the nugget component which reflects the measurement errors (or a small range structure that is not reproduced by CHIMERE) is large.

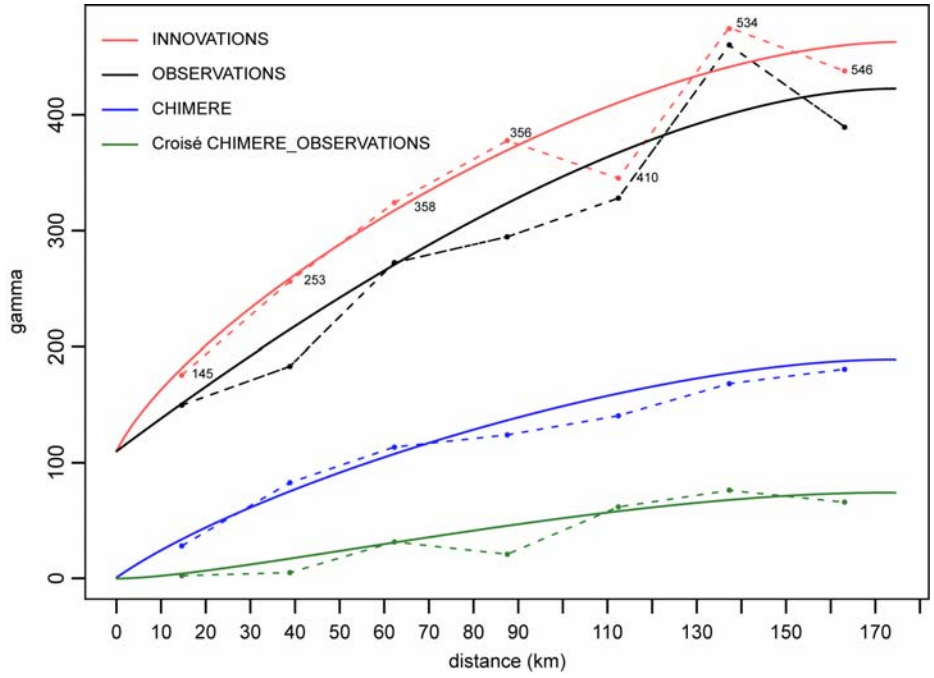

Figure 2. Automatic fitting of mean spatial variograms of hourly Ozone concentrations (here, the 27 July 2006, 9hUT) with linear co-regionalization model. The upper variogram is the one of innovations. The cross variogram between observations and CHIMERE increases proportionally less quickly near the origin than the CHIMERE variogram, showing that a bivariate model of residual type is unsuited.

