

Reconstruction des conditions limites amont de PROSE par co-krigeage

Lauriane Vilmin^{1,*}, Chantal de Fouquet¹, Nicolas Flipo¹, Michel Poulin¹

¹ Centre de Géosciences, Mines ParisTech, 35 rue Saint-Honoré, 77 305 Fontainebleau

* lauriane.vilmin@mines-paristech.fr

Introduction

Afin de simuler les processus hydro-écologiques en rivière, il est important de s'assurer de la cohérence entre les données employées comme conditions limite et celles utilisées pour la calibration du modèle le long du linéaire étudié. Si ces deux jeux de données ne sont pas en accord, des erreurs liées à la mesure de la condition limite amont peuvent se répercuter tout au long du linéaire simulé, même si le modèle reproduit correctement les variations de concentrations aux différentes échelles de temps.

Dans le cas des concentrations en nitrates, [Polus et al. \(2011\)](#) a montré que le simple fait de mettre en accord les conditions limite amont du modèle PROSE et les données de calibration (aux stations SIAAP) permet de réduire le biais de façon significative (celui-ci passe de 0.274 à 0.038 mgN/L sur l'ensemble des stations de mesure SIAAP pour l'année 2003).

Ce rapport présente une méthode permettant de mettre en accord les conditions limite et les données de calibration pour garantir un jeu de données cohérent et homogène. La méthode utilisée est le co-krigeage, pour lequel certaines contraintes doivent être respectées avant de pouvoir l'utiliser. En effet, pour la reconstitution d'une condition limite amont, une bonne corrélation entre la chronique à interpoler et la chronique auxiliaire, nous renseignant sur la variabilité à plus courts pas de temps de la variable, est nécessaire.

Ce rapport s'appuie sur l'exemple des concentrations en matières en suspension (MES) pour lesquelles la reconstitution de la condition limite amont par co-krigeage sur 2007-2011 fonctionne très bien. L'exemple de la variable orthophosphates, qui ne peut pas être estimée par co-krigeage des données SIAAP par les données SEDIF, est lui aussi traité. La formulation mathématique adoptée dans ce rapport s'inspire du cours de géostatistique de [Marcotte \(2011\)](#) et de la thèse de [Lefebvre \(2010\)](#).

1 Théorie du co-krigeage

1.1 Pourquoi effectuer un co-krigeage ?

Il est fréquent de connaître plusieurs variables (ou une seule variable provenant de plusieurs sources différentes) mesurées, soit en un même point d'échantillonnage, soit en des points dif-

férents, et qui présentent une forte corrélation. La variable à interpoler est désignée par **variable prioritaire**. Les autres variables sont des **variables auxiliaires** ; ici nous n'en considérerons qu'une seule. Dans le contexte présent, la variable auxiliaire doit être échantillonnée plus finement que la variable principale. La méthode du co-krigeage nous permet ainsi de tirer parti d'une densité plus forte de prélèvement de la variable auxiliaire, qui nous renseigne sur la variabilité temporelle, tout en restant cohérents avec les valeurs mesurées de la variable principale.

1.2 Principe

Soit $Z_1(t)$ la variable temporelle prioritaire à interpoler, connue sur l'ensemble de points S_1 , et $Z_2(t)$ la variable auxiliaire, connue sur S_2 . Le co-krigeage consiste à estimer $Z_1(t)$ à partir des $Z_1(t_\alpha)$ et des $Z_2(t_\beta)$, avec $\alpha \in S_1$ et $\beta \in S_2$ (Equation 1). On note $Z_1^*(t_0)$ la valeur estimée de Z_1 au temps t_0 .

$$Z_1^*(t_0) = \sum_{\alpha \in S_1} \lambda_{1\alpha} Z_1(t_\alpha) + \sum_{\beta \in S_2} \lambda_{2\beta} Z_2(t_\beta) \quad (1)$$

1.3 Variogrammes simples et croisés

Définition Le variogramme temporel $\gamma(\tau)$ correspond à l'espérance du demi-écart quadratique entre deux variables séparées d'un intervalle de temps τ . Par exemple le variogramme de la variable Z_1 s'exprime :

$$\gamma_{11}(\tau) = \frac{1}{2} E [(Z_1(t + \tau) - Z_1(t))^2] \quad (2)$$

où E correspond à l'espérance mathématique.

