



---

# Exploration de l'assimilation de données multivariée pour la chimie atmosphérique

---

**Antoine PERROT**, 2ème année de Master MAPI3 (Mathématiques Appliquées pour l'Ingénierie,  
l'Industrie et l'Innovation)

encadré par **Olivier PANNEKOUCKE** (CERFACS, CNRM) et **Vincent GUIDARD** (CNRM).

Enseignant tuteur : **François BACHOC**

25 août 2021

# Sujet

À Météo-France, la prévision de la composition chimique de l'atmosphère est réalisée à l'aide du modèle de chimie transport MOCAGE (CTM) dont l'état initial est estimé à l'aide d'une chaîne d'assimilation de données propre au modèle. Actuellement, l'assimilation de données des espèces chimiques est construite en assimilant les espèces indépendamment les unes des autres, ce qui correspond à une assimilation univariée. Il s'agit là d'une approximation très grossière qu'on souhaite relaxer. Pour cela il faut envisager une assimilation multivariée, ce qui passe par la modélisation et l'estimation des covariances d'erreur de prévision inter-espèces (Emili et al. 2016). L'objectif du stage sera d'étudier différentes approches pour la modélisation et l'estimation des fonctions de covariance d'erreur de prévision multivariée, et la caractérisation de leur évolution temporelle. On s'appuiera pour cela sur le formalisme du filtre de Kalman paramétrique (Pannekoucke et al. , 2016 ; Pannekoucke et al. 2018) et sur un outil de génération automatique de réseau de neurone pour apprendre la dynamique à partir des données (Pannekoucke and Fablet, 2020 ; Pannekoucke, 2020 ; github repository).

## Remerciements

Je tiens à remercier mes encadrants de stage, **Vincent Guidard** et **Olivier Pannekoucke** qui m'ont d'abord accordé leur confiance en me désignant pour ce stage particulièrement intéressant, puis guidé tout au long de ces six mois dans mes travaux. Leur aide a également joué un grand rôle dans l'obtention de la ma bourse de thèse, qui sera encadrée par Olivier. Enfin, je remercie Olivier pour la qualité de son écoute et de ses conseils quant à mon orientation ainsi que pour toute la culture météorologique qu'il m'a apportée au cours de nos échanges.

# Table des matières

<b>1</b>	<b>Différentes méthodes d'assimilation de données</b>	<b>3</b>
1.1	L'assimilation de données dans la prévision de la composition chimique de l'atmosphère .	3
1.2	Filtre de Kalman et introduction aux méthodes d'ensembles . . . . .	3
1.2.1	Le filtre de Kalman (KF) . . . . .	3
1.2.2	Le filtre de Kalman d'ensemble (EnKF) . . . . .	4
1.3	Illustration avec un modèle chimique jouet : Lotka-Volterra . . . . .	5
1.3.1	Le modèle chimique de Lotka-Volterra . . . . .	5
1.3.2	Exemple d'application de l'EnKF avec Lotka-Volterra . . . . .	6
1.4	Une nouvelle méthode en assimilation de données : le filtre de Kalman paramétrique (PKF)	8
1.4.1	Principe . . . . .	8
1.4.2	Illustration du PKF pour la prévision avec Lotka-Volterra . . . . .	8
<b>2</b>	<b>Transport 1D d'un polluant dans l'atmosphère</b>	<b>10</b>
2.1	Formuler le PKF . . . . .	10
2.2	Modèle de covariance univarié . . . . .	11
2.3	Simulation numérique . . . . .	11
2.3.1	Paramètres de simulation . . . . .	11
2.3.2	Génération des champs d'erreurs aléatoires . . . . .	12
2.3.3	Résultats : comparaison des prévisions du PKF et de l'EnKF . . . . .	13
2.3.4	Résultats : comparaison des erreurs de prévisions pour le PKF et l'EnKF . . . . .	14
2.4	Influence de l'erreur de modèle sur les diagnostics de l'EnKF . . . . .	15
2.4.1	Mise en évidence d'un cas de divergence du filtre de Kalman d'ensemble . . . . .	15
2.4.2	Impact de la réduction des longueurs de portées initiales sur les diagnostics . . . . .	15
2.5	Conclusion dans le cadre univarié de l'advection 1D . . . . .	16
<b>3</b>	<b>Un modèle de chimie transport simplifié pour l'exploration d'une formulation multivariée du PKF</b>	<b>17</b>
3.1	Formulation multivariée du PKF avec le SymPKF . . . . .	17
3.2	Simplification du système PKF . . . . .	19
3.3	Simulation numérique : une prévision multivariée du système PKF . . . . .	21
3.4	Définition d'un modèle de corrélation multivarié . . . . .	22
3.5	Assimilation d'une observation : PKF et EnKF . . . . .	25
<b>4</b>	<b>Vers un modèle chimique transport plus réaliste</b>	<b>28</b>
4.1	Introduction du modèle chimique GRS . . . . .	28
4.2	Comparaison prévision PKF et EnKF . . . . .	29
<b>5</b>	<b>Conclusion et perspectives</b>	<b>32</b>

# 1 Différentes méthodes d'assimilation de données

## 1.1 L'assimilation de données dans la prévision de la composition chimique de l'atmosphère

L'assimilation de données est au coeur des problématiques de chimie atmosphérique, ou même en prévision numérique du temps (PNT). Il s'agit, partant à  $t = 0$  d'une connaissance *a priori* de l'état de l'atmosphère, notée  $\mathcal{X}_0 = \mathcal{X}^b$  ( $b$  pour *background*), et d'une erreur lui étant associée  $\Sigma_0 = \Sigma^b$ , d'estimer les états futurs de l'atmosphère  $(\mathcal{X}_t)_{t>0}$  à l'aide d'un modèle de prévision  $\mathcal{M}$  et leurs erreurs associées  $(\Sigma_t)_{t>0}$ . À intervalles réguliers (toutes les 6h), il faut mettre à jour la prévision à l'aide des observations recueillies entre temps. Ces observations peuvent provenir de capteurs sur terre, en mer, dans les airs (ballons ou avions de ligne) ou encore de satellites. L'étape de prise en compte des observations est appelée "phase d'analyse". Ainsi, l'assimilation de données consiste en une suite de cycles de prévision/analyse.

Le filtre de Kalman (Kalman and Others [1960]) (KF) est un algorithme à la base de l'assimilation de données bien que non utilisé dans les problématiques étudiées ici, pour des raisons qui seront explicitées plus tard. Il existe principalement deux alternatives au KF utilisées aujourd'hui et qui ont été développées au début des années 90 : les méthodes d'ensembles (EnKF) (Evensen [1994]) et les méthodes variationnelles (Courtier et al. [1994]). De plus en plus, des hybridations de ces deux méthodes sont formulées afin de tirer le meilleur de chacune des deux approches. En particulier l'hybridation permet de fournir des statistiques d'erreurs de prévision du jour dans les méthodes variationnelles (Berre et al. [2007]). Nous nous intéresserons ici principalement aux méthodes d'ensembles.

Plus récemment une nouvelle implémentation du filtre de Kalman a été proposée, le filtre de Kalman paramétrique (PKF) (Pannekoucke et al. [2016, 2018]), qui sera l'objet principal de ce stage et dont nous détaillerons plus tard le principe, les apports et les difficultés qui lui sont propres.

## 1.2 Filtre de Kalman et introduction aux méthodes d'ensembles

Nous rappelons le principes du filtre de Kalman, afin d'ensuite présenter le filtre de Kalman d'ensemble (EnKF), ses avantages et ses limites.

### 1.2.1 Le filtre de Kalman (KF)

Dans le cas d'opérateurs linéaires et d'incertitudes gaussiennes (entièrement caractérisées par une moyenne et une matrice de covariance), les équations du filtre de Kalman décrivent l'évolution de l'incertitude en caractérisant la dynamique de la moyenne et de la matrice de covariance tout au long des cycles de prévision et d'analyse.

Formellement, on peut modéliser une dynamique observée par le système :

$$\begin{cases} \mathcal{Y}_t = \mathbf{H}_t \mathcal{X}_t + \mathbf{v}_t, & \mathbf{v}_t \sim \mathcal{N}_{m_t}(0, \mathbf{R}_t) \\ \mathcal{X}_t = \mathbf{M}_t \mathcal{X}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t \sim \mathcal{N}_n(0, \mathbf{Q}_t) \\ \mathcal{X}_0 = \mathcal{X}^b, \Sigma^a = \Sigma^b \end{cases}$$

où  $\mathcal{Y}_t$  est le vecteur  $m_t$  dimensionnel d'observations au temps  $t$ ,  $\mathbf{H}_t$  l'opérateur d'observation,  $\mathbf{M}_t$  le modèle de prévision au temps  $t$ ,  $\mathcal{X}_t$  le vecteur  $n$  dimensionnel d'un état qu'on cherche à estimer. On fait l'hypothèse que les matrices  $\mathbf{R}_t, \mathbf{Q}_t$  sont connues. Tous les opérateurs (modèle, observation, erreur de modèle, erreur d'observation) peuvent dépendre du temps, et sont donc indexés par  $t$ .

Un cycle d'assimilation, qui consiste en deux étapes, une de prévision ( $f$  pour *forecast*), puis une phase d'analyse ( $a$  pour *analysis*), suit les équations du filtre de Kalman :

$$\text{KF - forecast : } \begin{cases} \mathcal{X}_t^f = \mathbf{M}_t \mathcal{X}_{t-1}^a & (1a) \\ \Sigma_t^f = \mathbf{M}_t \Sigma_{t-1}^a \mathbf{M}_t^T + \mathbf{Q}_t & (1b) \end{cases}$$

$$\text{KF - analysis : } \begin{cases} \mathbf{K}_t = \Sigma_t^f \mathbf{H}_t^T (\mathbf{H}_t^T \Sigma_t^f \mathbf{H}_t + \mathbf{R}_t)^{-1} & (2a) \\ \mathcal{X}_t^a = \mathcal{X}_t^f + \mathbf{K}_t (\mathcal{Y}_t - \mathbf{H}_t \mathcal{X}_t^f) & (2b) \\ \Sigma_t^a = (\mathbf{I}_n - \mathbf{K}_t \mathbf{H}_t) \Sigma_t^f & (2c) \end{cases}$$

En résumé,  $\mathcal{X}_t^f$  est la prévision fournie par le modèle  $\mathbf{M}_t$ ,  $\mathcal{X}_t^a$  est la correction de celle-ci obtenue après prise en compte judicieuse grâce au gain de Kalman  $\mathbf{K}_t$  des observations  $\mathcal{Y}_t$ .  $\mathbf{K}_t$  le gain de Kalman détermine le niveau de prise en compte des observations en faisant la balance entre erreurs de prévision, erreur de modèle et erreur d'observation. La matrice de covariance d'erreur de prévision  $\Sigma_t^f$  est finalement mise à jour en fonction de la matrice d'erreur d'observation (2c). Les coûts numériques trop importants de son calcul (1b) ainsi que de sa mise à jour (2c) limite l'utilisation du filtre de Kalman dans les applications réelles en grande dimension. Voyons à présent les mécanismes du filtre de Kalman d'ensemble qui offrent une approximation du filtre de Kalman utilisable dans les problèmes de grande dimension.

### 1.2.2 Le filtre de Kalman d'ensemble (EnKF)

Dans les modèles géophysiques, l'opérateur  $\mathbf{M}$  n'est en général pas linéaire et la dynamique

$$\partial_t \mathcal{X} = \mathcal{M}(t, \mathcal{X}) \quad (3)$$

décrit l'évolution de l'état qui appartient en général à un espace de grande dimension ( $10^6$  à  $10^7$ ).

Le filtre de Kalman d'ensemble est alors une extension non linéaire du filtre de Kalman qui consiste en une approximation de la matrice de covariance d'erreur de prévision  $\Sigma_t^f$  par la matrice de covariance empirique d'un échantillon, appelé "ensemble", de prévisions  $(\tilde{\mathcal{X}}_t^i)_{i=1,\dots,N}$  suivant la dynamique :

$$\text{EnKF : } \begin{cases} \mathcal{X}_0^i \sim \mathcal{N}(\mathcal{X}^b, \Sigma^b) & (4a) \\ \mathcal{X}_t^{f,i} = \mathcal{M}_{t-1 \rightarrow t}(\mathcal{X}_{t-1}^{a,i}) + \mathbf{w}_t^i & (4b) \\ \mathcal{X}_t^{a,i} = \mathcal{X}_t^{f,i} + \mathbf{K}_t (\mathcal{Y}_t + \mathbf{v}_t^i - \mathbf{H}_t \mathcal{X}_t^{f,i}) & (4c) \end{cases}$$

où les  $\mathbf{v}_t^i$  correspondent à des perturbations des observations.

Notons que cette méthode d'assimilation présente des avantages importants : elle permet de considérer des modèles non linéaires sans difficulté, elle est très simple à mettre en oeuvre puisqu'il ne suffit de faire que plusieurs fois une même prévision, de ce fait elle est parallélisable, ce qui permet une réduction considérable des temps de calcul. Cependant, les ensembles utilisés en opérationnel sont en général de petite taille (une douzaine de membres, parfois une centaine) conduisant à un fort bruit d'échantillonnage. De plus, pour rester compétitives en opérationnel, les prévisions sont souvent réalisées à basse résolution spatiale.

Nous illustrons à présent l'utilisation de ce filtre dans le cas d'une dynamique chimique simplifiée : le modèle de Lotka-Volterra.

### 1.3 Illustration avec un modèle chimique jouet : Lotka-Volterra

Pour s’initier à l’assimilation multivariée, on introduit un modèle suffisamment compliqué pour représenter le multivarié et les interactions non-linéaires rencontrées en chimie, et suffisamment simple pour limiter le coût de calcul et faciliter la compréhension. Pour cela notre choix s’est focalisé sur le modèle de Lotka-Volterra qui décrit l’interaction non-linéaire de deux espèces chimiques.

Dans cette section, nous introduisons les équations de cette dynamique puis une illustration de son assimilation avec l’EnKF.

#### 1.3.1 Le modèle chimique de Lotka-Volterra

On considère quatre espèces chimiques  $A, B, X$  et  $Y$  soumises aux réactions chimiques suivantes :



Cette suite de réactions revient à la réaction  $X \rightarrow Y$ . La dynamique des concentrations de  $A$  et  $B$  peut-être modélisée par le système d’équations différentielles suivant :

$$\frac{d[A]}{dt} = k_1[X][A] - k_2[A][B] \quad (6a)$$

$$\frac{d[B]}{dt} = k_2[A][B] - k_3[B] \quad (6b)$$

où  $[A]$  désigne la concentration de l’espèce chimique  $A$ . Si la concentration dans l’atmosphère de l’espèce chimique  $X$  est constante (c’est-à-dire en excès), le système se réécrit :

$$\text{LV} : \begin{cases} \frac{d[A]}{dt} = k_1[A] - k_2[A][B] \\ \frac{d[B]}{dt} = k_2[A][B] - k_3[B] \end{cases} \quad (7a)$$

$$(7b)$$

Les paramètres  $k_1, k_2$  et  $k_3$  correspondent respectivement aux vitesses de reproduction de  $A$ , de réaction pour les réactifs  $A$  et  $B$ , et de désintégration de  $B$  en  $Y$ . On considèrera arbitrairement que  $k_1 = 0.075, k_2 = 0.065$  et  $k_3 = 0.085$  pour notre modèle  $\mathcal{M}$ .

**Méthode de résolution numérique :** afin de simuler le système (7) qui est un problème de Cauchy, on développe une classe en Python capable de simuler tous les problèmes de cette forme. On résout numériquement ces équations en appliquant au choix les méthodes d’Euler explicite, Runge-Kutta d’ordre 2 ou 4.

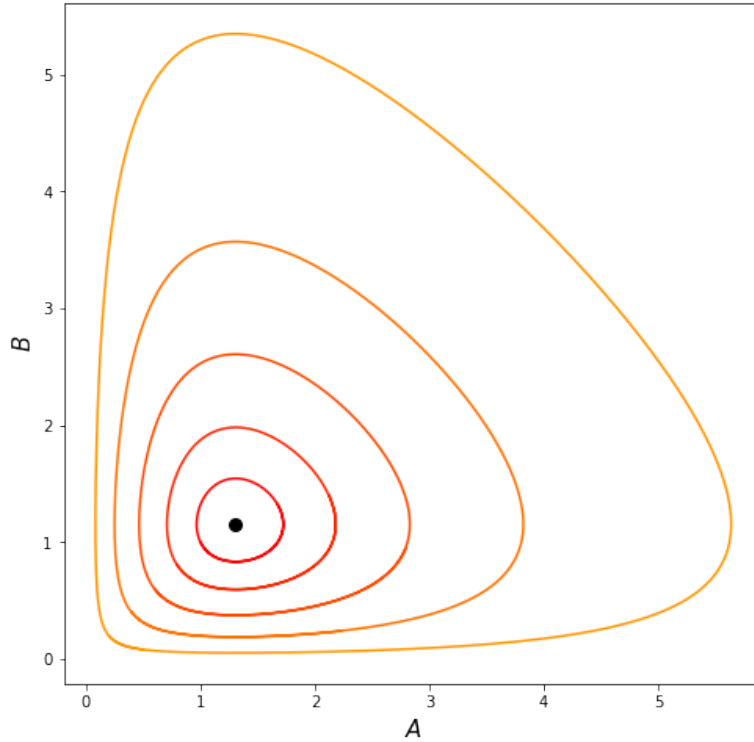


FIGURE 1 – Exemples de simulations.

