Boundary Conditions for the Parametric Kalman Filter forecast

Martin Sabathier[1], Olivier Pannekoucke[2], Vincent Maget[1], and Nour Dahmen[1]

[1]ONERA
[2]CNRM UMR 3589, Météo-France/CNRS, Toulouse, France

May 1, 2023

Abstract

This paper is a contribution to the exploration of the parametric Kalman filter (PKF), which is an approximation of the Kalman filter, where the error covariance are approximated by a covariance model. Here we focus on the covariance model parameterized from the variance and the anisotropy of the local correlations, and whose parameters dynamics provides a proxy for the full error-covariance dynamics. For this covariance mode, we aim to provide the boundary condition to specify in the prediction of PKF for bounded domains, focusing on Dirichlet and Neumann conditions when they are prescribed for the physical dynamics. An ensemble validation is proposed for the transport equation and for the heterogeneous diffusion equations over a bounded 1D domain. This ensemble validation requires to specify the auto-correlation time-scale needed to populate boundary perturbation that leads to prescribed uncertainty characteristics. The numerical simulations show that the PKF is able to reproduce the uncertainty diagnosed from the ensemble of forecast appropriately perturbed on the boundaries, which show the ability of the PKF to handle boundaries in the prediction of the uncertainties. It results that Dirichlet condition on the physical dynamics implies Dirichlet condition on the variance and on the anisotropy.

Boundary Conditions for the Parametric Kalman Filter forecast

version of: October 28, 2022, 10:06am

M. Sabathier[1], O. Pannekoucke[2,3,4], V. Maget[1], N. Dahmen[1]

[1]ONERA, Toulouse, France
[2]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[3]CERFACS, Toulouse, France
[4]INPT-ENM, Toulouse, France

Key Points:

- Data assimilation
- Parametric Kalman Filter
- Boundary conditions

Corresponding author: Vincent Maget, `vincent.maget@onera.fr`

Abstract

This paper is a contribution to the exploration of the parametric Kalman filter (PKF), which is an approximation of the Kalman filter, where the error covariance are approximated by a covariance model. Here we focus on the covariance model parameterized from the variance and the anisotropy of the local correlations, and whose parameters dynamics provides a proxy for the full error-covariance dynamics. For this covariance mode, we aim to provide the boundary condition to specify in the prediction of PKF for bounded domains, focusing on Dirichlet and Neumann conditions when they are prescribed for the physical dynamics. An ensemble validation is proposed for the transport equation and for the heterogeneous diffusion equations over a bounded 1D domain. This ensemble validation requires to specify the auto-correlation time-scale needed to populate boundary perturbation that leads to prescribed uncertainty characteristics. The numerical simulations show that the PKF is able to reproduce the uncertainty diagnosed from the ensemble of forecast appropriately perturbed on the boundaries, which show the ability of the PKF to handle boundaries in the prediction of the uncertainties. It results that Dirichlet condition on the physical dynamics implies Dirichlet condition on the variance and on the anisotropy.

Plain Language Summary

This work addresses the question of the uncertainty prediction in bounded domains. It contributes to explore a theoretical formulation of the uncertainty prediction that opens the way to data assimilation in real applications where the boundaries are important as in radiation belts predictions, air quality, atmosphere-ocean coupling, or wild-land fire ; while these applications are not discussed here.