Par définition, $\gamma_{11}(0) = 0$.

Le variogramme croisé décrit la variabilité conjointe de deux variables $Z_1(t)$ et $Z_2(t)$. Il peut être négatif lorsque les deux variables sont négativement corrélées :

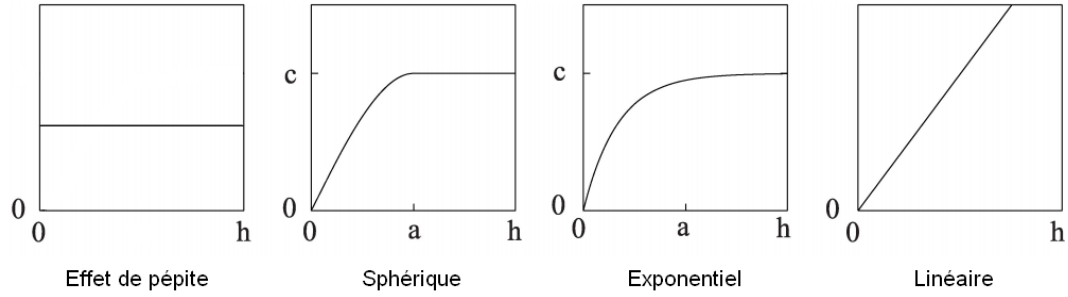
$$\gamma_{12}(\tau) = \frac{1}{2} E [(Z_1(t + \tau) - Z_1(t)) \cdot (Z_2(t + \tau) - Z_2(t))] \quad (3)$$

Variogramme expérimental En pratique, on calcule un variogramme expérimental à partir des données disponibles aux points expérimentaux t_i . Le variogramme expérimental de $Z_1(t)$ (γ_{11}^* , aussi couramment noté γ_1^*) vaut :

$$\gamma_{11}^*(\tau) = \frac{1}{2N(\tau)} \sum_{i=1}^{N(\tau)} [(Z_1(t_i + \tau) - Z_1(t_i))^2] \quad (4)$$

Le variogramme de $Z_2(t)$ (γ_{22}^* ou γ_2^*) peut être calculé de la même manière. Le variogramme croisé de $Z_1(t)$ et $Z_2(t)$ vaut :

$$\gamma_{12}^*(\tau) = \frac{1}{2N(\tau)} \sum_{i=1}^{N(\tau)} [(Z_1(t_i + \tau) - Z_1(t_i)) \cdot (Z_2(t_i + \tau) - Z_2(t_i))] \quad (5)$$


 FIGURE 1 – Exemples de modèles de variogrammes courants (Lefebvre, 2010). c : palier, a : portée.

Modèle de variogramme : le modèle linéaire de co-régionalisation Les variogrammes expérimentaux simples et croisés sont ensuite approchés par un modèle variographique, nécessaire pour les calculs en tout point. Le modèle variographique retenu est le modèle linéaire de co-régionalisation. Dans ce modèle, les deux variables sont des combinaisons linéaires de facteurs temporels associées à différentes structures élémentaires (Figure 1). Les variogrammes simples et croisés calculés sont donc des combinaisons linéaires des mêmes composantes élémentaires. Ce type de modèle variographique a été employé par Polus et al. (2011) dans le cas de la reconstruction de conditions limite amont en nitrates pour PROSE.

1.4 Système de co-krigeage

L'erreur d'estimation sur la variable prioritaire en un point t_0 s'écrit :

$$\epsilon = Z_1^*(t_0) - Z_1(t_0) \quad (6)$$

$Z_1^*(t_0)$ est la valeur estimée de $Z_1(t_0)$ au temps t_0 . La variance de l'erreur d'estimation s'écrit :

$$\begin{aligned} \text{Var}(Z_1^*(t_0) - Z_1(t_0)) &= \text{Var}(Z_1(t_0)) \\ &+ \sum_{i \in S_1} \sum_{j \in S_1} \lambda_{1i} \lambda_{1j} \text{Cov}(Z_1(t_i), Z_1(t_j)) \\ &+ \sum_{i \in S_2} \sum_{j \in S_2} \lambda_{2i} \lambda_{2j} \text{Cov}(Z_2(t_i), Z_2(t_j)) \\ &+ 2 \sum_{i \in S_1} \sum_{j \in S_2} \lambda_{1i} \lambda_{2j} \text{Cov}(Z_1(t_i), Z_2(t_j)) \\ &- 2 \sum_{i \in S_1} \lambda_{1i} \text{Cov}(Z_1(t_0), Z_1(t_i)) \\ &- 2 \sum_{i \in S_2} \lambda_{2i} \text{Cov}(Z_1(t_0), Z_2(t_i)) \end{aligned} \quad (7)$$