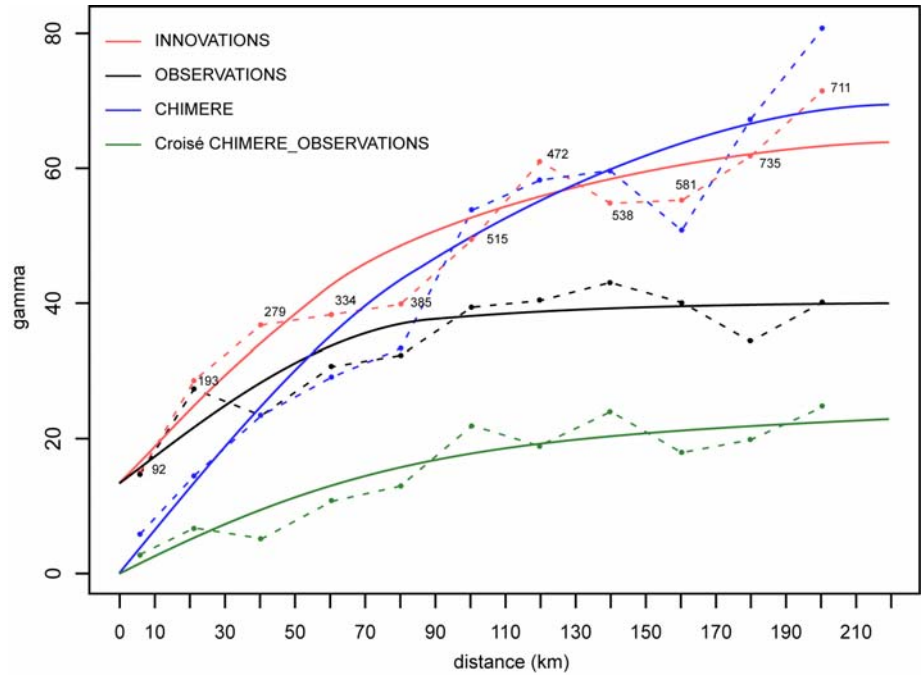

Figure 3. Automatic fitting of spatial variograms of particles PM10 daily concentrations (here, the 3 may 2009) with linear co-regionalization model. The CHIMERE variogram intersects the variograms of observations and innovations respectively.

Lastly, for another variable, the « PM10 » particles, the situation described by the linear co-regionalization model appears to be more complex. In the example of Figure 3 CHIMERE simulation highly exaggerates the amplitude of the large range structure, which is however correlated with the observations. The variogram of CHIMERE intersects the variograms of the observations and innovations respectively. In this example too, the variogram of the innovations is greater than the variogram of the observations, as an effect of the components of CHIMERE which are spatially not correlated with the observations.

II.2. Rivers: ProSe model and observations

Measurements of nutrients and dissolved oxygen concentrations in the Seine river and its main affluent, the Marne river, are compared with the results of the hydro-ecological model ProSe (Even, Poulin et al., 1998; Even, Mouchel et al., 2007; Flipo, Even et al., 2004; Flipo, Rabouille et al.,2007). Polus-Lefèbvre (2010) has examined how to improve the spatial and temporal estimation of concentrations from measurements using the deterministic simulation ProSe as co-variable. For nitrates Polus, Flipo et al. (2011) have shown the pertinence of the linear co-regionalization model between observations and deterministic simulations on the measurement time series at the water control stations. The spatial evolution of the fitted parameters (ω_u, ρ_u) makes it possible to clarify the system behavior and the imperfections of the model.

For dissolved oxygen, the linear co-regionalization model fits very well the temporal simple and cross variograms of « instantaneous » concentrations time series at measurement stations (Figure 4), whereas spatial variograms are fitted within the intrinsic correlation model (Figure 5).

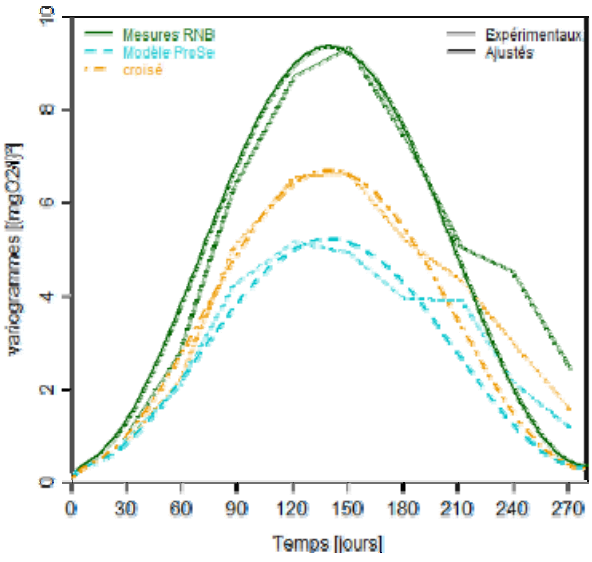

Figure 4. Simple and cross temporal variograms of instantaneous dissolved oxygen at one measurement station (n° 15) on the Seine, fitted within the linear co-regionalization model.

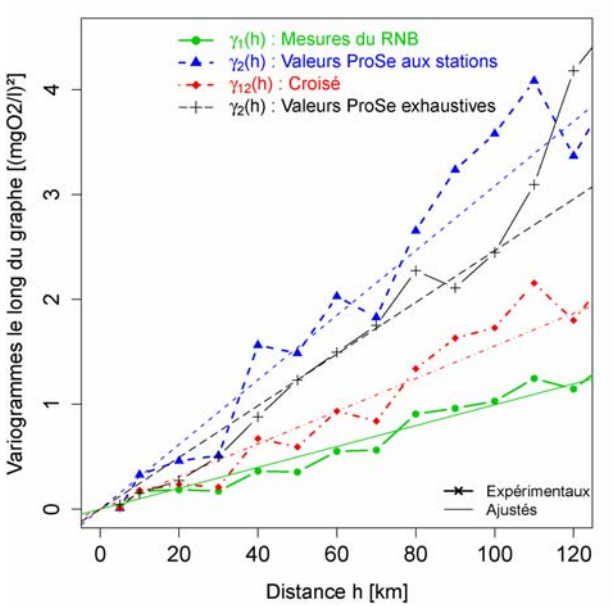

Figure 5. Dissolved oxygen concentrations along the Seine. Fitting of the global simple and cross variograms within the intrinsic correlation model.

