Assimilation séquentielle de données en géosciences

Thomas Romary

Centre de Géosciences, Equipe Géostatistique

June 13, 2019

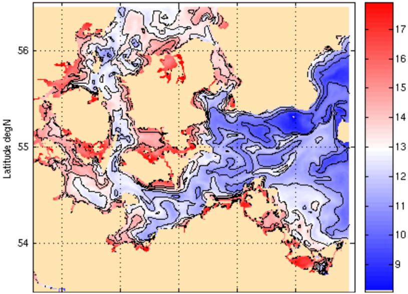
Qu'est-ce que l'assimilation de données ?

Tentative de définition

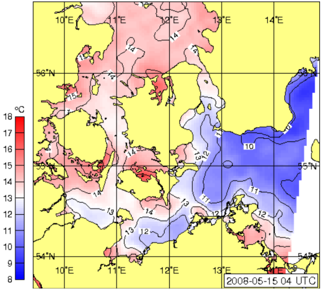
L'assimilation de données est la technique selon laquelle des observations et la sortie d'un modèle numérique sont combinées pour produire une estimation optimale de l'état d'un système évolutif

Exemple : température de surface de la mer

SST: Simulation (BSHcmod)



SST: Satellite (AVHRR)



Information : modèle

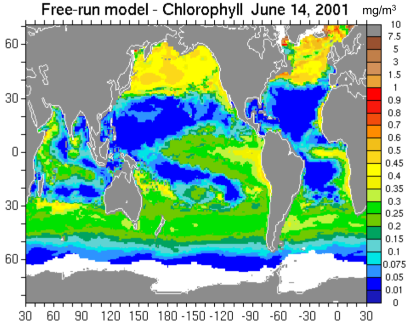
Information : données

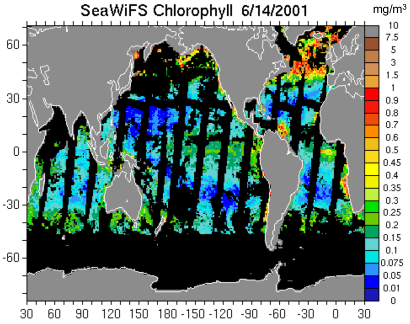
On combine les deux types d'information par assimilation de données

$\Longrightarrow$ meilleures données d'analyse (rétrospectives), prédictions pour la température, la couverture de glace, etc.

extrait de "An overview of DA", L. Nerger

Exemple : Chlorophylle dans l'océan





Information : modèle

Information : données

On combine les deux types d'information par assimilation de données

$\Longrightarrow$ meilleures données d'analyse (rétrospectives), prédictions

extrait de "An overview of DA", L. Nerger

Exemples d'application

Historiquement : météorologie, puis océanographie

Aujourd'hui, de plus en plus de domaines d'application

- glaciologie
- sismologie
- fusion nucléaire
- épidémiologie
- agronomie
- qualité de l'air
- etc.

Pour quoi faire ?

Historiquement : estimation d'un état initial, pour la prédiction de la météo

Aujourd'hui, beaucoup d'autres types d'applications

- conditions initiales pour la prédiction
- calibration et validation de modèles
- conception, validation et surveillance de systèmes d'observation
- réanalyse
- meilleure compréhension des systèmes (erreurs de modélisation, d'observation, interaction de processus physiques, paramètres, etc.)
- etc.

Formalisation

On cherche à caractériser l'état $(x_1, \ldots, x_T)$ du système au cours du temps discrétisé entre 1 et T, à partir des observations $(y_1, \ldots, y_T)$

Deux ingrédients

1. un modèle d'évolution F

$$x_t = F(x_{t-1}) + \varepsilon_t^F$$

2. un opérateur d'observation H

$$y_t = H(x_t) + \varepsilon_t^H$$

où x_0 est connu (conditions initiales), ε^F et ε^H sont les erreurs de modélisation et d'observation

NB x_t ne dépend que de x_{t-1} (Markov)

Formalisation

On cherche à caractériser l'état $(x_1, \ldots, x_T)$ du système au cours du temps discrétisé entre 1 et T, à partir des observations $(y_1, \ldots, y_T)$

Deux ingrédients

1. un modèle d'évolution F

$$x_t = F(x_{t-1}) + \varepsilon_t^F$$

2. un opérateur d'observation H

$$y_t = H(x_t) + \varepsilon_t^H$$

où x_0 est connu (conditions initiales), ε^F et ε^H sont les erreurs de modélisation et d'observation