On impose une condition de non biais sur l'estimateur Z_1^* , ce qui induit :

- si l'on estime directement par co-krigeage une combinaison linéaire des variables, la valeur co-krigée sera égale à la même combinaison linéaire appliquée aux valeurs co-krigées de chaque variable ;
- le co-krigeage constitue un interpolateur exact : si l'on estime un point connu, on retrouve la valeur connue ;
- les points les plus proches du point estimé reçoivent généralement les poids les plus importants ;
- le co-krigeage tient compte de la taille du support de la variable à estimer et de la position des points entre eux ;
- la continuité du phénomène étudié est prise en compte (effet de pépite, anisotropie, etc.) ;
- les estimations sont généralement moins variables que les teneurs réelles que l'on cherche à estimer ;
- si la variable auxiliaire est échantillonnée aux mêmes points (ou en un sous ensemble de points) que la variable principale et que les covariances croisées et directes sont proportionnelles, alors le co-krigeage est identique au krigeage.

1.7 Dans quel cas le co-krigeage est-il possible ?

Pour réaliser un co-krigeage d'une variable prioritaire par une variable auxiliaire, dans le contexte présent, il est important que ces deux variables soient bien corrélées.

Pour s'assurer de la fiabilité du co-krigeage et valider le voisinage, une validation croisée peut-être réalisée. Le principe de la validation croisée est de retirer tour à tour chaque point de la chronique à interpoler et de l'estimer par co-krigeage. Si la corrélation entre la valeur mesurée et la valeur estimée par co-krigeage de chacun de ces points est bonne, le co-krigeage peut être considéré comme fiable.

2 Co-krigeage des données SIAAP de MES à Choisy par les données SEDIF

Afin d'illustrer la théorie exposée précédemment, nous prendrons l'exemple des mesures de concentrations en matières en suspension (MES) à Choisy (Figure 2), sur la Seine, utilisées comme condition limite amont du modèle ProSe. Deux jeux de mesures sont disponibles : l'un fourni par le SIAAP, à pas de temps hebdomadaires (235 valeurs sur 2007-2011), et l'autre fourni par le SEDIF, à pas de temps plus fins (788 valeurs sur 2007-2011).

On observe que les concentrations estivales mesurées par le SEDIF sont inférieures à celles du SIAAP (Figure 3). Sur les cinq années étudiées, les moyennes des mesures sont de 21.98 mg/L pour le SEDIF et de 21.59 mg/L pour le SIAAP aux dates de mesure communes, et de 4.42 mg/L pour le SEDIF et 5.86 mg/L pour le SIAAP en période de basses eaux (débit à Austerlitz inférieur au débit médian sur 2007-2011).

Il semblerait qu'une différence entre les protocoles de mesures ou de prélèvement utilisés par les deux organismes entraîne un écart dans leurs mesures. Il est donc souhaitable, pour pouvoir comparer les résultats de simulation aux mesures du SIAAP, de recalculer la condition limite amont