Les trajectoires dans le plan de phase de Lotka-Volterra correspondent à des orbites non circulaires dans le cadran  $\{\mathbf{x} > 0, \mathbf{y} > 0\}$  et autour du point stationnaire  $\left(\frac{k_3}{k_2}, \frac{k_1}{k_2}\right)$ .

### 1.3.2 Exemple d'application de l'EnKF avec Lotka-Volterra

À présent, on souhaite simuler des cycles de prévision/assimilation à l'aide de l'EnKF. Notre modèle de prévision  $\mathcal{M}$  reste inchangé, mais l'on considère par souci de réalisme qu'il ne représente pas les vraies lois de la nature  $\mathcal{N}$ . Dans un esprit de simplification, on suppose que  $\mathcal{N}$  est un modèle de type Lotka-Volterra dont les paramètres diffèrent légèrement de ceux de notre modèle  $\mathcal{M}$ .

On simule alors quelques cycles de prévision/assimilation. Pour cela, on considère un ensemble de 400 membres, un état initial  $\mathcal{X}^b$  et matrice de covariance initiale  $\Sigma^b$  tous deux connus. À intervalles réguliers, on recueille une observation de la réalité dont on connaît le bruit. Finalement, on applique les équations (4).

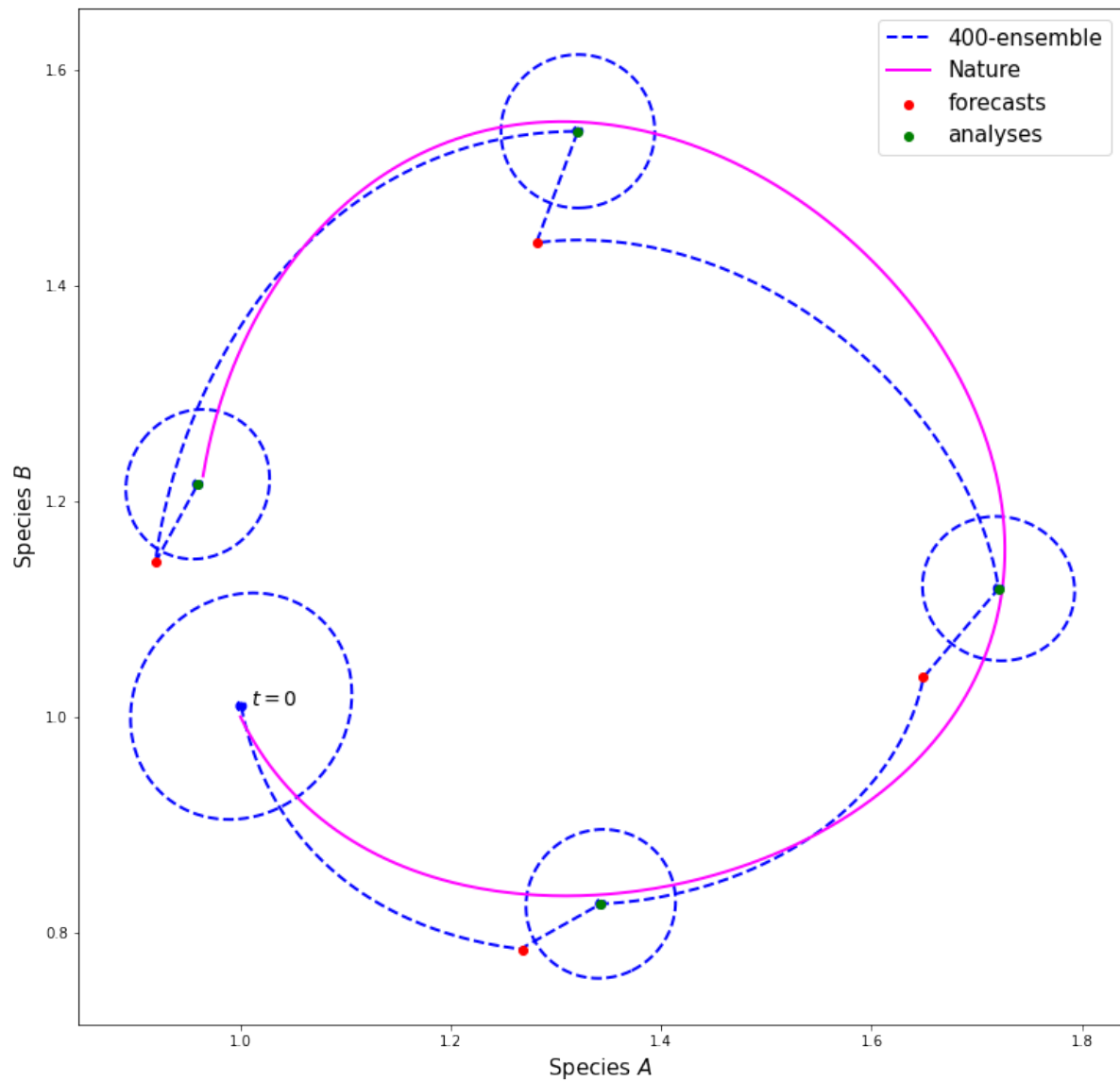


FIGURE 2 – Simulation de 4 cycles d'assimilation avec l'EnKF.

Les étapes d'assimilation des observations nous permettent de rester sur une trajectoire raisonnablement proche de la réalité.



## 1.4 Une nouvelle méthode en assimilation de données : le filtre de Kalman paramétrique (PKF)

### 1.4.1 Principe

Après avoir évoqué (section 1.2.2) la nécessité de considérer les méthodes d'ensembles comme alternative au filtre de Kalman classique, il convient de dire que le filtre de Kalman d'ensemble souffre à son tour de certaines limites.

De part sa nature, il est sujet à un bruit d'échantillonnage d'autant plus important que les ensembles considérés sont de petites tailles (de 30 à 100 membres). Ce bruit d'échantillonnage vient donc polluer les diagnostics des prévisions (moyenne empirique des prévisions) et des matrices de covariances empiriques d'erreurs de prévision. Pour pallier à ces défauts, des méthodes de localisation ont été mises au point afin d'éviter que le bruit d'échantillonnage ne viennent indiquer des corrélations entre des paires de points trop distants sur la carte. A l'inverse, les diagnostics effectués sur l'ensemble peuvent être grandement erronés par l'erreur de modèle, en pratique inconnue, comme nous le verrons dans la section 2.4. Enfin, en raison de la trop grande dimension des problèmes, il est impossible de manipuler directement les matrices de covariance. Estimer correctement l'erreur de prévision est pourtant essentiel afin de fournir des prévisions de la meilleure qualité possible.

Le PKF (Pannekouce et al. [2016]) est une nouvelle méthode d'assimilation qui vient décrire la dynamique des incertitudes d'une manière novatrice : les matrices de covariance y sont approchées par des modèles de covariance reposant sur un jeu léger de paramètres. Le filtre de Kalman paramétrique consiste alors en la description de la dynamique de ces paramètres afin d'ensuite estimer les fonctions de covariance et permettre l'assimilation des observations. On illustre en suivant le PKF dans un cadre adimensionnel et le compare à l'EnKF, afin de mieux en saisir les apports.

### 1.4.2 Illustration du PKF pour la prévision avec Lotka-Volterra

Dans un cadre bivarié (à deux espèces chimiques) et adimensionnel, on illustre le PKF. Dans un tel contexte, l'état que l'on cherche à prévoir est de la forme  $(A^f, B^f)$ , et sa matrice de covariance  $\Sigma^f$  : s'écrit :

$$\Sigma^f = \begin{pmatrix} V_A^f & V_{AB}^f \\ V_{AB}^f & V_B^f \end{pmatrix} \quad (8)$$

Ainsi, l'incertitude associée la prévision est caractérisée par les 3 quantités  $V_A^f, V_B^f$  et  $V_{AB}^f$ . Afin d'estimer la dynamique de ces paramètres, on procède par décomposition de Reynolds. On décompose un état initial  $(A, B)$  comme étant un état moyen perturbé, c'est-à-dire :  $(A, B) = (\bar{A}, \bar{B}) + (\varepsilon_A, \varepsilon_B)$  où  $(\bar{\cdot})$  désigne la moyenne, l'espérance. On peut ensuite définir les paramètres d'incertitude de  $\Sigma^f$  par  $(V_A, V_B, V_{AB}) = (\overline{\varepsilon_A^2}, \overline{\varepsilon_B^2}, \overline{\varepsilon_A \varepsilon_B})$ . Enfin, en utilisant les équations du problème (7) et des manipulations algébriques, on parvient au système suivant :

$$\text{PKF-LV} : \begin{cases} \partial_t A = k_1 A - k_2 AB - k_2 V_{AB} & (9a) \\ \partial_t B = -k_3 B + k_2 AB + k_2 V_{AB} & (9b) \\ \partial_t V_A = 2[V_A(k_1 - k_2 B) - k_2 A V_{AB}] + k_2 \overline{\varepsilon_A^2 \varepsilon_B} & (9c) \\ \partial_t V_B = 2[V_B(-k_3 + k_2 A) + k_2 B V_{AB}] + k_2 \overline{\varepsilon_B^2 \varepsilon_A} & (9d) \\ \partial_t V_{AB} = V_{AB}(k_1 - k_2 B - k_3 + k_2 A) + k_2 V_A B - k_2 V_B A + k_2 \overline{\varepsilon_A^2 \varepsilon_B} - k_2 \overline{\varepsilon_A \varepsilon_B^2} & (9e) \end{cases}$$

Notez que dans toutes les équations (9), on entend par  $A$  et  $B$  les états moyens : par souci de lisibilité, on omettra toujours le symbole  $\overline{(\cdot)}$  pour désigner les états moyens dans les systèmes du PKF. Ensuite, les termes en rouge correspondent à des termes non fermés d'ordre 3. Le PKF se limite à une description des moments d'ordre au plus 2 (tangent-linéaire) des paramètres d'incertitudes, ainsi ces termes seront négligés.

Enfin, le calcul à la main de la dynamique des paramètres du PKF peut vite s'avérer fastidieux et source d'erreur, en particulier dès que l'on considère des problèmes d'une plus grande complexité. Pour cette raison a été développé un package en langage Python, nommé SymPKF (Pannekoucke and Arbogast [2021]), dont nous nous servons dans la suite de ce rapport, et qui se base sur du calcul symbolique pour déterminer de manière automatisée la dynamique du système du PKF pour un problème donné.

On compare à présent les prévisions fournies par le PKF avec celle fournies par un ensemble de taille 1600 (erreur d'échantillonnage 2.5%). Initialement, on considère que  $\mathbf{x}_0 = (1, 1)$  et  $\Sigma_0 = 0.01 \mathbf{I}_2$  et l'on génère l'ensemble initial selon la loi  $\mathcal{N}(\mathbf{x}_0, \Sigma)$ . La simulation s'arrête à  $t = 7$  et utilise un schéma d'Euler explicite avec un pas de temps de 0.01.

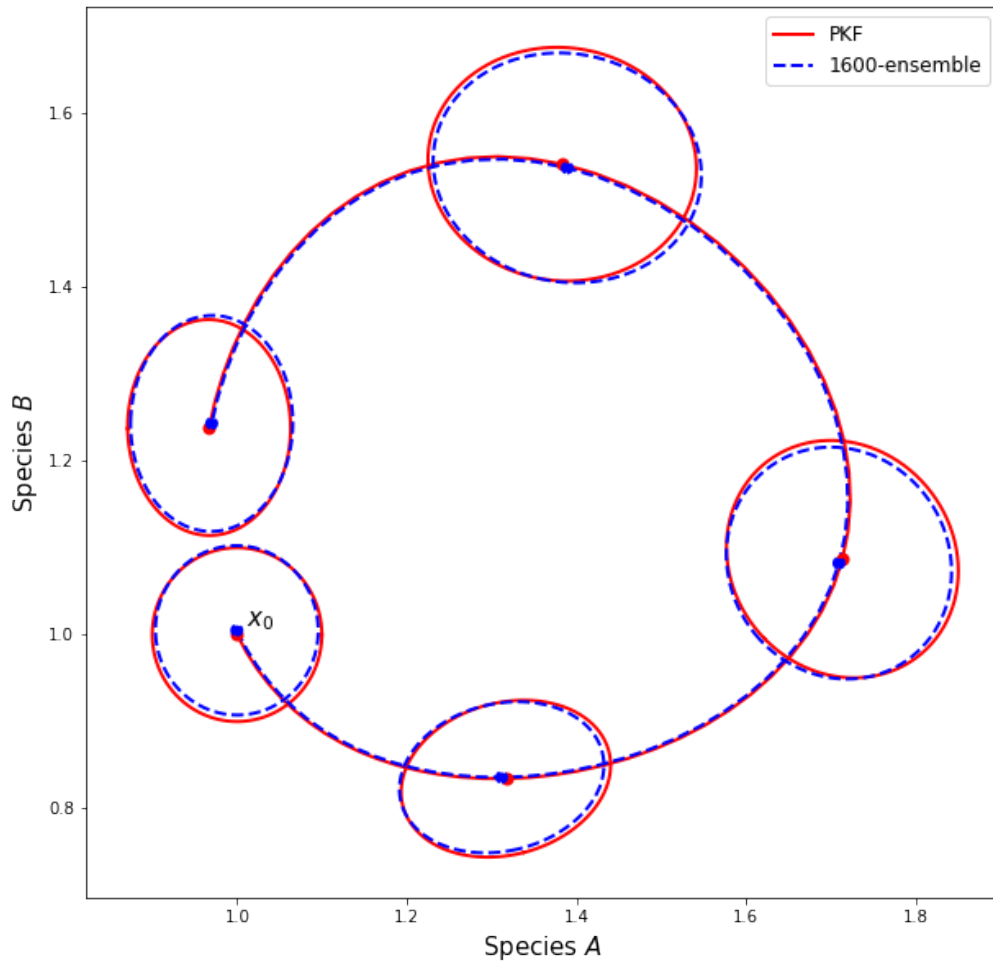


FIGURE 3 – PKF et EnKF - prévisions pour le système de Lotka-Volterra.

Le PKF réalise ici des prévisions de l'état moyen et des statistiques d'erreurs parfaitement recevables. Notons qu'en plus d'être de qualité comparable à celle d'un grand ensemble, le PKF et dans ce cas environ 1600 fois moins coûteux en calculs.

## 2 Transport 1D d'un polluant dans l'atmosphère

A présent, nous nous intéressons au transport par le vent (noté  $u$ ) d'un polluant  $A$  (une espèce chimique : *univarié*) dans l'atmosphère. On considère ici le cas d'une espèce chimique n'interagissant pas avec les autres composants chimique de l'atmosphère, ce peut être le cas par exemple des cendres volcaniques. On modélise ici l'espace dans lequel évolue ce polluant par le segment  $]0, 1[$ , que l'on périodise. La concentration du polluant  $A$  suit alors la dynamique suivante :

$$(Adv) : \partial_t A + u \partial_x A = -A \partial_x u \quad (10)$$

Disposant d'une connaissance *a priori* de l'état du système  $\mathcal{X}^b$  et de l'erreur lui étant associée  $\Sigma^b$ , il s'agit de fournir une prévision de l'état du système et l'erreur :  $\mathcal{X}^f$  et  $\Sigma^f$ . Considérant maintenant un problème spatialisé (1D) un nouveau paramètre est nécessaire au PKF afin de modéliser correctement les fonctions de covariance entre les points de l'espace. Ceci est fait en détail dans la partie suivante.

### 2.1 Formuler le PKF

Dans ce cadre 1D univarié, la modélisation des fonctions de covariance repose sur un champ de variance  $V_{\mathbf{x}}$ , et un champ de longueurs de portées  $L_{\mathbf{x}}$ , tous deux définis en tout point  $\mathbf{x}$  du domaine. Le champ de variance  $V_{\mathbf{x}}$  donne l'amplitude des erreurs en chaque point d'espace, et la longueur de portée elle caractérise l'étendue de la fonction de corrélation en un point de l'espace. On peut faire l'analogie entre la longueur de portée et le rôle de l'écart-type dans la forme d'une gaussienne. Notons qu'en dimension supérieure (2D, 3D), cette longueur de portée devient ce que l'on nomme le tenseur d'aspect noté  $s_{\mathbf{x}}$ , lui même relié au tenseur métrique noté  $g_{\mathbf{x}}$  par  $g_{\mathbf{x}} = s_{\mathbf{x}}^{-1}$ ; là encore l'analogie avec la matrice de covariance d'une gaussienne multi-dimensionnelle peut-être faite pour appréhender le rôle du tenseur d'aspect. Le champ des tenseurs d'aspect est plus couramment désigné comme la "carte d'anisotropies". En 1D, la longueur de portée  $L_{\mathbf{x}}$  est reliée au tenseur d'aspect (qui est en fait un scalaire) par  $L_{\mathbf{x}} = \sqrt{s_{\mathbf{x}}}$ . Nous verrons section 2.2 comment modéliser les fonctions de covariance à l'aide du champ de variance et de la carte d'anisotropies.