III. ESTIMATION OF THE MAIN VARIABLE FROM OBSERVATIONS AND DETERMINISTIC SIMULATION

The deterministic simulation *S* is generally available at all a grid cells, whereas the measurements are made (continuously or episodically) at some sites. In this part, the deterministic simulation is supposed to be known at the measurement sites too. The associated estimation (or interpolation) incertitude will be neglected. The measurement errors are also neglected in a first step.

Let *(Y, S)* be a model with residual in which the deterministic simulation *S* is the master variable, to which *Y* is subordinate (Rivoirard, 2001 and 2012)*:* their cross variogram is proportional to γ_S. In this case, when *S* is known at all *Y* data points (and at all estimation grid cells), the cokriging of *Y* from *Y* and *S* data is coincides with the separated kriging of the residual. This case corresponds to the correction of the deterministic simulation as in Eq. 3.

The model with residual corresponds to the following particular cases:

- *Y* and *S* are intrinsically correlated (the variogram of the residual and that of the master variable are proportional) ;

- *Y* and *S* are linked by a linear co-regionalization model, in which all components present in *S* are present in *Y* (the set of indexes *u'* is empty), and such that $\rho_u \omega_u$ is constant for all the common structures indexed by *u***.** Indeed, as the γ_u constitute a free system, ones have $\gamma_{YS} = A\gamma_S$ with $\gamma_S = \sum_u \gamma_u$ and $\gamma_{YS} = \sum_u \omega_u \rho_u \gamma_u$.

External drift kriging corresponds to the linear regression of *Y* on *S* at neighborhood scale, in which the (spatial, temporal or spatio-temporal) correlation of the residuals is taken into account (Rivoirard, 2002). As the coefficients of the regression are not known, the variogram of the residuals can be indirectly fitted, comparing the theoretical variance of

authorized linear combinations (which “filter” the unknown coefficients of the linear regression) with the associated sample quadratic value (Chilès and Delfiner, 2012). In this estimation, the amplitude correction on the deterministic simulation *S* is identical for all components, via the coefficients of the linear regression: $Y^{KDE} = \alpha^{KDE} S + \beta^{KDE} + R^{KDE}$.

In the other cases *Y* (or the residual *R*) should be estimated by cokriging from the observations and the deterministic simulation *S* values, in order to take into account the amplitude ratio and the correlation coefficient which depends on the structural components.

The above results remain valid in presence of a nugget component for measurement errors on the observations.

Link with sequential data assimilation

Sequential data assimilation methods (for a synthetic presentation see for example Wackernagel, 2004) present high analogies with Eq. 3. Let Y_{n-1}^{a} be the best estimator of the system state at instant t_{n-1}. The system state at instant t_n is forecast by simulation: $Y_n^f = \mathbf{F}_n Y_{n-1}^a$. The « analysis » step, i.e. the correction of the forecast from the observations Z_n at step *n* is written

$$Y_n^a = Y_n^f + \mathbf{K}_n \left(Z_n - \mathbf{H} Y_n^f \right)$$

where the matrix **H** makes the link between the observations and the system state. The corrective term $\mathbf{K}_n \left(Z_n - \mathbf{H} Y_n^f \right)$ corresponds to the estimation of the forecast error $Y_n^a - Y_n^f$ from observed residuals. In the absence of spatial correlation between forecast Y_n^f and observed residuals, this linear correction is analogous to kriging of residual (Chilès and Delfiner, 2012).

When the observations are numerous enough, it would be interesting to control the correlation between predicted and observed residuals, in order to improve the “analysis” step, thanks to a cokriging from the forecasts and the observations.