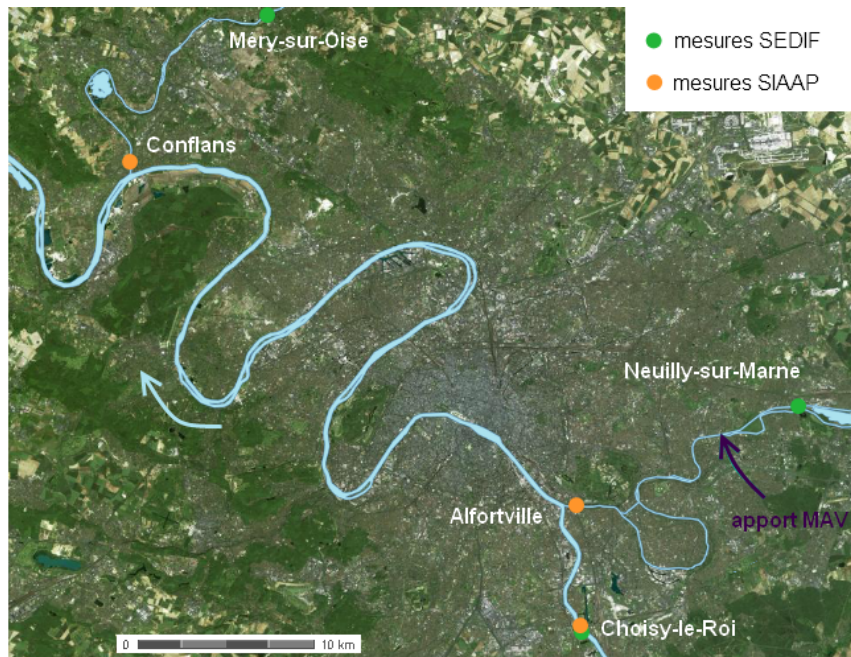


FIGURE 2 – Localisation des points de mesure amont du SIAAP et du SEDIF (Source : Géoportail).

sur les mesures du SIAAP à Choisy, cohérentes avec les mesures du SIAAP aux autres sites d'observation qui servent de points de validation pour le modèle.

Les deux chroniques, même si elles présentent des différences en termes d'amplitude des valeurs, ont des valeurs synchrones très fortement corrélées ($\rho = 0.97$, Figure 3), ce qui permet d'envisager un co-krigeage des données SIAAP par les données SEDIF, et de remplacer la condition limite SEDIF par la chronique interpolée. Cette interpolation permet de rester en cohérence avec les valeurs mesurées par le SIAAP, que nous considérons comme variable prioritaire, tout en utilisant l'information que fournissent les données plus denses du SEDIF sur la variabilité temporelle des concentrations.

Pour les besoins du co-krigeage, un modèle variographique (ici composé d'un effet de pépité, d'une composante exponentielle de portée 10 jours et d'une composante périodique de période 359 jours) a été ajusté aux variogrammes expérimentaux simples et croisés des concentrations en MES mesurées à Choisy par le SIAAP et le SEDIF (Figure 4).

Dans notre cas, nous nous sommes assurés de la validité d'un co-krigeage des données SIAAP de 2007 à 2011 par les données SEDIF avec un voisinage de 50 jours (permettant de prendre en compte une soixantaine de valeurs de concentrations du SEDIF, une trentaine de valeurs antérieures et une trentaine postérieures au point estimé). La validation croisée donne une corrélation de 0.95 entre concentrations mesurées et concentrations estimées par co-krigeage (Figure 5).

Le co-krigeage des données SIAAP de concentrations en matières en suspension à Choisy par les données SEDIF a permis d'interpoler la chronique des données SIAAP en prenant en compte la variabilité temporelle connue grâce aux données plus denses du SEDIF (Figure 6).