Le tenseur métrique, en dimension supérieure ou égale à 1, est relié au champ d'erreur  $\varepsilon$  (Daley [1991]) par :

$$g_{ij}(\mathbf{x}) = \mathbb{E} \left[ \partial_{\mathbf{x}^i} \left( \frac{\varepsilon}{\sqrt{V}} \right) \partial_{\mathbf{x}^j} \left( \frac{\varepsilon}{\sqrt{V}} \right) \right] (\mathbf{x}) \quad (11)$$

On note parfois  $\tilde{\varepsilon} = \varepsilon / \sqrt{V}$  l'erreur normalisée et  $\sigma = \sqrt{V}$  l'écart-type. Dans le cas 1D, cette définition se réécrit :

$$g_{\mathbf{x}} = \mathbb{E} \left[ \left( \partial_{\mathbf{x}} \left( \frac{\varepsilon}{\sqrt{V}} \right) \right)^2 \right] (\mathbf{x}). \quad (12)$$

Pour ce problème de transport univarié, le PKF est alors constitué des paramètres d'état moyen, de champ de variance et de carte d'anisotropies :  $(A, V, s)$ . Afin de déterminer la dynamique de ces paramètres, nous utilisons le package SymPKF évoqué section 1.4.2.

$$\text{PKF-Advection : } \begin{cases} \partial_t A + u \partial_x A = -A \partial_x u & (13a) \\ \partial_t V + u \partial_x V = -2V \partial_x u & (13b) \\ \partial_t s + u \partial_x s = 2s \partial_x u & (13c) \end{cases}$$

Le système du PKF est entièrement découplé, et exact : aucun problème de fermeture. Il nous permettra de ce fait de mettre en perspective les biais qui peuvent apparaître dans les diagnostics fournis par des ensembles lors des simulations numériques. Nous considérerons des ensembles de taille 1600, avec un bruit d'échantillonnage de l'ordre de 2.5% afin de comparer les deux méthodes lors d'une étape de prévision à horizon lointain.

## 2.2 Modèle de covariance univarié

Afin d'être en mesure d'assimiler des observations après une étape de prévision, il faut être en capacité d'estimer l'erreur associée à la prévision que l'on vient de réaliser. Dans le formalisme de l'EnKF, la matrice de covariance d'erreur de prévision est estimée empiriquement à partir de l'ensemble des prévisions  $(\mathcal{X}_i^f)_{i=1,\dots,N_e}$  où  $N_e$  est la taille de l'ensemble. On obtient alors une matrice  $\Sigma^f$  de taille  $N_x \times N_x$  dans laquelle l'entrée  $\Sigma_{i,j}^f$  correspond à la covariance prévue entre les points de grille  $\mathbf{x}_i$  et  $\mathbf{x}_j$ .

Dans le formalisme du PKF, cette covariance est estimée à l'aide d'un modèle de covariance qui repose sur le champ de variance  $V$  et la carte d'anisotropies  $s$ . D'après le théorème de Paciorek et Schervish (Paciorek and Schervish [2006]), pour une fonction de corrélation homogène  $\rho^h(\|\mathbf{x} - \mathbf{y}\|)$  dans  $\mathbb{R}^d$  de tenseur d'aspect  $\mathbf{I}_d$  et un champ différentiable de tenseurs définis positifs  $\mathbf{s}$ , il est possible de former une fonction de corrélation hétérogène définie par :

$$\rho(\mathbf{s})(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{s}_\mathbf{x}|^{1/4} |\mathbf{s}_\mathbf{y}|^{1/4}}{|\frac{1}{2}(\mathbf{s}_\mathbf{x} + \mathbf{s}_\mathbf{y})|^{1/2}} \rho^h\left(\|\mathbf{x} - \mathbf{y}\|_{[\frac{1}{2}(\mathbf{s}_\mathbf{x} + \mathbf{s}_\mathbf{y})]}^{-1}\right) \quad (14)$$

pour laquelle le tenseur d'aspect en  $\mathbf{x}$  est approximativement  $\mathbf{s}_\mathbf{x}$ , et où  $|\cdot|$  est le déterminant matriciel. Nous considérons un modèle de corrélation gaussien, le modèle de covariance univarié s'écrit donc :

$$\mathbf{P}^{\text{he.gauss}}(V, \mathbf{s})(\mathbf{x}, \mathbf{y}) = \sqrt{V_\mathbf{x} V_\mathbf{y}} \frac{|\mathbf{s}_\mathbf{x}|^{1/4} |\mathbf{s}_\mathbf{y}|^{1/4}}{|\frac{1}{2}(\mathbf{s}_\mathbf{x} + \mathbf{s}_\mathbf{y})|^{1/2}} \exp\left(-\frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_{[\frac{1}{2}(\mathbf{s}_\mathbf{x} + \mathbf{s}_\mathbf{y})]}^2\right) \quad (15)$$

À l'aide de ce modèle de covariance, nous sommes en mesure de fournir une approximation de la matrice de covariance associée à une prévision reposant sur les paramètres du PKF.

## 2.3 Simulation numérique

### 2.3.1 Paramètres de simulation

**Paramètres communs à l'EnKF et au PKF :** Initialement, on fixe de manière homogène le champ d'état moyen (ou concentration moyenne), de variance, et de longueur de portée à respectivement  $A_0 = 1$ ,  $V_0 = 0.125^2$  et  $s_0 = l_h^2 = (19\Delta x)^2$  (ce qui correspondra finalement à une longueur de portée de 26km). On considère que le segment  $]0, 1[$  est une dilatation d'un segment  $]0, D[$  où  $D = 1000\text{km}$ . Le champ de vent, constant dans le temps, est défini par  $u(\mathbf{x}) = (35 + 15 \cos(2\pi\mathbf{x}))/D$ , modélisant un vent d'intensité moyenne  $35\text{km}\cdot\text{h}^{-1}$  avec une intensité maximale  $U_{max} = 50\text{km}\cdot\text{h}^{-1}$  (figure 4). On définit ensuite le temps caractéristique  $T$  comme étant le temps moyen nécessaire à un polluant pour effectuer une révolution autour du domaine, *i.e.*  $T = D/\bar{u}$ . L'horizon de prévision  $t_{end}$  est fixé à  $3T$  ( $\simeq 85\text{h}$ ).

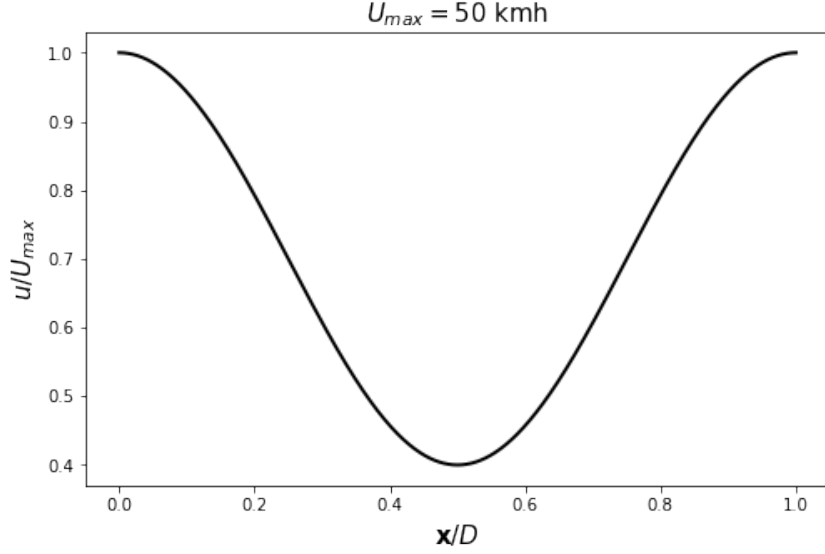


FIGURE 4 – champ de vent utilisé sur le domaine périodique.

**Schéma numérique employé :** Le segment  $]0, 1[$  est discrétisé en  $N_x = 723$  points de grille. Les dérivés spatiales sont approchées par différences finies avec un schéma centré. Un schéma explicite Runge-Kutta d'ordre 4 (RK4) est utilisé pour l'intégration temporelle. Le pas de temps  $\Delta t$  considéré est déterminé par  $\Delta t = \Delta x / U_{max}$ , où  $\Delta x = 1/N_x$  est la taille d'une maille (1.38km).

### 2.3.2 Génération des champs d'erreurs aléatoires

Afin de générer des champs aléatoires pour initialiser un ensemble, on procède comme suit. Étant donné les paramètres d'incertitude initiale  $V_0$  et  $l_h$ , on définit la fonction de covariance  $c_h$  homogène par :

$$c_h(\mathbf{x}, \mathbf{y}) = V_0 \exp \left( -\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2l_h^2} \right) \quad (16)$$

où l'on a périodisé la distance euclidienne  $\|\cdot\|_2$  sur le segment  $]0, 1[$ . Soit ensuite  $N_x$  le nombre de points de grille de la discrétisation du segment  $]0, 1[$ , on définit la matrice  $\mathbf{P}^h$  par :

$$\mathbf{P}_{i,j}^h = \rho_h(\mathbf{x}_i, \mathbf{x}_j)$$

Enfin, on définit le champ aléatoire  $\varepsilon_h$  par :

$$\varepsilon_h = \mathbf{P}_h^{1/2} \varepsilon$$

où  $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_x})$ .

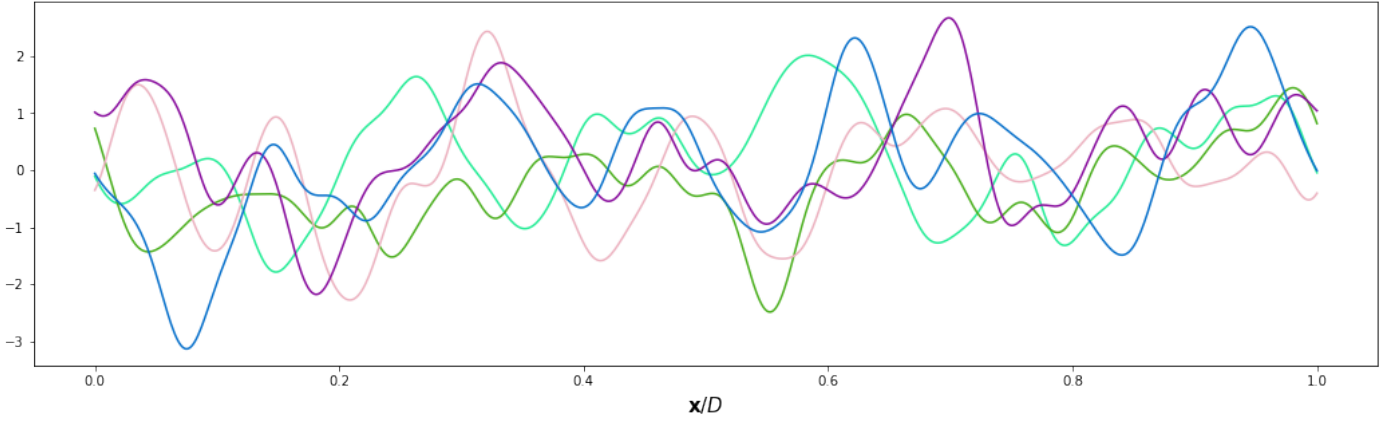


FIGURE 5 – Exemples de champs d’erreurs aléatoires (périodiques),  $l_h = 19\Delta x$ .

### 2.3.3 Résultats : comparaison des prévisions du PKF et de l’EnKF

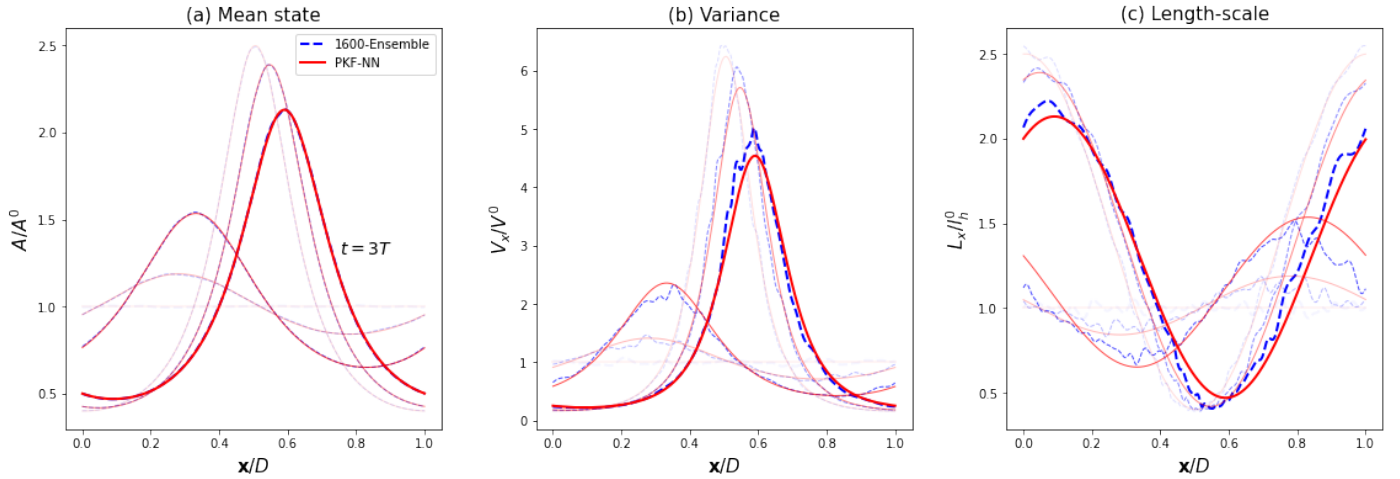


FIGURE 6 – Comparaison des prévisions du PKF et diagnostics de l’EnKF.

Les deux méthodes fournissent des prévisions similaires pour la concentration moyenne. Le champ de vent  $u$  que l’on s’est fixé plus haut entraîne une accumulation de la matière dans la région où le vent est le moins intense, le parallèle peut-être fait avec une situation d’embouteillage.

Les courbes de variance sont globalement identiques. L’incertitude a tendance à se concentrer dans les zones d’accumulations de matière.

Enfin, pour les longueurs de portées, on remarque une plus grande sensibilité à l’erreur d’échantillonnage pour l’ensemble ; les prévisions sont néanmoins identiques. Les incertitudes sont plus localisées là où la matière est en plus forte concentration.

### 2.3.4 Résultats : comparaison des erreurs de prévisions pour le PKF et l'EnKF.

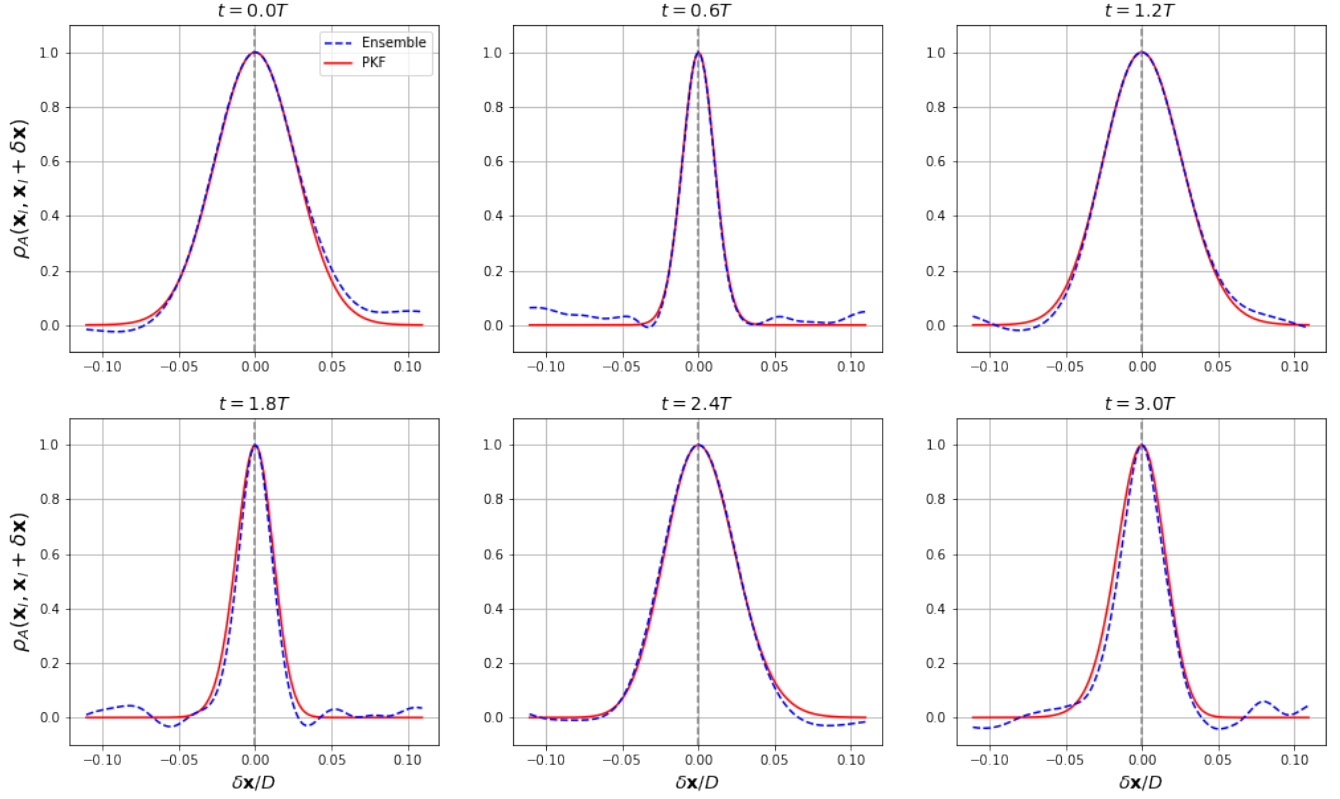


FIGURE 7 – Fonctions de corrélations à  $\mathbf{x}_l = 0.5$  et  $t = [0.0, 0.6, 1.2, 1.8, 2.4, 3.0]T$ . Affichage pour  $\delta\mathbf{x} \in [-0.1, 0.1]$ .