1 Introduction

Uncertainty prediction is a challenging topic, important in data assimilation as well as in probabilistic forecasting. One of the main theoretical backbone is given by the Kalman filter equations that applies for linear dynamics, but that fails to apply for large system where the numerical cost to predict the error covariance becomes prohibitive. Hence, approximations of the KF have been proposed, as the ensemble Kalman filter (EnKF) where the error covariance matrices are approximated from ensemble estimation (Evensen, 2009) ; or recently the parametric Kalman filter (PKF) where the error covariance matrices are approximated from a covariance model (Pannekoucke et al., 2016). In the PKF, the dynamics of the parameters provides a proxy for the dynamics of the full covariance matrix. For instance, covariance model parameterized from the variance and the anisotropy of the local correlation functions are able to predict the dynamics of the covariance matrix for transport equations (Cohn, 1993), but at a numerical cost equivalent to three time the integration of the transport (Pannekoucke et al., 2018; Pannekoucke, 2021). In addition, the PKF provides a view of the dynamics of the uncertainties that cannot be understood from an ensemble estimate alone. And this has open new ways to tackle difficult topics *e.g.* the dynamics of the model-error covariance (Pannekoucke et al., 2021; Ménard et al., 2021). Note that other insight in theoretical covariance dynamics have been recently proposed (Gilpin et al., 2022).

The EnKF is widely used and has shown to perform well for many applications in geosciences *e.g.* for the weather prediction (Houtekamer & Mitchell, 2001), or for the radiation belts prediction (Bourdarie & Maget, 2012). Radiation belts dynamics modeling consist in estimating quantitatively the fluxes of high energetic electrons and protons trapped in the Earth magnetic field using a typical advection-diffusion equation. This region spans from 1 Earth Radius up to 8, thus encompassing all typical satellites orbits, with which such particles can strongly interact and induced from minor to critical onboard anomalies. Compared with global prediction, radiation belts predictions are performed on a limited and non-periodic domain where the boundary imposes conditions to the dynamics of the electrons and protons. Indeed, on one side, the outer boundary condition is considered as the prime access for fresh materials, coming from the so-called magneto-tail, and is typically modeled as an imposed Dirichlet condition at this altitude (8 Earth radii) that can evolve as a function of solar activity (e.g. energy spectrum reshaping from time to time) (Maget et al., 2015). On the other side, close to the Earth, the atmosphere implies a necessary fixed Dirichlet condition too, as all radiation belts particles coming down there are absorbed (e.g. distribution always equal to 0). Finally, for low energy boundary we expect to rely on a Neumann condition to limit naturally any escape of particles or artificial source. Nonetheless, when performed on limited area models, atmospheric prediction also present such kind of boundaries.

An appropriate specification of boundary uncertainties is crucial because it tells how uncertainty will enter the domain, while the dynamics will transport and modify it. Hence, appropriate specification of boundary condition for uncertainty prediction and assimilation is a crucial issue.

In an astonishing way, it is relatively easy to introduce uncertainty at the boundary in LAM for ensemble methods, by considering a forcing from an ensemble of global forecast even if the consistency across multiple domains is difficult to handle (Houtekamer & Zhang, 2016, sec. 6.a); while at a theoretical level, the Kalman forecast equation, in its matrix algebra, is less appropriate to introduce such an uncertainty at boundaries. Note that the difficulty to control the error at the boundary also exists for variational data assimilation Gustafsson (2012).

Until now, the PKF has been explored on periodic 1D or 2D domains, where it has been shown to reproduce interesting features of the uncertainty dynamics in linear problem *e.g.* for the transport (Pannekoucke, 2021), as well as for non-linear dynamics at the second order *e.g.* for the non-linear advection-diffusion equation (Pannekoucke et al., 2018).

However, to go ahead toward real applications, and especially applications in bounded domains, appropriate specification of boundary condition of the error statistics is needed for the PKF dynamics. To do so, we propose to explore the specification of the boundary conditions for the PKF when Dirichlet and Neumann conditions are considered in the physical dynamics. This exploration is focus of two dynamics of interest for our applications: the transport equation *e.g.* for air quality or weather prediction ; and the diffusion equation *e.g.* for radiation belts prediction or uncertainty dynamics in boundary layer for air quality. The paper focuses on the forecast step, and the assimilation step is not addressed here.