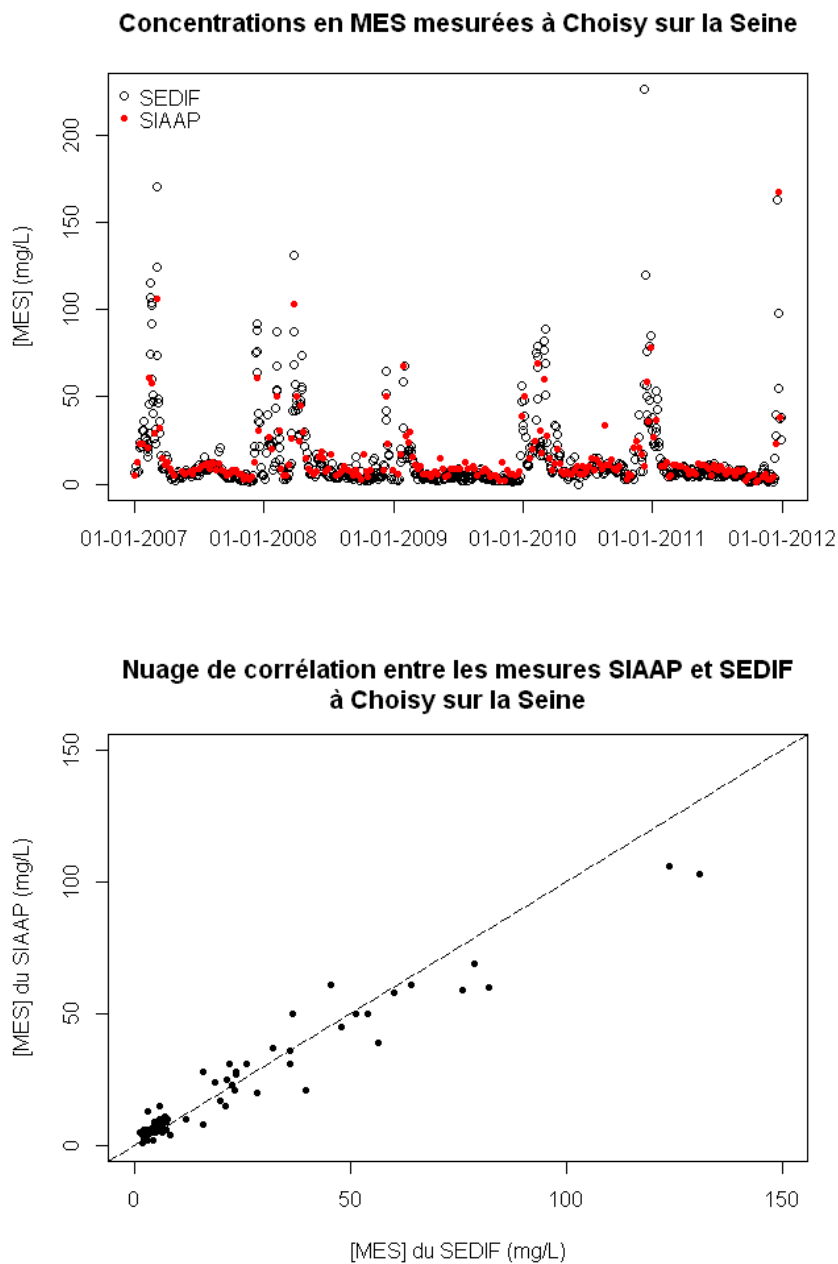


FIGURE 3 – Comparaison des chroniques en matières en suspension mesurées à Choisy par la SEDIF et par le SIAAP.

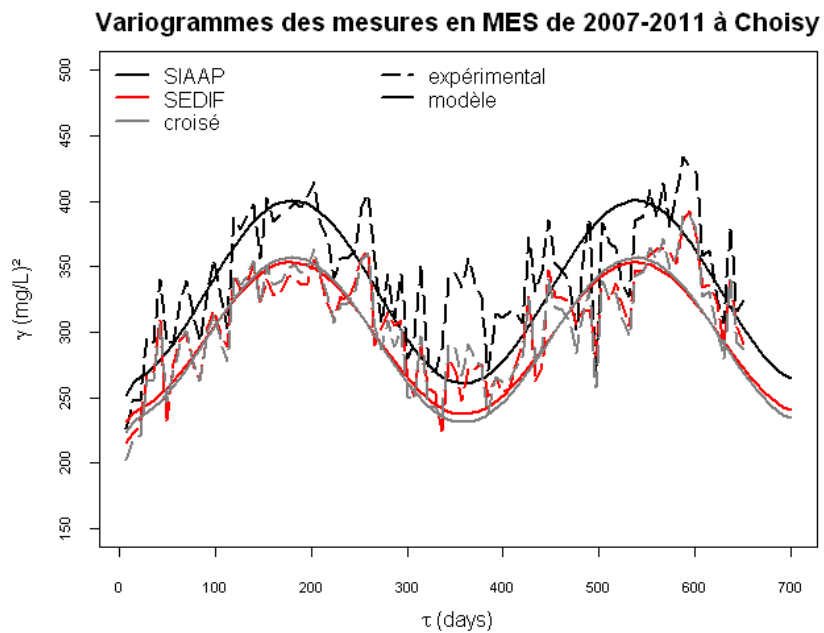


FIGURE 4 – Variogrammes expérimentaux et modèles variographiques des concentrations en MES mesurées à Choisy par le SIAAP et le SEDIF.