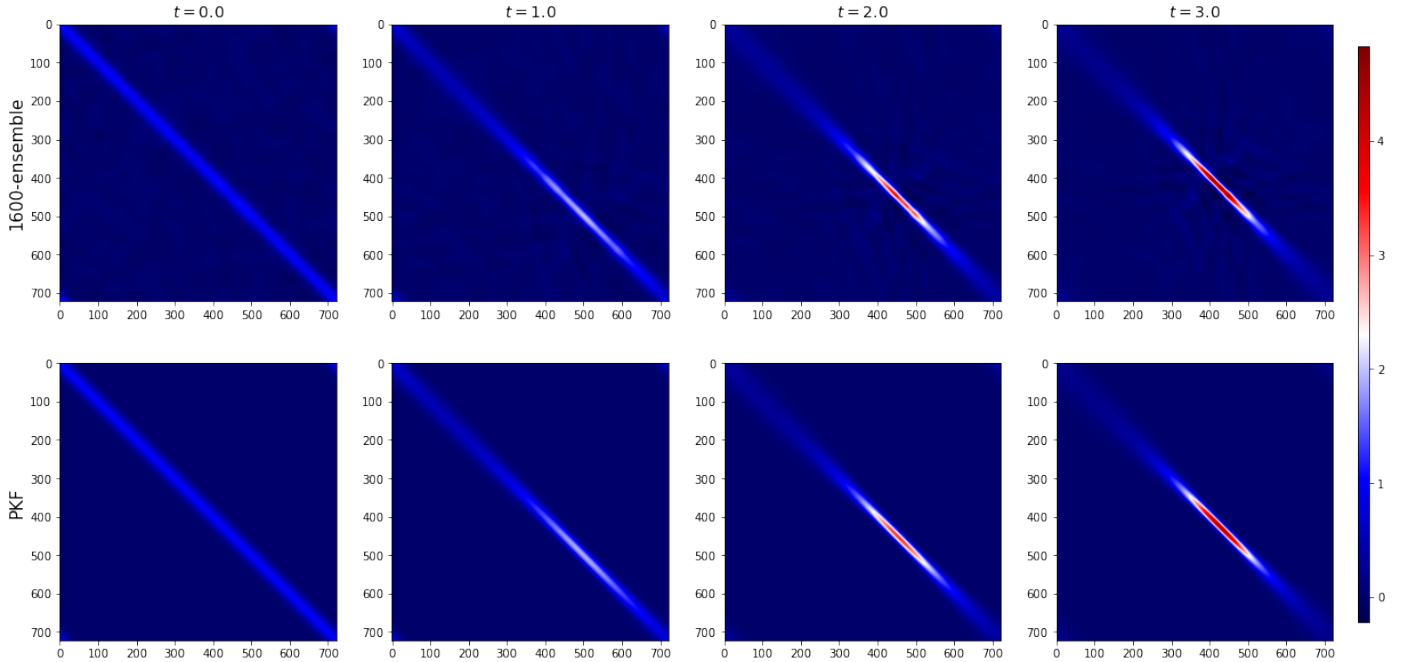


FIGURE 8 – Matrices de covariance d'erreur de prévision à  $t = [0.0, 1.0, 2.0, 3.0]T$ .

## 2.4 Influence de l'erreur de modèle sur les diagnostics de l'EnKF

### 2.4.1 Mise en évidence d'un cas de divergence du filtre de Kalman d'ensemble

On souhaite ici mettre en avant l'influence que peut avoir l'erreur de modèle sur la qualité des diagnostics fournis par un ensemble. Les résultats obtenus précédemment contenaient relativement peu d'erreur : on avait quasi-superposition des courbes d'ensemble avec celles du PKF (exactes). Cela était dû au schéma centré d'ordre 2 utilisé pour l'estimation des dérivées spatiales. Nous allons maintenant approximer les dérivés spatiales non plus par un schéma centré, mais par un schéma *upwind*, implicitement diffusif et d'ordre 1. On met ensuite en perspective les diagnostics de l'ensemble avec ceux fournis par le PKF.

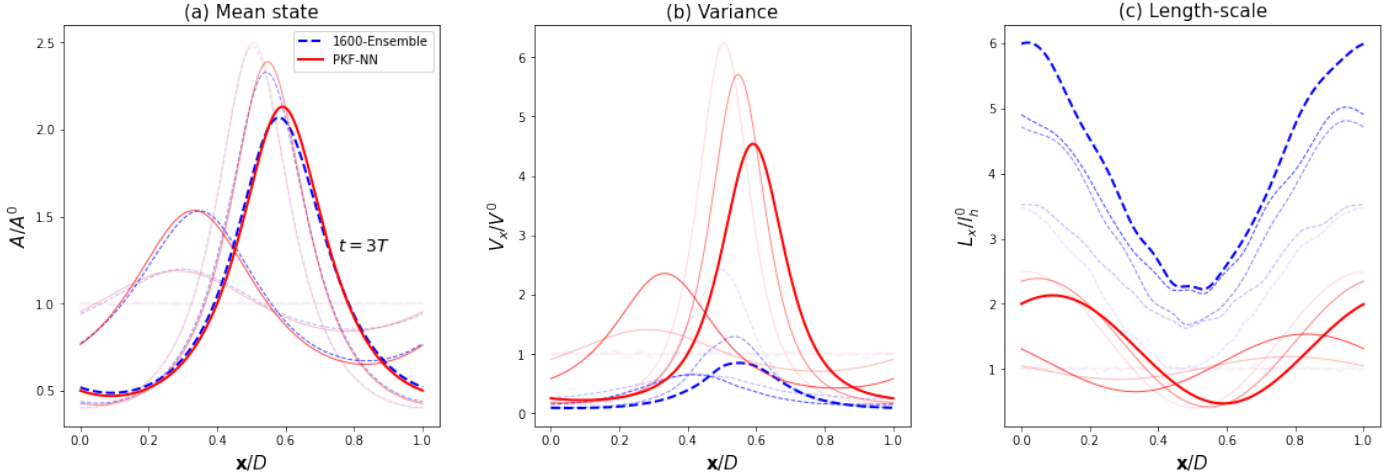


FIGURE 9 – Comparaison des prévisions du PKF et diagnostics de l'EnKF. La prévision de l'EnKF est réalisée avec un schéma upwind explicite.

On voit ici que le diagnostic obtenu pour la variance indique une incertitude significativement sous-estimée. Cela s'explique par la diffusion que comporte le schéma upwind explicite, qui a tendance à lisser les trajectoires, les rendant de plus en plus homogènes et ce faisant diminuant la variance. Dans le cas extrême où cette variance deviendrait nulle, on en viendrait à ne pas tenir compte des nouvelles observations lors de l'étape d'analyse. On parle alors de divergence du filtre de Kalman d'ensemble. En d'autres termes, on "perd le signal".

### 2.4.2 Impact de la réduction des longueurs de portées initiales sur les diagnostics

Jusqu'ici les longueurs de portées initiales considérées étaient "adaptées" au niveau de discrétisation, avec  $l_h = 19\Delta x \simeq 26\text{km}$  pour  $N_x = 723$  points de grille. En conservant le même maillage mais en considérant des longueurs de portées initiales plus petites, de l'ordre de  $4\Delta x \simeq 5.5\text{km}$ , nous sommes en mesure de décrire des phénomènes plus localisés, mais s'avère alors source d'une importante erreur de modèle (voir figures 10 et 11). On considère un schéma centré pour approcher les dérivées spatiales, et un schéma RK4 pour l'intégration temporelle. Les résultats sont mis en perspective avec les prévisions du PKF, exactes quelque soit la longueur de portée initiale.



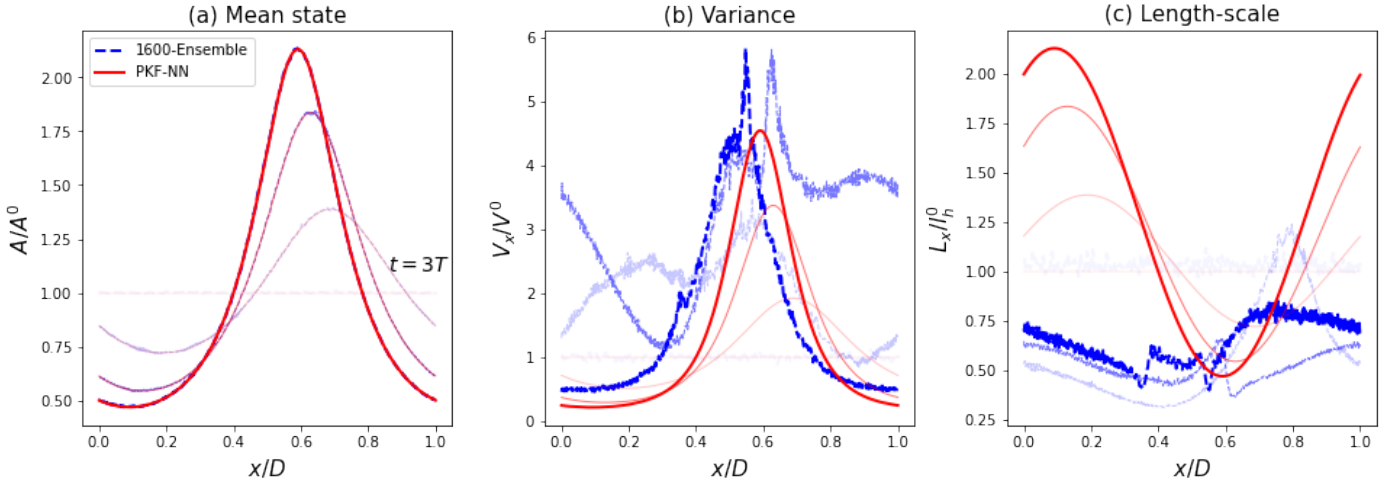


FIGURE 10 – Comparaison des prévisions du PKF et diagnostics de l’EnKF, avec  $l_h = 4\Delta x \simeq 5.5\text{km}$ .

Ces graphiques mettent en évidence le fait que l’erreur de modèle liée à la discrétisation spatio-temporelle, accentuée par les petites longueurs de portées considérées, nous a conduit à des prévisions très erronées pour les statistiques d’erreurs.

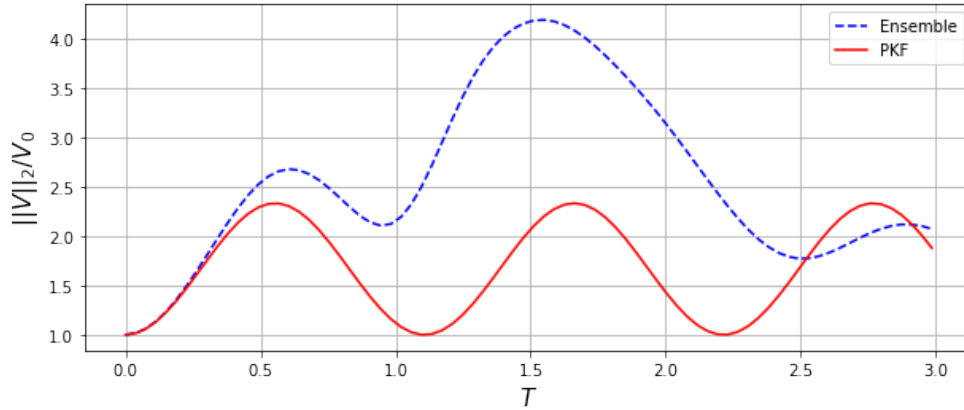


FIGURE 11 – Quantité totale de variance au cours du temps pour l’EnKF et le PKF, avec  $l_h = 4\Delta x \simeq 5.5\text{km}$ .

Ces courbes soulignent l’impact de l’erreur de modèle sur la dynamique de la variance dans le cas de l’EnKF. Le PKF permet de remarquer la périodicité de l’incertitude totale.

## 2.5 Conclusion dans le cadre univarié de l’advection 1D

On peut se rendre compte au travers de cet exemple simple des apports du PKF. D’abord, il permet à des coûts de calculs moindres (1600 fois inférieurs) l’obtention de diagnostics de meilleure qualité que l’ensemble car dépourvu d’erreur d’échantillonnage. Ensuite, grâce à la modélisation des fonctions d’auto-corrélations, l’information normalement contenue dans une matrice de covariance gigantesque est résumée en quelques paramètres légers, nous dispensant du besoin d’une capacité immense de stockage. Enfin, on a vu que le PKF nous a permis de caractériser l’erreur de modèle associé à un schéma de discrétisation spatiale. De ce point de vue, le PKF est un outil d’exploration offrant de nouveaux points de vue.

### 3 Un modèle de chimie transport simplifié pour l'exploration d'une formulation multivariée du PKF

À terme, on aimerait être capable de formuler un modèle de chimie-transport (CTM) reposant sur le formalisme du PKF. Un CTM consiste en la modélisation des interactions chimiques des espèces présentes dans l'atmosphère (120 espèces considérées dans MOCAGE) ainsi que de leur transport par le vent. Pour explorer la possibilité d'une formulation multivariée du PKF, on considère un modèle de chimie transport simplifié, qui contient néanmoins des composantes essentielles à un modèle plus réaliste (non-linéarités, plusieurs espèces, transport). Ce modèle consiste donc en la modélisation du transport dans l'atmosphère de deux espèces chimiques  $A$  et  $B$  dont les interactions chimiques sont modélisées par le système non-linéaire de Lotka-Volterra. Il s'agit donc de conjuguer les processus d'advection (10) et de Lotka-Volterra (7), donnant lieu au système suivant :

$$\text{Advection-LV} : \begin{cases} \partial_t A + u \partial_x A = -A \partial_x u + k_1 A - k_2 AB \\ \partial_t B + u \partial_x B = -B \partial_x u + k_2 AB - k_3 B \end{cases} \quad (17a)$$

$$(17b)$$

#### 3.1 Formulation multivariée du PKF avec le SymPKF

Dans ce contexte, le système du PKF est constitué de sept paramètres : état moyen, champ de variance et carte d'anisotropies pour chaque espèce, ainsi qu'un champ de covariance propre au caractère multivarié. On détermine la dynamique des paramètres du PKF à l'aide du package SymPKF présenté dans la section 1.4.2. On illustre ici la simplicité d'obtention de la dynamique du PKF grâce au package.

```
from sympy import symbols, Function, Derivative
from sympkf.sympkf import PDESystem, Eq, SymbolicPKF, t

x = symbols("x")
k1, k2, k3 = symbols("k_1 k_2 k_3")
u = Function("u")(x) #wind
A = Function("A")(t, x) #tracer A
B = Function("B")(t, x) #tracer B
dynamics = [
    Eq(Derivative(A, t), -u*Derivative(A, x) - A*Derivative(u, x) + k1*A - k2*A*B),
    Eq(Derivative(B, t), -u*Derivative(B, x) - B*Derivative(u, x) - k3*B + k2*A*B)
]
display(dynamics[0]), display(dynamics[1]); 3
```

$$\frac{\partial}{\partial t} A(t, x) = k_1 A(t, x) - k_2 A(t, x) B(t, x) - A(t, x) \frac{d}{dx} u(x) - u(x) \frac{\partial}{\partial x} A(t, x)$$

$$\frac{\partial}{\partial t} B(t, x) = k_2 A(t, x) B(t, x) - k_3 B(t, x) - B(t, x) \frac{d}{dx} u(x) - u(x) \frac{\partial}{\partial x} B(t, x)$$

FIGURE 12 – Définition d'un problème avec le SymPKF.