NB x_t ne dépend que de x_{t-1} (Markov)

Analyse Bayésienne

Formule de Bayes

Soit x l'état du système et y une observation, on a

$$\begin{aligned}\mathbb{P}(x|y) &= \frac{\mathbb{P}(y|x)\mathbb{P}(x)}{\mathbb{P}(y)} \\ &\propto \mathbb{P}(y|x)\mathbb{P}(x)\end{aligned}$$

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \ldots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \ldots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\mathbb{P}(x_{1:t}|y_{1:t}) \propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \dots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \dots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\mathbb{P}(x_{1:t}|y_{1:t}) \propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

On caractérise ainsi séquentiellement la loi de x_t sachant les observations passées
pour plus de détails, voir Wikle and Berliner [2007]

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \ldots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \ldots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\mathbb{P}(x_{1:t}|y_{1:t}) \propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

On caractérise ainsi séquentiellement la loi de x_t sachant les observations passées
pour plus de détails, voir Wikle and Berliner [2007]

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \ldots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \ldots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\mathbb{P}(x_{1:t}|y_{1:t}) \propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

$$\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})$$

On caractérise ainsi séquentiellement la loi de x_t sachant les observations passées
pour plus de détails, voir Wikle and Berliner [2007]

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \dots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \dots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\begin{aligned}
\mathbb{P}(x_{1:t}|y_{1:t}) &\propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1}) \\
&\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1}) \\
&\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})
\end{aligned}$$

On caractérise ainsi séquentiellement la loi de x_t sachant les observations passées

pour plus de détails, voir Wikle and Berliner [2007]

Analyse Bayésienne

Notons $x_{1:T} = (x_1, \ldots, x_T)$ la *trajectoire* de l'état du système et $y_{1:T} = (y_1, \ldots, y_T)$ celle des observations

On a $\forall t \leq T$

$$\begin{aligned}
\mathbb{P}(x_{1:t}|y_{1:t}) &\propto \mathbb{P}(y_t|x_{1:t}, y_{1:t-1})\mathbb{P}(x_{1:t}|y_{1:t-1}) \\
&\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{1:t-1}, y_{1:t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1}) \\
&\propto \mathbb{P}(y_t|x_t)\mathbb{P}(x_t|x_{t-1})\mathbb{P}(x_{1:t-1}|y_{1:t-1})
\end{aligned}$$

On caractérise ainsi séquentiellement la loi de x_t sachant les observations passées
pour plus de détails, voir Wikle and Berliner [2007]

Analyse Bayésienne

Quels problèmes peut-on résoudre ?

1. *filtrage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:t})$
2. *prédiction*, $\Longrightarrow \mathbb{P}(x_{T+n}|y_{1:T})$
3. *lissage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:T})$

Analyse Bayésienne

Quels problèmes peut-on résoudre ?

1. *filtrage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:t})$
2. *prédiction*, $\Longrightarrow \mathbb{P}(x_{T+n}|y_{1:T})$
3. *lissage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:T})$

Analyse Bayésienne

Quels problèmes peut-on résoudre ?

1. *filtrage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:t})$
2. *prédiction*, $\Longrightarrow \mathbb{P}(x_{T+n}|y_{1:T})$
3. *lissage*, $\Longrightarrow \mathbb{P}(x_t|y_{1:T})$

Analyse Bayésienne

Quels problèmes peut-on résoudre ?

1. *filtrage*, $\Longrightarrow \mathbb{P}(x_t | y_{1:t})$
2. *prédiction*, $\Longrightarrow \mathbb{P}(x_{T+n} | y_{1:T})$
3. *lissage*, $\Longrightarrow \mathbb{P}(x_t | y_{1:T})$

Résolution

Séquentiellement et généralement en deux étapes

1. **Propagation** (*forecast*): on laisse évoluer le système selon F sans tenir compte des observations
2. **Correction** (*analysis*): on corrige l'état propagé en fonction des observations

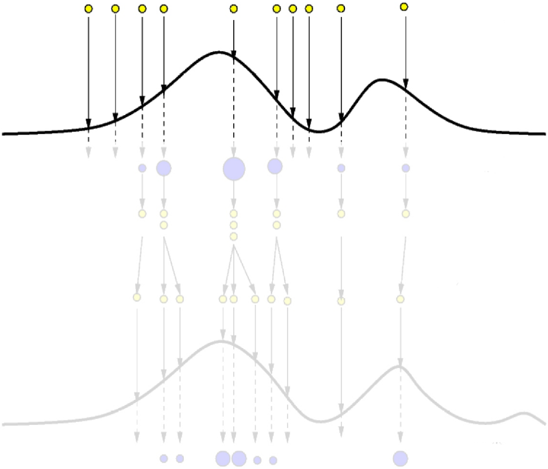
Cas particulier

Si F et H sont linéaires et les erreurs gaussiennes, alors on peut calculer explicitement toutes les distributions conditionnelles