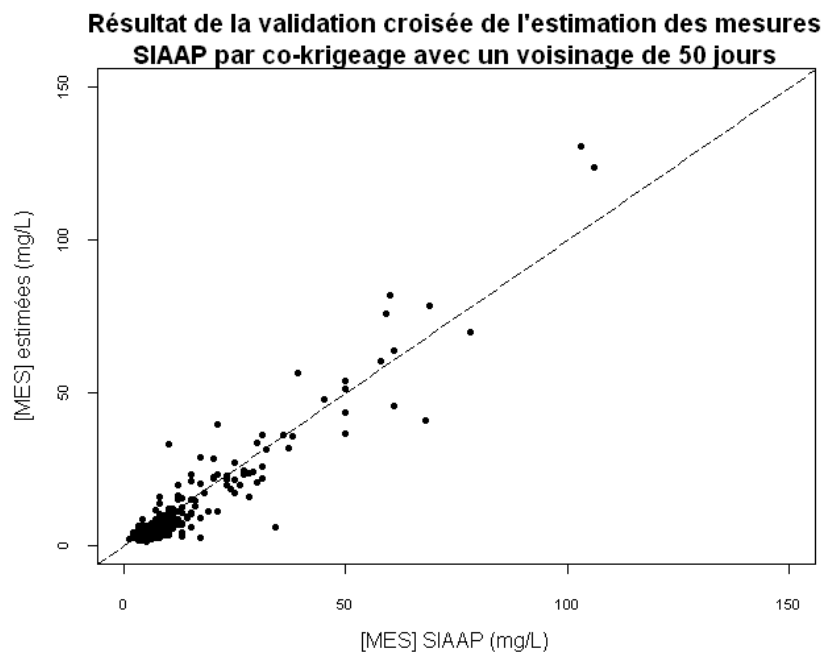


FIGURE 5 – Résultat de la validation croisée du co-krigeage des mesures SIAAP en MES par les mesures SEDIF sur 2007-2011 avec un voisinage de 50 jours.

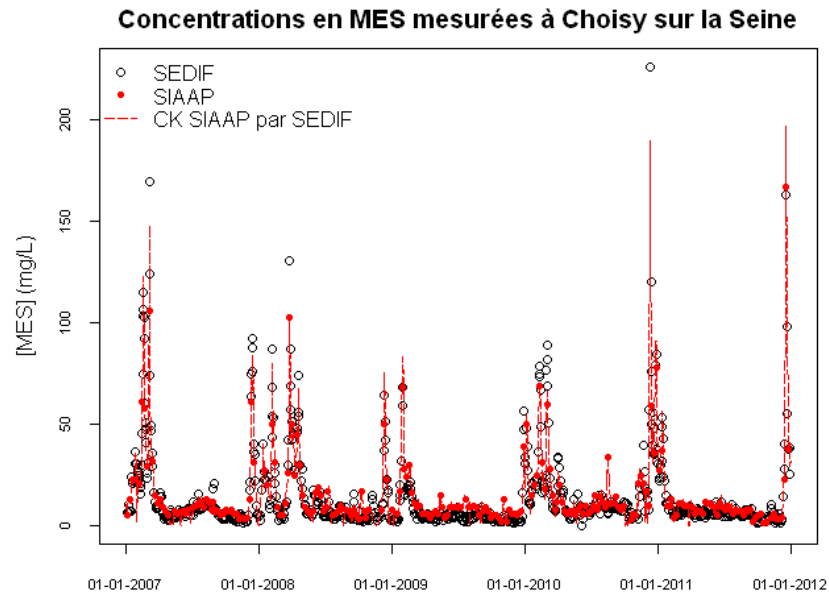


FIGURE 6 – Chroniques mesurées en MES à Choisy et chronique des mesures SIAAP co-krigées par les données SEDIF sur 2007-2011.

3 Limites du co-krigeage : cas des orthophosphates

Pour la période simulée (2007-2011), les différences entre orthophosphates mesurés par le SEDIF et orthophosphates mesurés par le SIAAP sont trop grandes et ne permettent pas de réaliser un co-krigeage pour le problème posé (Figure 7). La reconstruction de la condition limite en orthophosphates par co-krigeage des données SIAAP par les données SEDIF n’apporterait rien. En effet, la corrélation entre ces deux chroniques est de 0.27, ce qui est particulièrement visible lorsque l’on trace leurs variogrammes simples et croisés. Les variabilités mesurées ne sont clairement pas les mêmes, quel que soit le pas de temps considéré (Figure 3). Ceci a pour conséquence de donner très peu de poids à la variable auxiliaire dans le système de co-krigeage.