Toute la partie fastidieuse et sujette à erreur de calcul à la main des dynamiques de chacune des quantités est alors traitée automatiquement par le SymPKF.

```

system = PDESystem(dynamics)
pkf_system = SymbolicPKF(system)
for equation in pkf_system.in_aspect: display(equation)

```

$$\begin{aligned}
\frac{\partial}{\partial t} A(t, x) &= k_1 A(t, x) - k_2 A(t, x) B(t, x) - k_2 V_{AB}(t, x) - A(t, x) \frac{d}{dx} u(x) - u(x) \frac{\partial}{\partial x} A(t, x) \\
\frac{\partial}{\partial t} B(t, x) &= k_2 A(t, x) B(t, x) + k_2 V_{AB}(t, x) - k_3 B(t, x) - B(t, x) \frac{d}{dx} u(x) - u(x) \frac{\partial}{\partial x} B(t, x) \\
\frac{\partial}{\partial t} V_A(t, x) &= 2k_1 V_A(t, x) - 2k_2 A(t, x) V_{AB}(t, x) - 2k_2 B(t, x) V_A(t, x) - u(x) \frac{\partial}{\partial x} V_A(t, x) - 2 V_A(t, x) \frac{d}{dx} u(x) \\
\frac{\partial}{\partial t} V_B(t, x) &= 2k_2 A(t, x) V_B(t, x) + 2k_2 B(t, x) V_{AB}(t, x) - 2k_3 V_B(t, x) - u(x) \frac{\partial}{\partial x} V_B(t, x) - 2 V_B(t, x) \frac{d}{dx} u(x) \\
\frac{\partial}{\partial t} V_{AB}(t, x) &= k_1 V_{AB}(t, x) + k_2 A(t, x) V_{AB}(t, x) - k_2 A(t, x) V_B(t, x) - k_2 B(t, x) V_{AB}(t, x) + k_2 B(t, x) V_A(t, x) - k_3 V_{AB}(t, x) \\
&\quad - \frac{V_{AB}(t, x) u(x) \frac{\partial}{\partial x} V_B(t, x)}{2 V_B(t, x)} - \frac{V_{AB}(t, x) u(x) \frac{\partial}{\partial x} V_A(t, x)}{2 V_A(t, x)} - 2 V_{AB}(t, x) \frac{d}{dx} u(x) - u(x) \sqrt{V_A(t, x)} \sqrt{V_B(t, x)} \mathbb{E} \left( \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right) \\
&\quad - u(x) \sqrt{V_A(t, x)} \sqrt{V_B(t, x)} \mathbb{E} \left( \varepsilon_B(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \right) \\
\frac{\partial}{\partial t} s_{A,xx}(t, x) &= - \frac{2k_2 A(t, x) V_{AB}(t, x) s_{A,xx}(t, x)}{V_A(t, x)} + \frac{2k_2 A(t, x) \sqrt{V_B(t, x)} s_{A,xx}^2(t, x) \mathbb{E} \left( \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right)}{\sqrt{V_A(t, x)}} \\
&\quad + \frac{k_2 A(t, x) s_{A,xx}^2(t, x) \mathbb{E} \left( \varepsilon_B(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \right) \frac{\partial}{\partial x} V_B(t, x)}{\sqrt{V_A(t, x)} \sqrt{V_B(t, x)}} - \frac{k_2 A(t, x) \sqrt{V_B(t, x)} s_{A,xx}^2(t, x) \mathbb{E} \left( \varepsilon_B(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \right) \frac{\partial}{\partial x} V_A(t, x)}{V_A^{\frac{3}{2}}(t, x)} \\
&\quad + \frac{2k_2 \sqrt{V_B(t, x)} s_{A,xx}^2(t, x) \mathbb{E} \left( \varepsilon_B(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \right) \frac{\partial}{\partial x} A(t, x)}{\sqrt{V_A(t, x)}} - u(x) \frac{\partial}{\partial x} s_{A,xx}(t, x) + 2 s_{A,xx}(t, x) \frac{d}{dx} u(x) \\
\frac{\partial}{\partial t} s_{B,xx}(t, x) &= \frac{2k_2 B(t, x) V_{AB}(t, x) s_{B,xx}(t, x)}{V_B(t, x)} - \frac{2k_2 B(t, x) \sqrt{V_A(t, x)} s_{B,xx}^2(t, x) \mathbb{E} \left( \frac{\partial}{\partial x} \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right)}{\sqrt{V_B(t, x)}} \\
&\quad + \frac{k_2 B(t, x) \sqrt{V_A(t, x)} s_{B,xx}^2(t, x) \mathbb{E} \left( \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right) \frac{\partial}{\partial x} V_B(t, x)}{V_B^{\frac{3}{2}}(t, x)} - \frac{k_2 B(t, x) s_{B,xx}^2(t, x) \mathbb{E} \left( \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right) \frac{\partial}{\partial x} V_A(t, x)}{\sqrt{V_A(t, x)} \sqrt{V_B(t, x)}} \\
&\quad - \frac{2k_2 \sqrt{V_A(t, x)} s_{B,xx}^2(t, x) \mathbb{E} \left( \varepsilon_A(t, x, \omega) \frac{\partial}{\partial x} \varepsilon_B(t, x, \omega) \right) \frac{\partial}{\partial x} B(t, x)}{\sqrt{V_B(t, x)}} - u(x) \frac{\partial}{\partial x} s_{B,xx}(t, x) + 2 s_{B,xx}(t, x) \frac{d}{dx} u(x)
\end{aligned}$$

FIGURE 13 – Calcul de la dynamique des paramètres du PKF avec le SymPKF.

Note : dans les notations du package SymPKF,  $\varepsilon$  désigne une erreur normalisée. Nous continuerons de dénoter une erreur de normalisée par  $\tilde{\varepsilon}$  et une erreur par  $\varepsilon$ .

Le système obtenu se réécrit plus simplement :

$$\partial_t A + u \partial_{\mathbf{x}} A = -A \partial_{\mathbf{x}} u + k_1 A - k_2 AB - k_2 V_{AB} \quad (18a)$$

$$\partial_t B + u \partial_{\mathbf{x}} B = -B \partial_{\mathbf{x}} u - k_3 B + k_2 AB + k_2 V_{AB} \quad (18b)$$

$$\partial_t V_{AB} + u \partial_{\mathbf{x}} V_{AB} = -2V_{AB} \partial_{\mathbf{x}} u + V_{AB}(k_1 - k_2 B - k_3 + k_2 A) + k_2 V_A B - k_2 V_B A \quad (18c)$$

$$\partial_t V_A + u \partial_{\mathbf{x}} V_A = -2V_A \partial_{\mathbf{x}} u + 2[V_A(k_1 - k_2 B) - k_2 A V_{AB}] \quad (18d)$$

$$\partial_t V_B + u \partial_{\mathbf{x}} V_B = -2V_B \partial_{\mathbf{x}} u + 2[V_B(-k_3 + k_2 A) + k_2 B V_{AB}] \quad (18e)$$

$$\begin{aligned} \partial_t s_A = & \underbrace{-u \partial_{\mathbf{x}} s_A + 2s_A \partial_{\mathbf{x}} u}_{\text{advection}} - \frac{2k_2 s_A A V_{AB}}{V_A} + \frac{2k_2 A s_A^2 \sigma_A}{\sigma_B} \overline{\partial_{\mathbf{x}} \tilde{\varepsilon}_A \partial_{\mathbf{x}} \tilde{\varepsilon}_B} + \\ & \overline{\tilde{\varepsilon}_B \partial_{\mathbf{x}} \tilde{\varepsilon}_A} \left( -2k_2 A s_A^2 \sigma_B \partial_{\mathbf{x}} \sigma_A - \frac{2k_2 A s_A^2 V_B \partial_{\mathbf{x}} \sigma_B}{\sigma_A} + \frac{2k_2 \sigma_B s_A^2 \partial_{\mathbf{x}} A}{\sigma_A} \right) \end{aligned} \quad (18f)$$

$$\begin{aligned} \partial_t s_B = & \underbrace{-u \partial_{\mathbf{x}} s_B + 2s_B \partial_{\mathbf{x}} u}_{\text{advection}} + \frac{2k_2 s_B B V_{AB}}{V_B} - \frac{2k_2 B s_B^2 \sigma_A}{\sigma_B} \overline{\partial_{\mathbf{x}} \tilde{\varepsilon}_A \partial_{\mathbf{x}} \tilde{\varepsilon}_B} + \\ & \overline{\tilde{\varepsilon}_A \partial_{\mathbf{x}} \tilde{\varepsilon}_B} \left( 2k_2 B s_B^2 \sigma_A \partial_{\mathbf{x}} \sigma_B - \frac{2k_2 B s_B^2 V_A \partial_{\mathbf{x}} \sigma_A}{\sigma_B} - \frac{2k_2 \sigma_A s_B^2 \partial_{\mathbf{x}} B}{\sigma_B} \right) \end{aligned} \quad (18g)$$

Plusieurs remarques. D'abord notons que pour les dynamiques (18a - 18e), il s'agit en fait du couplage du système PKF pour Lotka-Volterra (9) avec le PKF pour l'advection (13). La notation  $\tilde{\varepsilon}$  correspond à l'erreur normalisée  $\tilde{\varepsilon} = \varepsilon/\sigma$ . Les termes en rouge sont exprimés en termes des erreurs et constituent donc des *problèmes de fermeture* pour les dynamiques des tenseurs d'aspect (18f),(18g). Dans ces équations, on peut faire la distinction entre les termes liés uniquement au processus d'advection, comme dans le système (13), des autres termes, liés au modèle chimique. Nous voyons dans la section suivante comment simplifier la dynamique des tenseurs d'aspects afin de résoudre le problème de fermeture.

### 3.2 Simplification du système PKF

En diagnostiquant à partir d'une prévision d'ensemble les paramètres du PKF, l'on a observé (figures 6 et 15) que les diagnostics des tenseurs d'aspects des deux espèces  $A$  et  $B$  étaient semblables à celui du tenseur d'aspect dans le cas univarié sans interaction chimique (eq 13). Cette observation laisse à penser que les tenseurs d'aspects sont majoritairement mus par le processus d'advection.

Afin de vérifier cette intuition, on a donc, à partir d'une prévision d'ensemble, diagnostiqué les trajectoires des tenseurs d'aspects, soit un ensemble de diagnostics  $(s_A(t_k))_k$  et  $(s_B(t_k))_k$  où  $t_k = k\Delta t$ . Ces trajectoires nous permettent ensuite d'établir *a posteriori* la dynamique des tenseurs d'aspect  $(\partial_t s_A(t_k))_k$  et  $(\partial_t s_B(t_k))_k$  en procédant par différences finies. Enfin, on compare ces quantités aux deux composantes de la dynamique globale : celle liée à l'advection  $(-u \partial_{\mathbf{x}} s + 2s \partial_{\mathbf{x}} u)$  et la partie restante liée à la chimie (différence entre globale et advection).

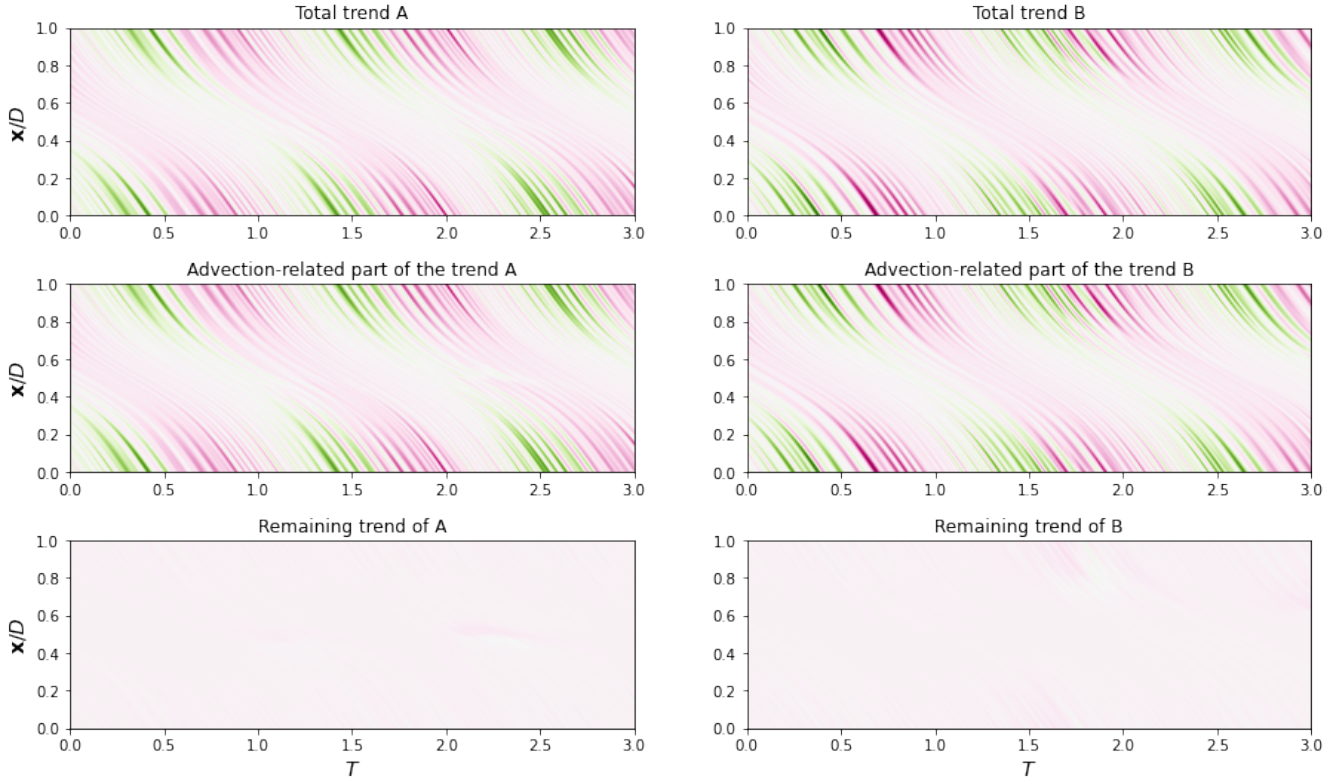


FIGURE 14 – Décomposition de la dynamique des tenseurs d’aspects - les échelles sont identiques pour chacun des graphiques.

On observe bien que l’advection explique en très grande partie la dynamique des tenseurs d’aspects. Ce résultat nous permet alors de simplifier grandement la dynamique des tenseurs d’aspect dans le système du PKF, pour donner lieu au système fermé suivant :

$$\text{PKF-Adv-LV : } \left\{ \begin{array}{l}
 \partial_t A + u \partial_x A = -A \partial_x u + k_1 A - k_2 AB - k_2 V_{AB} \\
 \partial_t B + u \partial_x B = -B \partial_x u - k_3 B + k_2 AB + k_2 V_{AB} \\
 \partial_t V_{AB} + u \partial_x V_{AB} = -2V_{AB} \partial_x u + V_{AB}(k_1 - k_2 B - k_3 + k_2 A) + k_2 V_A B - k_2 V_B A \\
 \partial_t V_A + u \partial_x V_A = -2V_A \partial_x u + 2[V_A(k_1 - k_2 B) - k_2 A V_{AB}] \\
 \partial_t V_B + u \partial_x V_B = -2V_B \partial_x u + 2[V_B(-k_3 + k_2 A) + k_2 B V_{AB}] \\
 \partial_t s_A + u \partial_x s_A = 2s_A \partial_x u \\
 \partial_t s_B + u \partial_x s_B = 2s_B \partial_x u
 \end{array} \right.$$

(19a)

(19b)

(19c)

(19d)

(19e)

(19f)

(19g)

### 3.3 Simulation numérique : une prévision multivariée du système PKF

Le système du PKF (19) étant formulé pour (17), on peut désormais comparer les prévisions fournies par un ensemble avec celles du PKF.

**Paramètres du problème :** Les paramètres de simulation (champ de vent, horizon de simulation, schéma numérique) sont identiques à ceux de la section 2. On définit les vitesses de réactions du modèle chimiques  $k_1, k_2$  et  $k_3$  respectivement à 0.075, 0.065 et 0.085. Également, une erreur initiale  $\varepsilon_0$  se décompose maintenant comme deux champs d'erreurs  $\varepsilon_0 = (\varepsilon_0^A, \varepsilon_0^B)$  que l'on considère décorrélés, *ie*  $V_0^{AB} = 0$ , et qui sont initialisés de la même manière que dans le problème d'advection (*ie* selon (16), avec  $V_0^A = V_0^B = 0.125^2$ , et  $s_0^A = s_0^B = 0.08^2$ ).