CONCLUSION

The linear co-regionalization model can be used to describe a great variety of situations in order to characterize the spatial or temporal link between a deterministic simulation and observations. The examination of the fitted parameters of the bivariate model is very effective in analyzing complex systems, and detecting and quantifying the imperfections of the deterministic simulations.

The bivariate model makes it possible to estimate the studied variable from the observations, using the deterministic simulation as a covariate. This can be used to make a deterministic simulation fit the observations, or to fill the gaps of a time series.

In the presented examples, the difference of spatial support between deterministic simulation and observations is not explicitly examined. The modeling of the links between deterministic simulations and observations can thus be improved in order to better take the difference of support into account.

The presented examples are either in a temporal or in a spatial context. For a spatio-temporal modeling, it can be necessary to introduce delayed correlation in order to take into account approximation of the displacement velocity on deterministic forecasts (for air, rivers or groundwater).

ACKNOWLEDGMENTS

Authors thank Philippe Le Caër for graphic help.

BIBLIOGRAPHY

Blond N., Bel L., Vautard R. 2003. Three-dimensional ozone data analysis with an air quality model over the Paris area. Journal of geophysical research 108(D23), 4744.

Chilès J.-P., Delfiner P. 2012. Geostatistics: modelling spatial uncertainty. 2nd edition. N.Y. Wiley.

Chilès J.-P., Séguret S., Riboud P.-M., 2008. Geostatistical analysis of validation data of an air pollution simulator. VIII International Geostatistics Congress, GEOSTATS 2008, Santiago Chile - 1-5 December 2008 - Vol. 2 - pp 861-870.

Even S., Poulin M., Garnier J., Billen G., Servais P., Chesterikoff A. et al. 1998. River ecosystem modelling. Applications of the PROSE model to the Seine river (France), Hydrobiologia, 373/374, p. 27–45.

Even S., Mouchel J.-M., Servais P., Flipo N., Poulin M., Blanc S., Chabanel M., Paffoni C. 2007. Modeling the impacts of combined sewer overflows on the river Seine water quality. Science of the Total Environment 375(1–3): 140–151.

Flipo N., Even S., Poulin M., Tusseau-Vuillemin M.-H., Améziane T. and Dauta A. 2004. Biogeochemical modelling at the river scale: plankton and periphyton dynamics - Grand Morin case study, France. Ecological Modelling, 176,. 333–47.

Flipo N., Rabouille C., Poulin M., Even S., Tusseau-Vuillemin M.-H., Lalande M. 2007. Primary production in headwater streams of the Seine basin: the Grand Morin case study. Science of the Total Environment 375:98–109.

de Fouquet C., Malherbe L., Ung A. 2011. Geostatistical analysis of the temporal variability of ozone concentrations. Comparison between CHIMERE model and surface observations. Atm. Env., 45(20) 3343-3514

Polus-Lefèbvre E. 2010. Apports croisés de la modélisation géostatistique and déterministe. Exemple des nitrates and de l'oxygène dissous dans un réseau hydrographique. Thèse de Doctorat, Ecole Nationale Supérieure des mines de Paris (Mines ParisTech).

Polus E., de Fouquet C., Flipo N., Poulin M. 2010. Caractérisation spatiale et temporelle des « Masses d'Eau Cours d'Eau ». Revue des sciences de l'eau. 23(4) 415-429

Polus E., Flipo N., de Fouquet C., Poulin M. 2011. Geostatistics for assessing the efficiency of distributed physically-based water quality model: application to nitrates in the Seine River. Hydrological Processes. 25(2) 217-233

Rivoirard, J. 2001. Which models for collocated cokriging? Mathematical Geology, 33(2) 117-131.

Rivoirard, J. 2002. On the structural link between variables in kriging with external drift. Mathematical Geology, 34(7) 797-808.

Rivoirard J. 2012. De l'analyse structurale à la modélisation des ressources naturelles : contributions géostatistiques. Mémoire d'Habilitation à Diriger des Recherches, Académie de Paris, Université Pierre-et-Marie Curie.

Wackernagel H. 2004. Géostatistique et assimilation séquentielle de données. Mémoire des Sciences de la Terre n°2004-02, Académie de Paris, Université Pierre-et-Marie Curie.