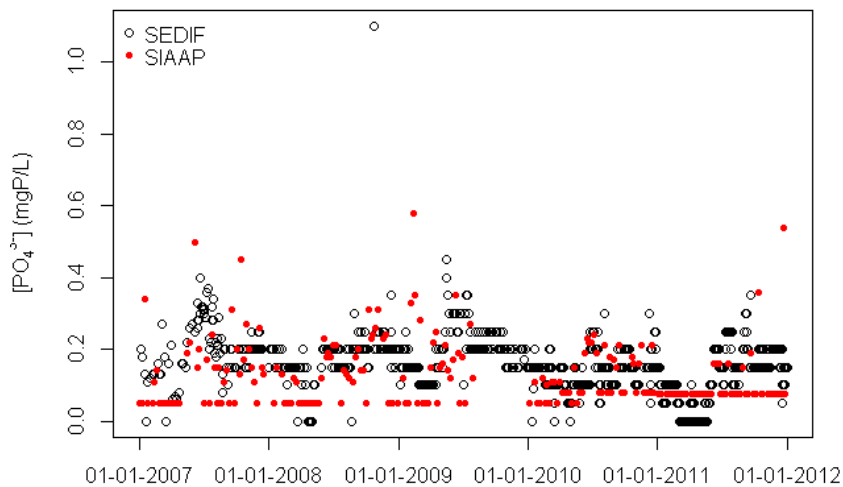
Les données SEDIF brutes, plus denses que celle du SIAAP, sont donc utilisées comme condition limite amont. Des pics de concentration, qui ne sont pas visibles sur les mesures du SIAAP, sont introduits au niveau des amonts et constituent certainement la source d’une partie de l’erreur sur les concentrations en orthophosphates calculées le long du linéaire simulé (Vilmin et al., 2013).

4 Conclusion

Polus et al. (2011) ont montré que cette méthode d’interpolation permet de reconstruire des jeux de données (conditions limite amont et données de validation) homogènes et cohérents en ce qui concerne les concentrations amont en nitrates. Cette reconstruction de données amont permet de s’affranchir des biais dus aux différences de protocoles de mesure pour les différentes chroniques disponibles, qui se répercutent tout au long du linéaire simulé.

Cette méthode est également employée pour construire les conditions limite amont en MES

Concentrations en orthophosphates mesurées à Choisy sur la Seine



Nuage de corrélation entre les mesures SIAAP et SEDIF à Choisy sur la Seine

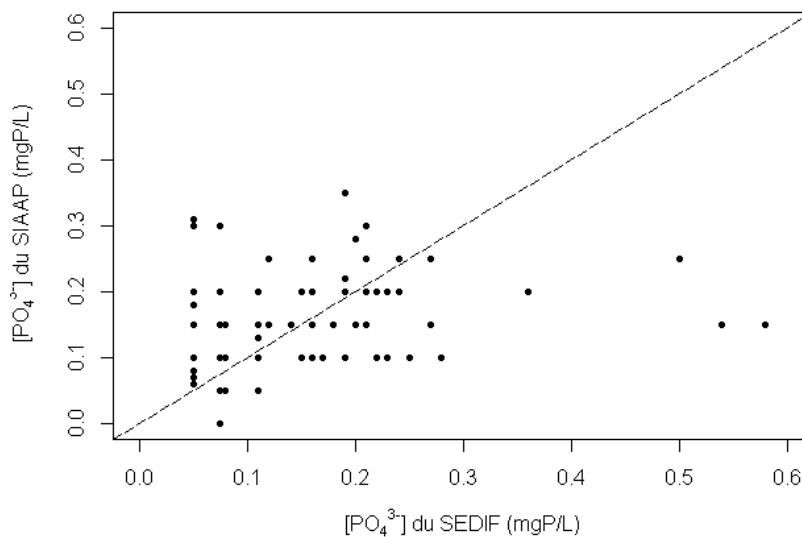


FIGURE 7 – Comparaison des chroniques en orthophosphates mesurées à Choisy par la SEDIF et par le SIAAP.