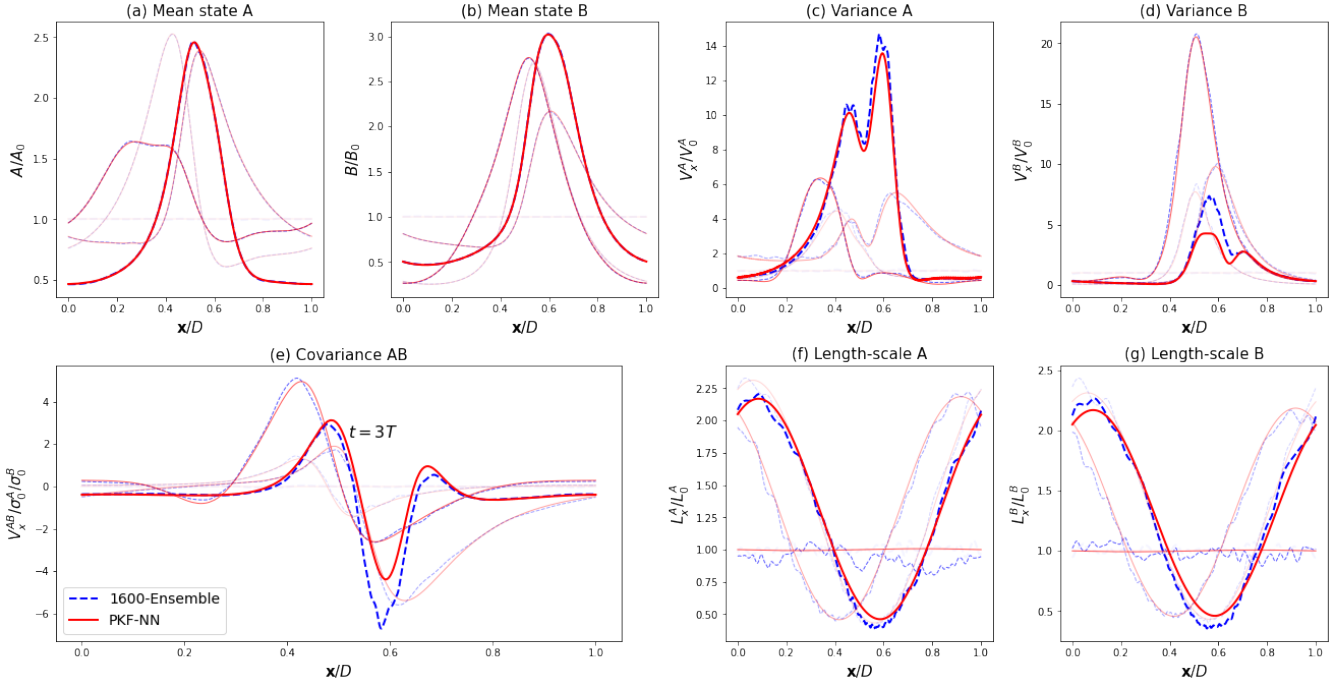


FIGURE 15 – Prévisions pour le système (17). PKF (lignes continues rouge) et EnKF (lignes pointillées bleu).

On observe d'abord que les prévisions pour les concentrations des deux espèces chimiques (panneaux (a) et (b)) sont identiques. Ensuite, les prévisions pour les variances et la covariance (panneaux (c), (d) et (e)) montrent des courbes qui diffèrent davantage, en particulier à  $t = 3T$ , mais dont les dynamiques sont globalement les mêmes. Enfin, on observe que les prévisions du PKF pour les longueurs de portées (panneaux (f) et (g)) coïncident parfaitement avec les diagnostics de l'ensemble, et cela sans bruit d'échantillonnage. Cela justifie la simplification de la dynamique des tenseurs d'aspect dans le système PKF réalisée dans la partie précédente (section 3.2).

Ce résultat démontre la capacité du PKF à réaliser une prévision des statistiques d'erreurs dans un cadre multivarié.

### 3.4 Définition d'un modèle de corrélation multivarié

Étant dans un cadre multivarié, on parle désormais de matrice de corrélation entre deux points de grille  $\mathbf{x}$  et  $\mathbf{y}$ , définie par :

$$\rho(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \rho_A(\mathbf{x}, \mathbf{y}) & \rho_{AB}(\mathbf{x}, \mathbf{y}) \\ \rho_{AB}(\mathbf{x}, \mathbf{y}) & \rho_B(\mathbf{x}, \mathbf{y}) \end{pmatrix} \quad (20)$$

Les termes présents sur la diagonale de cette matrice correspondent aux fonctions de corrélations univariées que l'on sait estimer grâce à la formule (15). La fonction de corrélations croisées  $\rho_{AB}(\mathbf{x}, \mathbf{y})$  n'est en revanche pas définie. Si connue, elle permet lors de l'assimilation d'une observation de l'espèce  $A$ , d'ajuster également la prévision sur l'espèce  $B$ . On utilise donc au maximum l'information contenue dans l'observation d'une espèce.

Dans certains cas, une estimation des covariances croisées peut être réalisée en utilisant des *opérateurs de balance* (Weaver et al. [2006], JC and Bouttier [2002]) établissant une relation entre les champs. En prévision numérique du temps par exemple, où le champ de vent  $u$  est relié au champ de pression  $\phi$  par  $u = \alpha \nabla \phi$ . À partir de cette relation de balance, une partie de l'erreur sur  $u$  est expliquée par l'erreur sur  $\phi$ .

En chimie atmosphérique, de tels opérateurs de balance ne sont pas connus. Nous cherchons donc à déterminer une approximation, dans l'idéal analytique, pour la fonction de corrélations croisées, qui est définie par :

$$\rho_{AB}(\mathbf{x}, \mathbf{y}) = \frac{\overline{\varepsilon_{\mathbf{x}}^A \varepsilon_{\mathbf{y}}^B}}{\sigma_{\mathbf{x}}^A \sigma_{\mathbf{y}}^B} \quad (21)$$

Cette formulation doit reposer uniquement sur les paramètres du système PKF.

En principe, le modèle de covariance complet (combinant auto-covariance et covariance croisées) devrait lors de la discrétisation, mener à des matrices de covariance définies positives. Néanmoins, dans le formalisme du PKF, il s'agit (voir section 3.5) simplement de mettre à jour les statistiques du PKF et non une matrice de covariance complète. Ainsi, rien n'indique *a priori* que cette contrainte doit être absolument respectée.

Afin de nous orienter dans la conception d'une formule d'approximation de (21), on s'appuie sur une approximation plus lisible de (15) donnée par :

$$\sqrt{V_{\mathbf{x}}} \sqrt{V_{\mathbf{y}}} \exp \left( -\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{s_{\mathbf{x}} + s_{\mathbf{y}}} \right) \quad (22)$$

Après avoir envisagé de nombreuses manières d'approximer les fonctions de corrélations croisées (combinaisons entre formulation analytique et réseaux de neurones, entraînés à partir d'ensembles), nous retenons la formulation analytique suivante :

$$\rho_{AB}(\mathbf{x}, \mathbf{y}) \simeq \frac{1}{2} \left( \frac{V_{\mathbf{x}}^{AB}}{\sigma_{\mathbf{x}}^A \sigma_{\mathbf{x}}^B} + \frac{V_{\mathbf{y}}^{AB}}{\sigma_{\mathbf{y}}^A \sigma_{\mathbf{y}}^B} \right) \exp \left( -\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{\frac{1}{2}(s_{\mathbf{x}}^A + s_{\mathbf{y}}^A + s_{\mathbf{x}}^B + s_{\mathbf{y}}^B)} \right) \quad (23)$$

Notons que cette formulation est symétrique, ce qui n'est pas vrai dans (21), mais que cela lui permet de tirer pleinement profit de l'information fournie par le système PKF aux points de grille  $\mathbf{x}$  et  $\mathbf{y}$ .

On a en fait pris la moyenne des coefficients de corrélations en  $\mathbf{x}$  et  $\mathbf{y}$  que l'on a multiplié par une



gaussienne. Pour le contenu de l'exponentielle, on s'est inspiré de l'approximation précédente (22) qui correspond à un cadre univarié, et on a simplement remplacé le tenseur d'aspect en un point de grille par la moyenne des tenseurs d'aspect des espèces chimiques en ce point.

Un des avantages majeurs de cette formule analytique est sa propension à être étendue à un problème de dimension supérieure et avec davantage d'espèces chimiques, comme ce sera certainement le cas par la suite.

À présent, évaluons les performances de cette formulation à l'aide de la prévision d'ensemble réalisée dans la sous-section précédente. Pour cela : on diagnostique les paramètres du système PKF à partir de l'ensemble et comparons les fonctions de corrélations croisées empiriques avec celle fournies par (23).

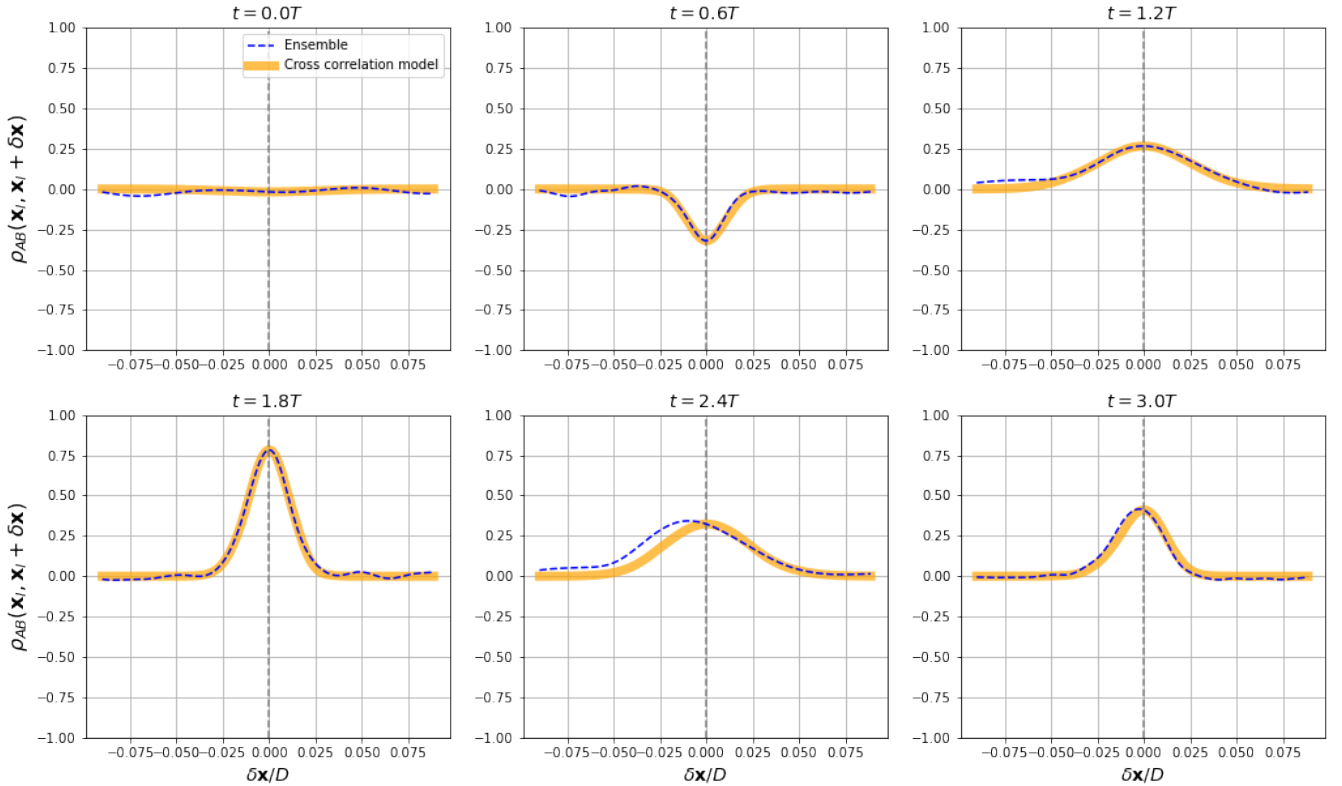


FIGURE 16 – Performance de la formulation (23) à  $\mathbf{x}_l = 0.5$  et aux temps  $t = [0.0, 0.6, 1.2, 1.8, 2.4, 3.0]T$ . Affichage pour  $\delta \mathbf{x} \in [-0.1, 0.1]$ .

La formulation (23) se montre capable d'approximer raisonnablement les fonctions de corrélations croisées à partir des paramètres du PKF.

Nous pouvons désormais utiliser cette formulation pour reconstruire les fonctions de corrélations croisées directement à partir de la prévision du PKF (voir figure 17). On note des différences d'estimation de ces fonctions. Cela s'explique d'abord par l'erreur de modélisation du modèle de corrélations croisées comme on vient de le voir précédemment, mais surtout par les écarts de prévision entre les deux méthodes pour les champs de variances et covariance (voir figure 15).



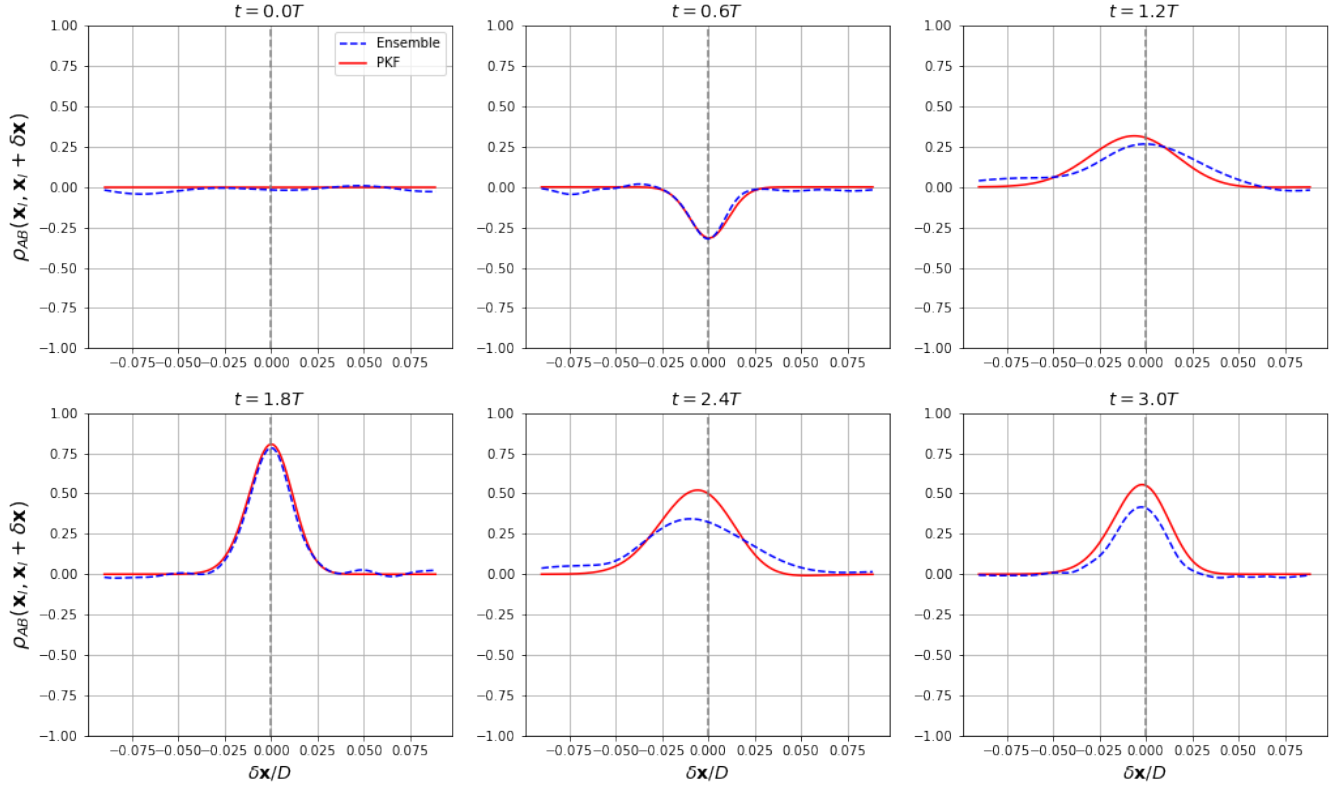


FIGURE 17 – Fonctions de corrélations croisées pour l’ensemble (courbes bleues discontinues) et le PKF (courbes continues rouges) à  $\mathbf{x}_l = 0.5$  et aux temps  $t = [0.0, 0.6, 1.2, 1.8, 2.4, 3.0]T$ . Affichage pour  $\delta\mathbf{x} \in [-0.1, 0.1]$ .

On vérifie expérimentalement que notre modèle de covariance ne s’éloigne que raisonnablement du caractère défini positif.