C'est le filtre de Kalman

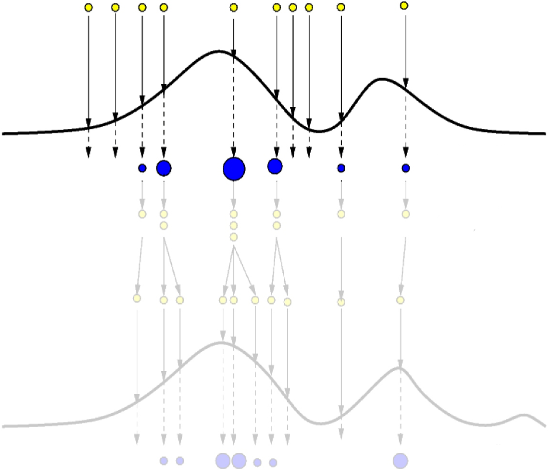
Filtrage particulaire

Principe



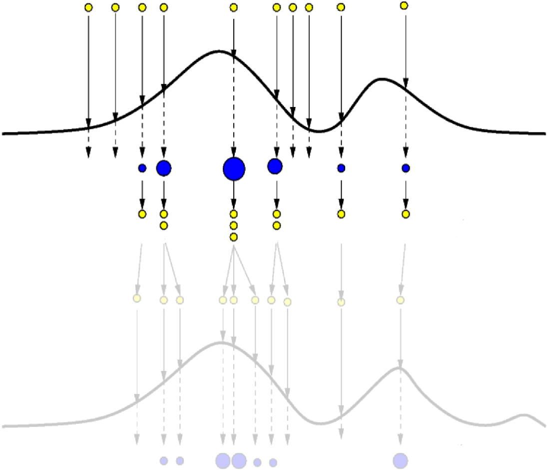
Filtrage particulaire

Principe



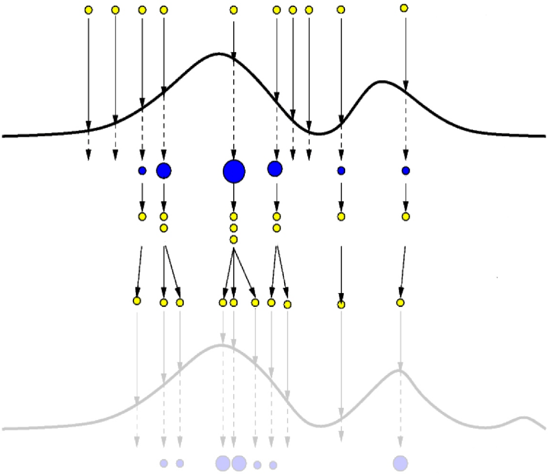
Filtrage particulaire

Principe



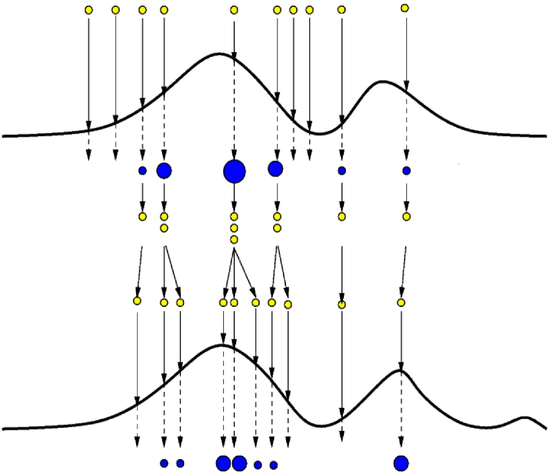
Filtrage particulaire

Principe



Filtrage particulaire

Principe



Un aperçu du zoo des méthodes d'assimilation de données

- Variationnelles
  - 3D-var
  - 4D-var
  - ...
- Stochastiques
  - KF-S
  - EKF-S
  - EnKF-S
  - PF-S (SMC)
  - ...
- méthodes hybrides