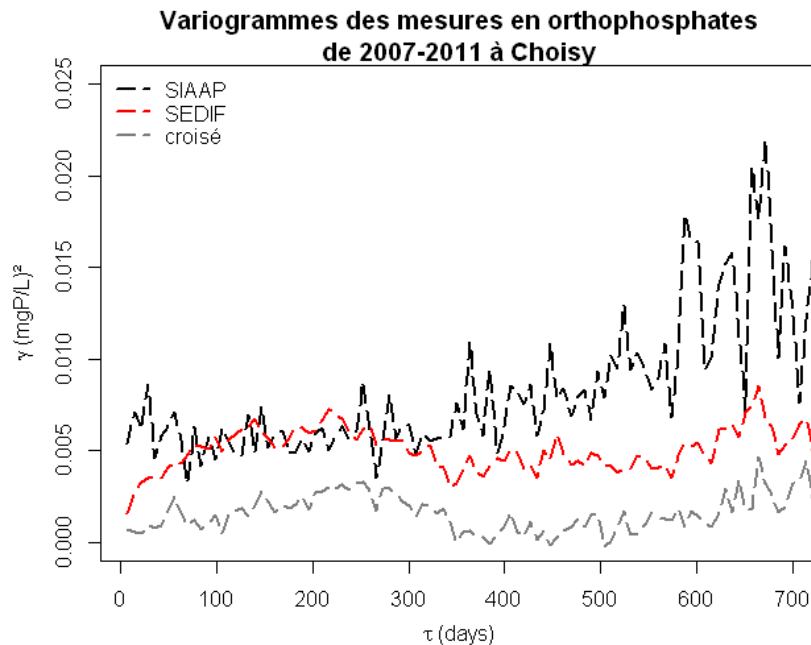


FIGURE 8 – Variogrammes expérimentaux simples et croisés des chroniques d’orthophosphates mesurées à Choisy par le SIAAP et le SEDIF en 2007-2011.

(Vilmin et al., 2012, 2013). Les données d’entrée en nitrates et en MES dans la Seine et dans l’Oise sont maintenant systématiquement reconstruites par co-krigeage. Les mesures amont sur la Seine sont effectuées sur la commune de Choisy aussi bien pour le SEDIF que pour le SIAAP. Sur l’Oise, le SEDIF réalise ses mesures sur des prélèvements effectués à Méry-sur-Oise, 20 km en amont de la confluence avec la Seine, tandis que le SIAAP prélève juste en amont de la confluence (Figure 2). Aucun rejet important ne figure entre les deux points de prélèvement et les deux chroniques de concentrations mesurées sont fortement corrélées aussi bien pour les MES que pour les nitrates. Elles peuvent donc faire l’objet d’un co-krigeage. Les conditions limite amont sur la Marne ne peuvent pas être estimées de la même manière, aucune donnée SIAAP n’étant disponible en amont de la station d’épuration de Marne Aval.

Les conditions limite amont des autres variables d’entrée de PROSE (ammonium, nitrites et orthophosphates) ne peuvent pas être reconstruites de la même manière. Les mesures effectuées par le SEDIF et le SIAAP ne sont effectivement pas suffisamment corrélées ($\rho < 0.5$ pour les trois variables citées) et aucun bénéfice ne serait tiré d’une interpolation des données SIAAP par co-krigeage.

Références

- Lefebvre, E., 2010. Apports croisés de la modélisation géostatistique et déterministe. exemple des nitrates et de l’oxygène dissous dans un réseau hydrographique. Ph.D. thesis, Ecole des Mines de Paris.
- Marcotte, D., 2011. Géostatistique — 9. Cokrigeage. Ecole Polytechnique de Montréal.

- Polus, E., Flipo, N., de Fouquet, C., Poulin, M., 2011. Geostatistics for assessing the efficiency of a distributed physically-based water quality model : application to nitrate in the seine river. *Hydrological Processes* 25 (2), 217–233.
- Vilmin, L., Flipo, N., Mouchel, J.-M., Grouz, N., Billen, G., Garnier, J., Poulin, M., de Fouquet, C., Groleau, A., Escoffier, N., 2013. Modélisation du fonctionnement écologique de la seine : sensibilité des résultats aux processus hydro-sédimentaires. Tech. rep., PIREN-Seine.
- Vilmin, L., Flipo, N., Poulin, M., Billen, G., de Fouquet, C., Groleau, A., Garnier, J., 2012. Sensibilité du modèle PROSE aux paramètres d'érosion de la matière en suspension : Application à la simulation du phosphore. Tech. rep., PIREN-Seine.