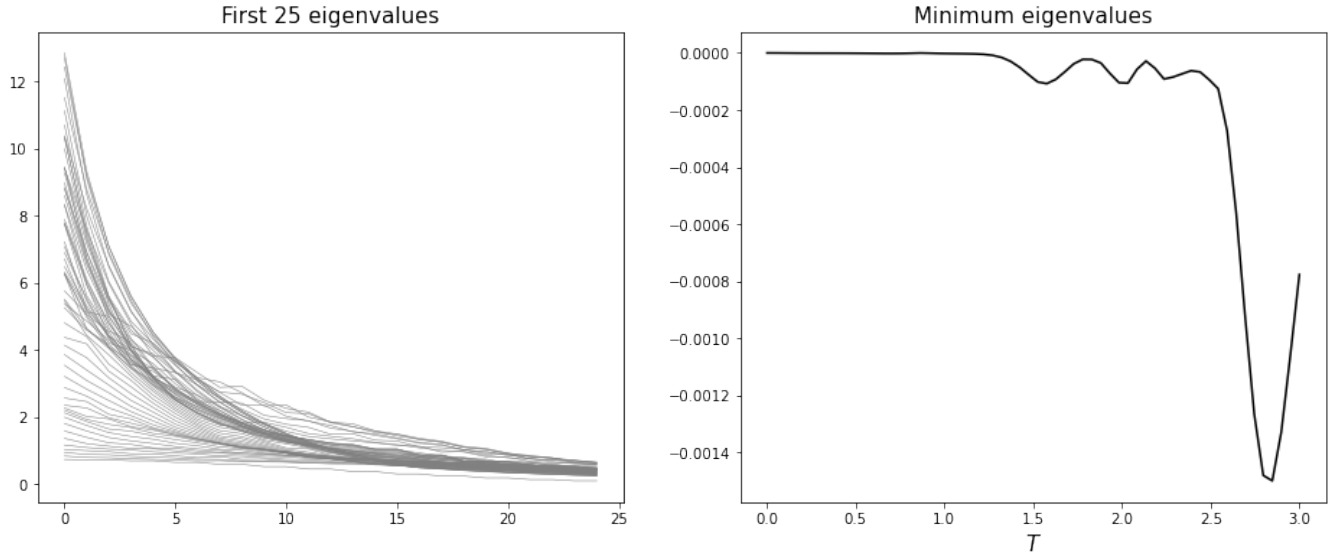


FIGURE 18 – 25 plus grandes valeurs propres des matrices de covariance générées par le modèle de covariance du PKF, à tous les temps (panneau gauche). Plus petite valeur propre des matrices de covariance du PKF en fonction du temps (panneau droit).

### 3.5 Assimilation d'une observation : PKF et EnKF

On s'intéresse à présent à la mise à jour des statistiques du PKF lors de l'assimilation d'une observation. On comparera la mise à jour de ces statistiques avec celles de l'ensemble.

La méthode de mise à jour des statistiques du PKF a été décrite dans (Pannekoucke [2021]). Les observations contenues dans un ensemble sont assimilées une à une. Pour une observation  $\mathcal{Y}_l^o$  au point de grille  $\mathbf{x}_l$ , de variance  $V_l^o$ , des formules de mises à jour viennent décrire comment sont modifiés l'état du système  $\mathcal{X}$ , la carte de variance  $V$  et la carte d'anisotropies  $\mathbf{s}$ . Il existe en fait deux formules intitulées "PKF O1" et "PKF O2" pour la mise à jour des cartes d'anisotropies, l'une d'ordre 1, la seconde d'ordre 2. On expose ici les formules de mise à jour :

$$\text{PKF O1 : } \begin{cases} \mathcal{X}_{\mathbf{x}}^a = \mathcal{X}_{\mathbf{x}}^f + \sigma_{\mathbf{x}}^f \rho_l(\mathbf{x}) \frac{\sigma_l^f}{V_l^f + V_l^o} (\mathcal{Y}_l^o - \mathcal{X}_l^f) & (24a) \\ V_{\mathbf{x}}^a = V_{\mathbf{x}}^f \left( 1 - [\rho_l(\mathbf{x})]^2 \frac{V_l^f}{V_l^f + V_l^o} \right) & (24b) \\ s_{\mathbf{x}}^a = \frac{V_{\mathbf{x}}^a}{V_{\mathbf{x}}^f} s_{\mathbf{x}}^f & (24c) \end{cases}$$

$$\text{PKF O2 : } \begin{cases} \mathcal{X}_{\mathbf{x}}^a = \mathcal{X}_{\mathbf{x}}^f + \sigma_{\mathbf{x}}^f \rho_l(\mathbf{x}) \frac{\sigma_l^f}{V_l^f + V_l^o} (\mathcal{Y}_l^o - \mathcal{X}_l^f) & (25a) \\ V_{\mathbf{x}}^a = V_{\mathbf{x}}^f \left( 1 - [\rho_l(\mathbf{x})]^2 \frac{V_l^f}{V_l^f + V_l^o} \right) & (25b) \\ g_{ij}^a(\mathbf{x}) = \frac{V_{\mathbf{x}}^f}{V_{\mathbf{x}}^a} g_{ij}^f(\mathbf{x}) + \frac{1}{4V_{\mathbf{x}}^f V_{\mathbf{x}}^a} (\partial_i V_{\mathbf{x}}^f)(\partial_j V_{\mathbf{x}}^f) \\ \quad - \frac{1}{V_{\mathbf{x}}^a} \partial_i (\rho_l(\mathbf{x}) \sigma_{\mathbf{x}}^f) \partial_j (\rho_l(\mathbf{x}) \sigma_{\mathbf{x}}^f) \frac{V_l^f}{V_l^f + V_l^o} \\ \quad - \frac{1}{4(V_{\mathbf{x}}^a)^2} (\partial_i V_{\mathbf{x}}^a)(\partial_j V_{\mathbf{x}}^a) & (25c) \end{cases}$$

où  $l, o, f, a$  et  $\rho_l$  désignent respectivement *location*, *observation*, *forecast*, *analysis* et la fonction de corrélation en  $\mathbf{x}_l$ . Notons que ces formules sont parfaitement valables dans un cadre multivarié (à plusieurs espèces chimiques), pour cela il faut remplacer  $(Z, \mathbf{x})$  et  $(l, \mathbf{x})$  au lieu de  $\mathbf{x}$  et  $l$  dans les formules, avec  $Z$  une espèce chimique.

Néanmoins, la formule de mise à jour pour le champ de covariance n'est pas décrite dans ces équations, il nous faut la déterminer de manière générique. L'on déduit en reprenant les équations de mise à jour dans le filtre de Kalman (2), que pour une observation de l'espèce  $X$  au point de grille  $\mathbf{x}_l$  de variance  $V_l^o$ , le champ de covariance des espèces  $X$  et  $Y$  est mise à jour comme suit :

$$V_{XY}^a(\mathbf{x}) = V_{XY}^f(\mathbf{x}) - \frac{c_{XY}^f(\mathbf{x}_l, \mathbf{x}) c_X^f(\mathbf{x}_l, \mathbf{x})}{V_X^f(\mathbf{x}_l) + V_l^o} \quad (26)$$

où  $c$  correspond à une fonction de covariance.

En regardant de près ces équations, on comprend en détail comment une observation et la variance qui lui est associée viennent modifier les différentes statistiques d'erreurs. On observe par exemple que

l'équation (24b) de mise à jour de la variance indique une réduction de la variance dans le voisinage de l'observation. Pour la méthode de mise à jour PKF O1, une assimilation en  $\mathbf{x}_l$  implique également une réduction des longueurs de portées au voisinage de l'observation.

On considère dans cette expérience une observation de l'espèce  $A$  en  $\mathbf{x}_l = 0.5$ , obtenue à la fin de la prévision réalisée section 3.3, figure 15. Cette observation a pour variance  $V_l^o = 0.125^2$ , identique aux champs de variances initiaux  $V_A^0$  et  $V_B^0$ . On détermine aléatoirement la valeur de l'observation  $\mathcal{Y}_{A,l}^o$  selon la loi  $\mathcal{N}(A_l^f, R_l + V_l^o)$ , où  $A_l^f$  est la prévision d'ensemble (quasi-identique à celle du PKF) pour l'espèce  $A$  en  $\mathbf{x}_l = 0.5$  et où  $R_l$  correspond à l'erreur de modèle (dans sa modélisation de la réalité) que l'on fixe arbitrairement à  $0.125^2$ .

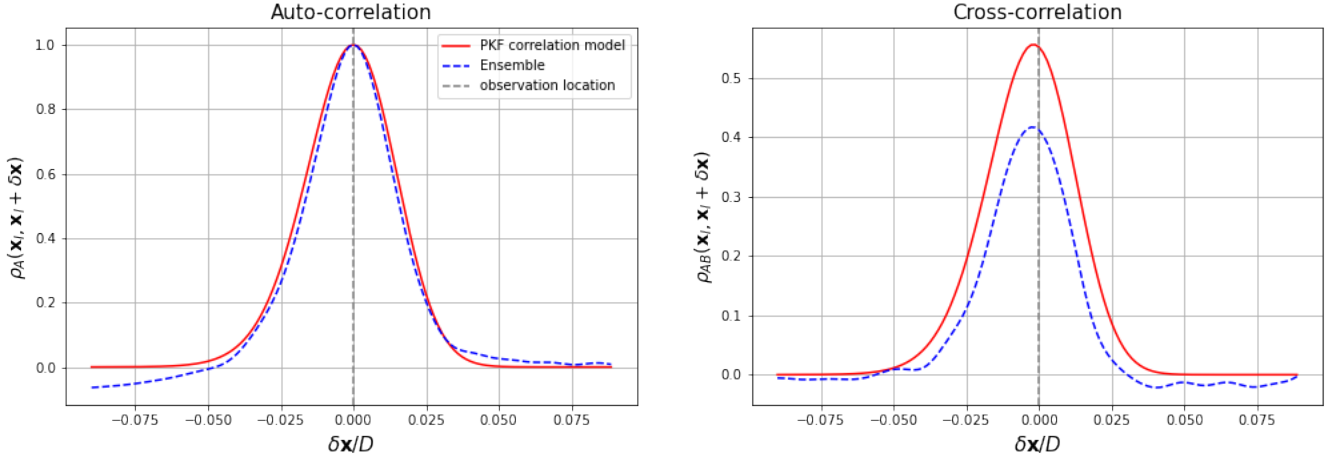


FIGURE 19 – Fonction de corrélation pour une observation à  $\mathbf{x}_l = 0.5$  de l'espèce  $A$ . Ensemble (ligne bleu pointillée) et PKF (ligne continue rouge). Affichage pour  $\delta\mathbf{x} \in [-0.1, 0.1]$ .

Voyons à présent comment l'observation a modifié les statistiques d'erreurs pour l'EnKF et le PKF (figure 20). D'abord, pour les états moyens  $A$  et  $B$ , les résultats des deux méthodes sont très proches (panneaux (a),(b),(h) et (i)). Pour les champs de variances et covariance, l'on ne partait pas au départ exactement des mêmes *forecast*, mais les résultats sont globalement semblables. On remarque davantage de différences pour les champs dans lesquels l'espèce  $B$  est impliquée, cela est dû aux différences observées dans la modélisation de la fonction de corrélations croisées en  $\mathbf{x}_l = 0.5$  (figure 19). Cet impact est également visible pour la mise à jour des longueurs de portées pour l'espèce  $B$ . Les panneaux (m) et (n) mettent en évidence les différences entre les méthodes de mise à jour PKF O1 et PKF O2 : la méthode d'ordre 2 permet une incrémentation plus juste au prix d'une complexité plus grande. Ceci n'est pas possible dans la formulation d'ordre 1, moins précise mais plus robuste, en particulier en dimension supérieure (*cf.* Pannekoucke [2021]). Notons enfin que la réduction des longueurs de portées, parfois de l'ordre de 70% (panneau (m)), pourrait lors d'une nouvelle prévision avec l'EnKF, altérer grandement l'estimation de l'erreur de prévision (voir section 2.4.2), ce dont ne souffrirait pas le PKF.

En conclusion, nous sommes parvenus à mettre correctement à jour les statistiques du PKF à un coût bien inférieur à celui de l'ensemble, qui comporte des multiplications de matrices de très grandes dimensions.

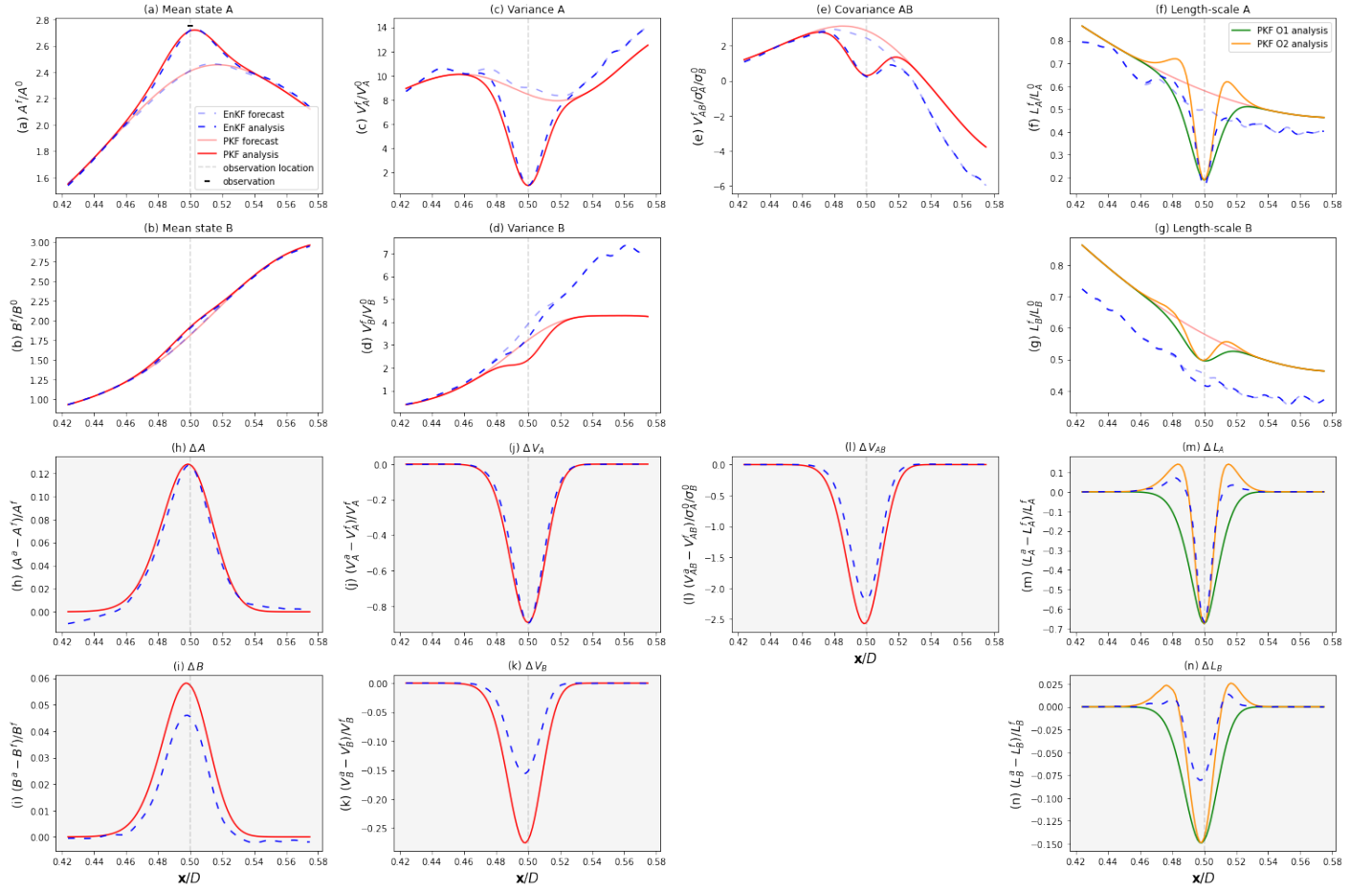


FIGURE 20 – Graphiques de l'étape d'assimilation. Comparaison PKF (lignes continues rouges, excepté pour les longueurs de portées en vert et orange) et EnKF (lignes pointillées bleues) pour les statistiques de *forecast* (lignes en légère transparence) et d'*analysis* (opaques) (panneaux blancs (a)-(f)). Incréments dus à l'assimilation de l'observation pour les différentes statistiques (panneaux gris (g)-(n)). Affichage pour  $\mathbf{x} \in [0.42, 0.58]$ .

## 4 Vers un modèle chimique transport plus réaliste

Afin de gagner en réalisme avec le PKF, on s'intéresse à un modèle chimique (dans un cadre 0D) plus réaliste et représentant des dynamiques plus complexes que celles qui peuvent être décrites par le système de Lotka-Volterra (7).

### 4.1 Introduction du modèle chimique GRS

Le modèle auquel on s'intéresse ici, intitulé GRS (Azzi et al. [1992], Haussaire and Bocquet [2016]) pour *generic reaction set*, décrit la dynamique d'un nombre réduit d'espèces ou méta-espèces chimiques présentes dans l'atmosphère.

Les réactions chimiques décrites par le modèle GRS sont les suivantes :



où ROC, RP et S(N)GN signifient respectivement *Reactive Organic Compound*, *Radical Pool* et *Stable (Non-) Gaseous Nitrogen product*.