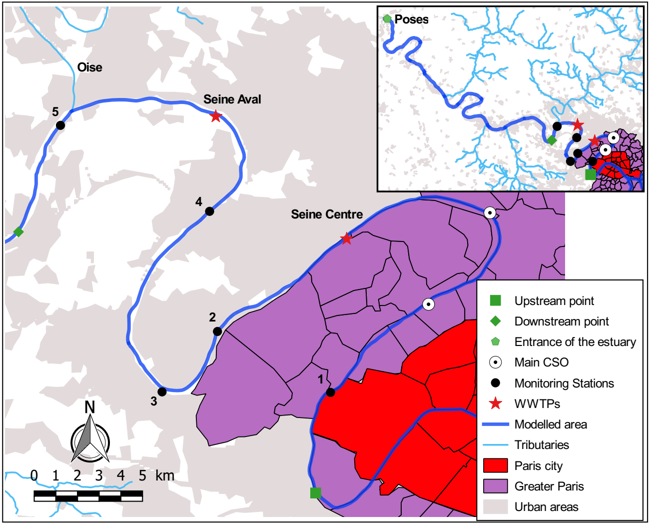
Exemples d'application

Historiquement : météorologie, puis océanographie

Aujourd'hui, de plus en plus de domaines d'application

- glaciologie
- sismologie
- fusion nucléaire
- épidémiologie
- agronomie
- qualité de l'air
- etc.
- **qualité de l'eau**

En qualité de l'eau



Modèle Prose

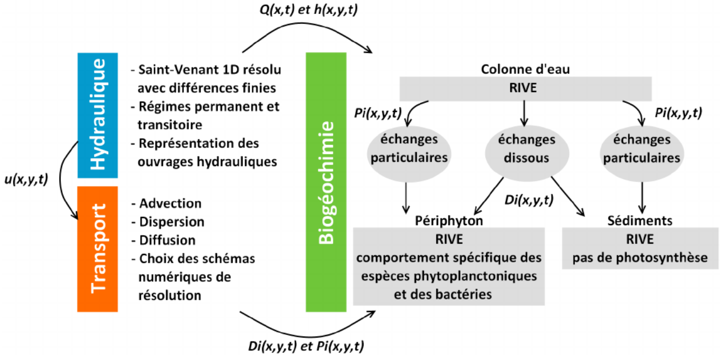
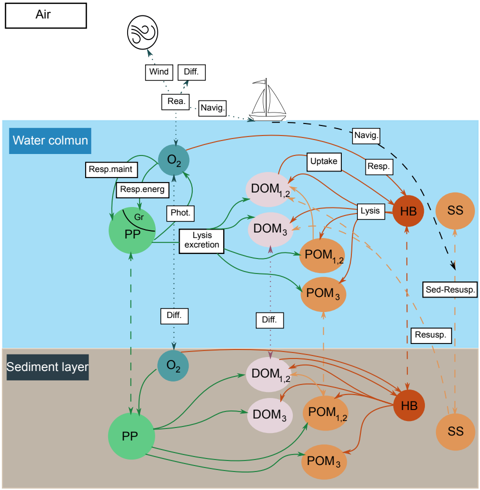


FIGURE A.1 – Structure du modèle PROSE. $u(x,y,t)$, $h(x,y,t)$, $P_i(x,y,t)$ et $D_i(x,y,t)$ correspondent à la vitesse d'écoulement, à la hauteur d'eau et flux d'espèces particulaires et dissoutes au point (x,y) à l'instant t. $Q(x,t)$ est le débit à la section d'abscisse x et au temps t.