Les émissions dans l'atmosphère de chacune de ces espèces sont également considérées, ainsi que leur dépôt au sol à un taux commun  $\lambda$ . Le système d'équations différentielles s'écrit ainsi :

$$\text{GRS : } \left\{ \begin{array}{l} \frac{d[ROC]}{dt} = -\lambda[ROC] + E_{ROC} \end{array} \right. \quad (28a)$$

$$\frac{d[RP]}{dt} = -\lambda[RP] + k_1[ROC] - [RP] (k_2[NO] + 2k_6[NO_2] + k_5[RP]) \quad (28b)$$

$$\frac{d[NO]}{dt} = -\lambda[NO] + E_{NO} + k_3[NO_2] - [NO] (k_2[RP] + k_4[O_3]) \quad (28c)$$

$$\frac{d[NO_2]}{dt} = -\lambda[NO_2] + E_{NO_2} + k_4[NO][O_3] + k_2[NO][RP] - [NO_2] (k_3 + 2k_6[RP]) \quad (28d)$$

$$\frac{d[O_3]}{dt} = -\lambda[O_3] + k_3[NO_2] - k_4[NO][O_3] \quad (28e)$$

$$\frac{d[S(N)GN]}{dt} = -\lambda[S(N)GN] + 2k_6[NO_2][RP] \quad (28f)$$

La vitesse de réaction  $k_1$  est reliée à la vitesse de photolyse  $k_3$  par  $k_1 = 0.152k_3$ . Cette vitesse de photolyse varie au cours d'une journée (voir figure 21).

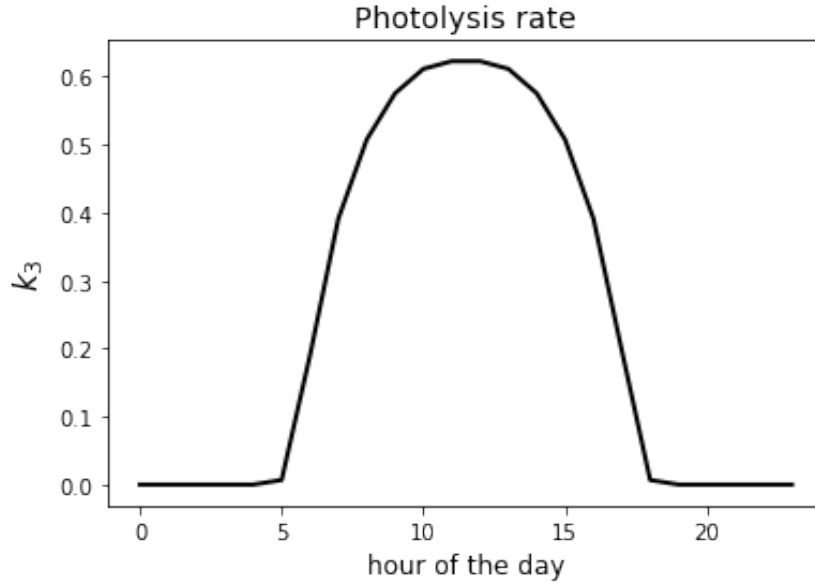


FIGURE 21 – Valeurs de  $k_3$  au cours d’une journée.

Toutes les valeurs des paramètres chimiques (concentrations initiales, facteurs d’émissions, vitesses de réaction, etc.) sont tirées de (Haussaire and Bocquet [2016]).

## 4.2 Comparaison prévision PKF et EnKF

On ne décrit pas ici l’étape de formulation du PKF (réalisée à l’aide du SymPKF). On rappelle qu’il est constitué dans ce contexte des états moyens et variances de chaque espèce ( $6 + 6$  paramètres) ainsi des covariances de toutes les paires d’espèces (15 paramètres), soit 27 paramètres au total.

**génération d’un état et d’une erreur initiale :** Toutes les concentrations étant initialement fixées à zéro dans l’article, il n’est pas possible de générer un ensemble. Nous lançons donc une prévision partant de zéro, stoppons la simulation dès que toutes les espèces sont en concentrations non nulles et décrétons une erreur de l’ordre de 10% des concentrations observées. On considère que les espèces sont indépendantes les unes des autres à ce stade. Ces paramètres constitueront nos données initiales pour le PKF et l’EnKF.

**paramètres de simulation numérique :** On considère un schéma d’Euler explicite avec un pas de temps  $\Delta t = 0.001$ . On arrête la simulation à  $T = 72\text{h}$ .

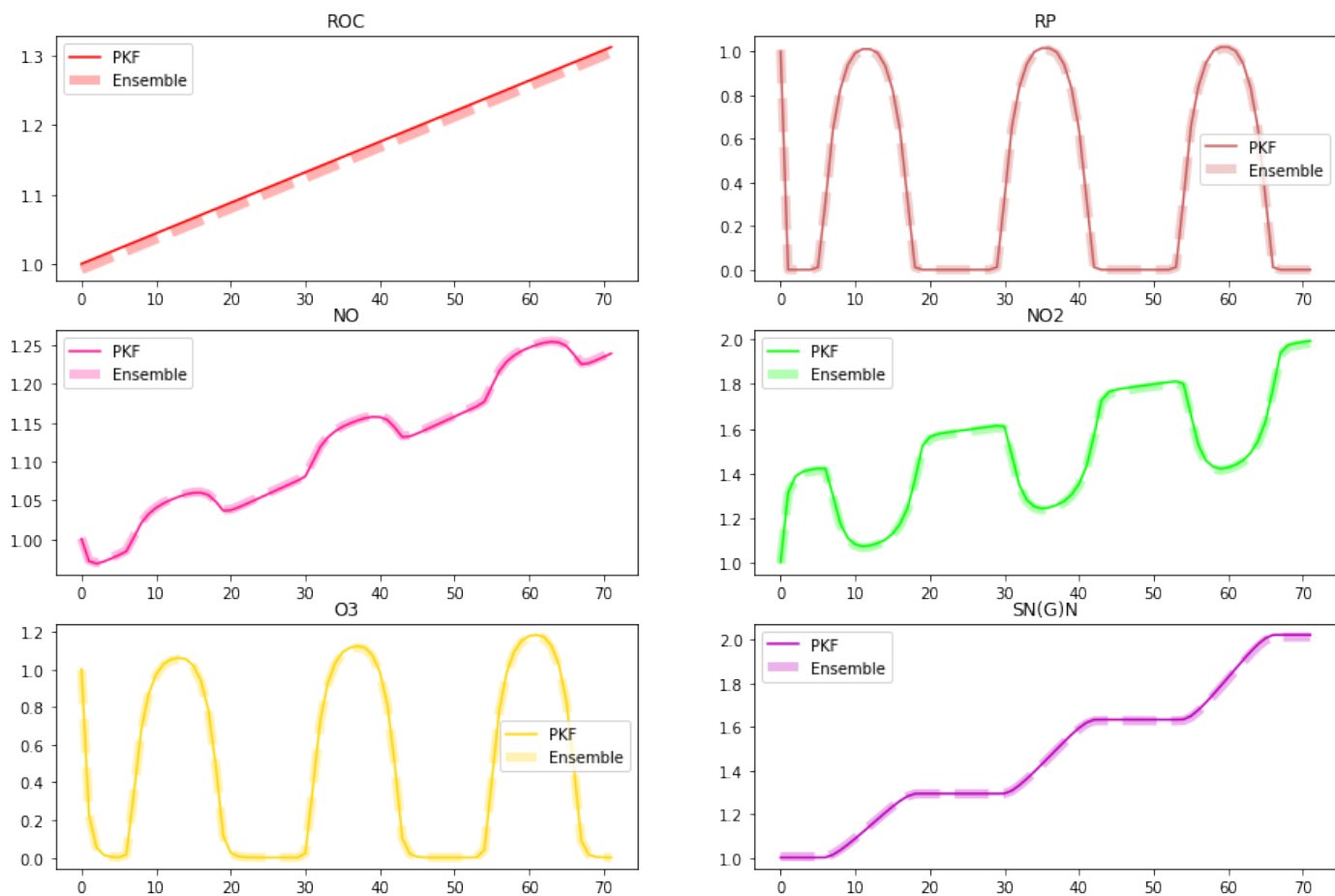


FIGURE 22 – Pr vision des concentrations. PKF (lignes continues) et EnKF (lignes  paisses en pointill s). Les pr visions sont normalis es par les concentrations initiales.

On observe que pour les deux m thodes les pr visions sont identiques. Est visible  galement l'impact de la variation du taux de photolyse  $k_3$  au long d'une journ e.

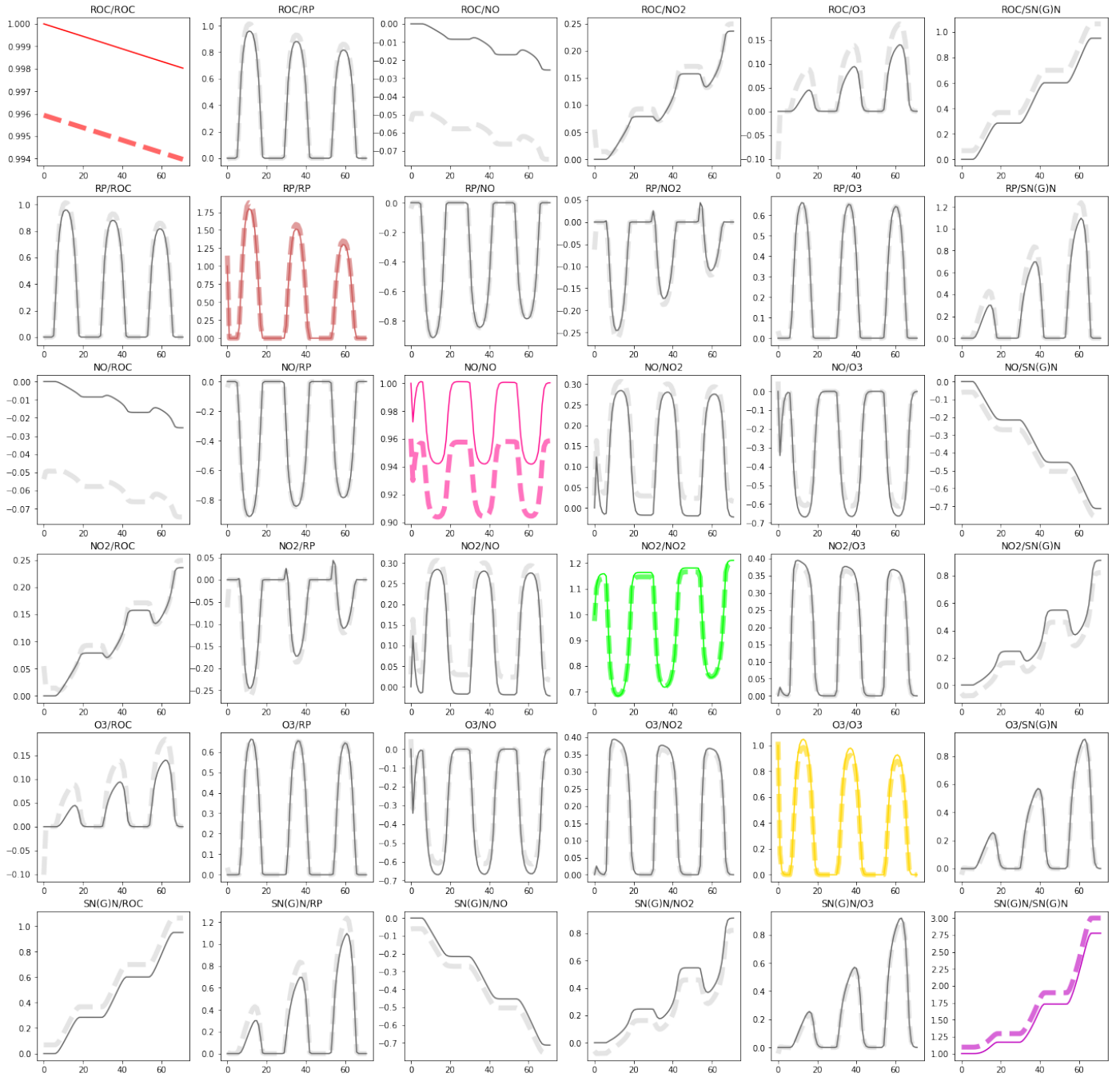


FIGURE 23 – Matrices de covariances d’erreurs de prévision. PKF (lignes continues) et EnKF (lignes épaisses en pointillés). Les variances sont normalisées par leur valeur initiale, les covariances par les écart-types des espèces correspondantes.

On observe que le PKF décrit parfaitement la dynamique des incertitudes : les tendances sont les mêmes pour les deux méthodes, le bruit d’échantillonnage en moins pour le PKF.

Ce résultat est motivant : dans un cadre bien plus complexe que celui de Lotka-Volterra (7), le PKF parvient à décrire la dynamique des incertitudes. En sachant que le processus d’advection est parfaitement décrit par le PKF (voir section 2), il est raisonnable de penser que le PKF puisse décrire la dynamique des incertitudes dans un modèle combinant GRS et transport, comme on l’a vu dans la section 3.



## 5 Conclusion et perspectives

L'objectif de ce travail était d'explorer l'assimilation multivariée pour les besoins en chimie atmosphérique. Pour cela nous avons utilisé et développé le formalisme du filtre de Kalman paramétrique dans le cadre simplifié du modèle de Lotka-Volterra (LV) sur un domaine 1D. Le modèle de LV est non-linéaire, ce qui correspond bien au cadre rencontré en chimie atmosphérique ; et représente l'interaction de deux espèces chimiques, ce qui permet d'aborder l'étude multivariée.

L'exploration du filtre de Kalman paramétrique, dans ce cadre simplifié, a permis de retrouver les résultats produits par un filtre de Kalman d'ensemble, mais à un coût numérique bien moindre et sans erreur d'échantillonnage. Ces résultats prometteurs font l'objet de la rédaction d'un article de recherche (Perrot et al.).

Il reste à transposer ces résultats dans le cadre de la prévision opérationnelle en chimie atmosphérique où il sera nécessaire de considérer une géométrie plus complexe (sphère ou domaine limité sur la sphère, avec extension verticale) et une chimie faisant intervenir des centaines d'espèces. Une première exploration vers une chimie plus complète, abordée en fin de ce travail, suggère que le passage à l'échelle opérationnelle semble bien possible.

L'ensemble de ces perspectives de travail feront l'objet de ma thèse.

# Références

- M. Azzi, G. Johnson, and M. Cope. An introduction to the generic reaction set photochemical smog mechanism. *Proceedings of the International Conference of the Clean Air Society of Australia and New Zealand*, 3 :451–462, 01 1992.
- L. Berre, O. Pannekoucke, G. Desroziers, S. Stefanescu, B. Chapnik, and L. Raynaud. A variational assimilation ensemble and the spatial filtering of its error covariances : increase of sample size by local spatial averaging. In ECMWF, editor, *ECMWF Workshop on Flow-dependent aspects of data assimilation, 11-13 June 2007.*, pages 151–168, Reading, UK, 11-13 June 2007.
- P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519) : 1367–1387, 1994. doi : 10.1002/qj.49712051912.
- Daley. Atmospheric data analysis. 1991.
- G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 :10143–10162, 1994.
- J.-M. Haussaire and M. Bocquet. A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes : L95-grs (v1.0). *Geoscientific Model Development*, 9(1) :393–412, 2016. doi : 10.5194/gmd-9-393-2016. URL <https://gmd.copernicus.org/articles/9/393/2016/>.
- D. JC and F. Bouttier. A reformulation of the background error covariance in the ecmwf global data assimilation. *Tellus A*, 51 :195 – 221, 09 2002. doi : 10.1034/j.1600-0870.1999.t01-2-00003.x.
- R. E. Kalman and Others. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1) :35–45, 1960.
- C. Paciorek and M. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17 :483–506, 08 2006. doi : 10.1002/env.785.
- O. Pannekoucke. An anisotropic formulation of the parametric kalman filter assimilation. *Tellus A : Dynamic Meteorology and Oceanography*, 73(1) :1–27, 2021. doi : 10.1080/16000870.2021.1926660. URL <https://doi.org/10.1080/16000870.2021.1926660>.
- O. Pannekoucke and P. Arbogast. Sympkf : a symbolic and computational toolbox for the design of parametric kalman filter dynamics. Mar. 2021.
- O. Pannekoucke, S. Ricci, S. Barthelemy, R. Ménard, and O. Thual. Parametric kalman filter for chemical transport models. *Tellus A : Dynamic Meteorology and Oceanography*, 68(1) :31547, 2016. doi : 10.3402/tellusa.v68.31547. URL <https://doi.org/10.3402/tellusa.v68.31547>.
- O. Pannekoucke, M. Bocquet, and R. Ménard. Parametric covariance dynamics for the nonlinear diffusive burgers’ equation. *Nonlinear Processes in Geophysics*, 2018 :1–21, 2018. doi : <https://doi.org/10.5194/npg-2018-10>. URL <https://www.nonlin-processes-geophys-discuss.net/npg-2018-10/>.
- A. Perrot, O. Pannekoucke, and V. Guidard. Toward a multivariate formulation of the pkf dynamics : application to a simplified chemical transport model. *Nonlinear Processes in Geophysics*.
- A. Weaver, C. Deltel, E. Machu, S. Ricci, and N. Daget. A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131 :3605 – 3625, 12 2006. doi : 10.1256/qj.05.119.