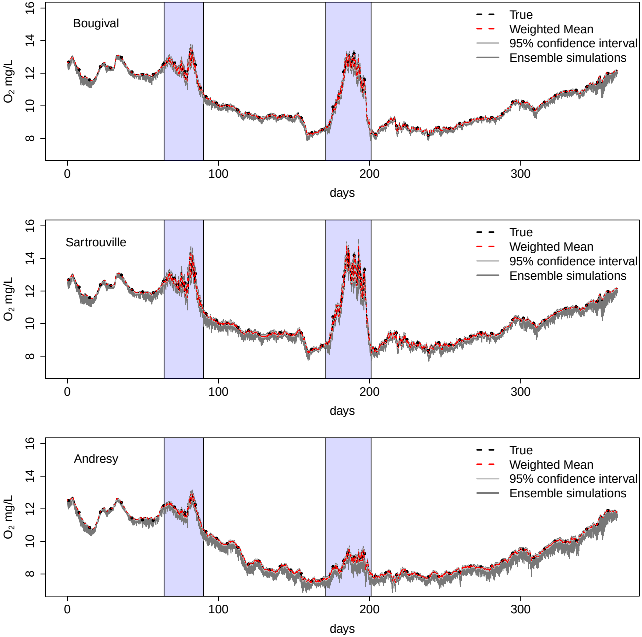
extrait de la thèse de L. Vilmin

Module biogéochimique C-RIVE

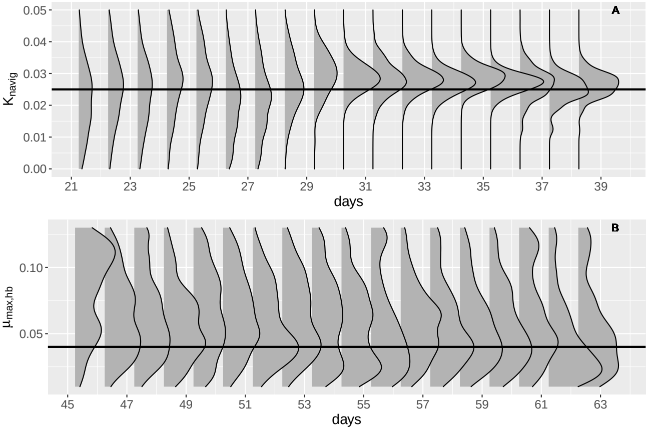


extrait de Wang et al. [2018]

Quelques résultats

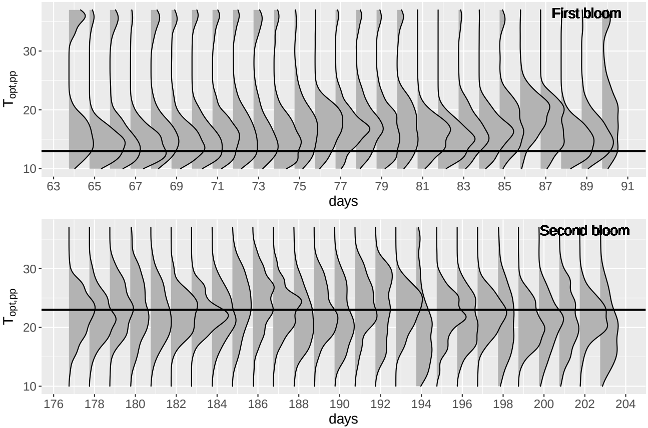


Quelques résultats



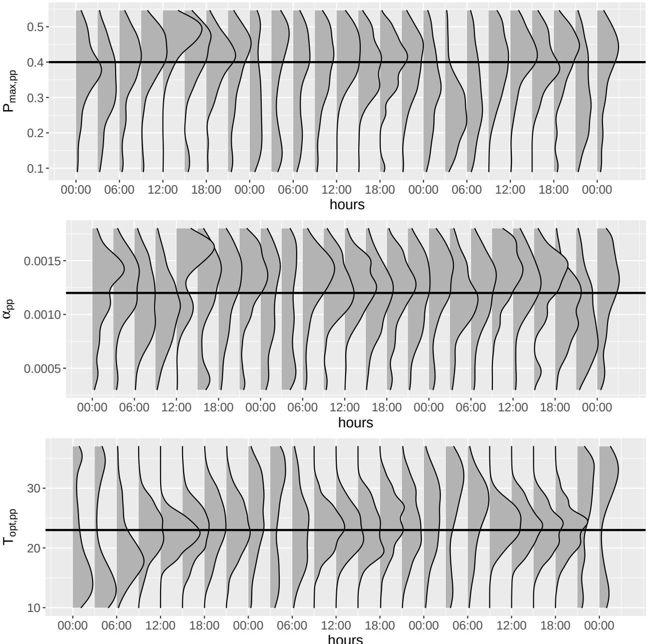
extrait de Wang et al. [2019]

Quelques résultats



extrait de Wang et al. [2019]

Quelques résultats



References

Shuaitao Wang, Nicolas Flipo, and Thomas Romary. Time-dependent global sensitivity analysis of the c-rive biogeochemical model in contrasted hydrological and trophic contexts. *Water research*, 144:341–355, 2018.

Shuaitao Wang, Nicolas Flipo, and Thomas Romary. Oxygen data assimilation for metabolism's parameter estimation in urban river systems. *submitted to Water research*, 2019.

Christopher K Wikle and L Mark Berliner. A bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16, 2007.