The paper is organized as follows. First, the background of the parametric Kalman filter (PKF) is reminded in Section 2. Then, Section 3 details how to specify the PKF conditions at the boundary for the forecast for the Dirichlet and the Neumann conditions. The ensemble validation of the boundary conditions for the PKF needs to create an ensemble of forecasts. To do so, an intermediate Section 4 will details how to specify boundary conditions in an EnKF experiment that produces desired error statistics. This is an important contribution of the paper so to validate the specification of the boundary conditions of the PKF, where the numerical validation is presented in Section 5. Conclusions and perspectives are given in the last Section 6.

2 Background on the PKF forecast step

This section gives a self-content introduction to the parametric Kalman filter, applied for a particular covariance model. First, the prediction step of the Kalman filter applied on a linear dynamics is reminded. Then, the formalism of the PKF is introduced, followed by the illustration on two dynamics: the transport equation, important in geosciences, and the diffusion equation important in radiation belt dynamics community.

2.1 Kalman filter forecast step

Here we consider the prediction of a univariate physical field $\chi(t, \mathbf{x})$ defined on a domain Ω of dimension d and coordinate system $\mathbf{x} = (x^i)_{i\in[1,d]}$, whose dynamics is given by

$$\partial_t \chi = \mathcal{M}(\chi, \partial\chi), \tag{1}$$

where $\mathcal{M}$ stands for a function of the state χ and of its spatial derivatives, $\partial\chi$, which is a shorthand for the partial derivative with respect to the spatial coordinates at any arbitrary orders. Thereafter, for the sake of simplicity, $\mathcal{M}$ is assumed linear but the formalism extends to the non-linear framework (Pannekoucke et al., 2018; Pannekoucke & Arbogast, 2021). Note that χ can be either continuous or discrete (the discretized version of the continuous field): the discrete case leads to matrix algebra relations *e.g.* $\mathcal{M}$ is replaced by its matrix formulation $\mathbf{M}$.

In real applications, the spatio-temporal heterogeneity of the observation network, as well as the model error, imply that χ is not known exactly and is modeled as a random field χ. The true state of the system is denoted by χ^t. The analysis state, that is the estimation of the true state knowing the observations until a given time, is denoted by χ^a. The deviation of the analysis state from the truth is the analysis error, $e^a = \chi^a - \chi^t$, and is often modeled as a random Gaussian vector of zero mean and covariance matrix $\mathbf{P}^a = \mathbb{E}\left[e^a (e^a)^{\mathrm{T}}\right]$, where $\mathbb{E}\left[\cdot\right]$ stands for the expectation operator and where the upper script $(\cdot)^{\mathrm{T}}$ stands for the transpose operator (later the adjoin operator for matrices). The forecast state at a time T, $\chi^f(T) = \mathbf{M}_{T\leftarrow 0}\chi^a$ provides an approximation of the true state at time T, where $\mathbf{M}_{T\leftarrow 0}$ denotes the propagator associated with the time integration of Eq. (1) over the period $[0, T]$. For linear dynamics and Gaussian uncertainty, the forecast error $e^f(T) = \chi^f(T) - \chi^t(T)$ is a Gaussian vector of zero mean and covariance matrix $\mathbf{P}^f(T) = \mathbb{E}\left[e^f (e^f)^{\mathrm{T}}\right](T)$, whose dynamics writes as

$$\partial_t e^f = \mathcal{M}(e^f, \partial e^f). \tag{2}$$

The forecast-error covariance matrix is related to the analysis-error covariance matrix by

$$\mathbf{P}^f(T) = \mathbf{M}_{T\leftarrow 0}\mathbf{P}^a \left(\mathbf{M}_{T\leftarrow 0}\right)^{\mathrm{T}}. \tag{3}$$

Equation (3) corresponds to the Kalman filter propagator of the error covariance matrix, whose the particular dynamics is given by

$$\frac{d\mathbf{P}^f}{dt} = \mathbf{M}\mathbf{P}^f + \mathbf{P}^f\mathbf{M}^{\mathrm{T}}, \tag{4}$$

integrated over the period $[0, T]$, starting from the initial condition $\mathbf{P}^f(0) = \mathbf{P}^a$.

While the KF forecast step Eq. (3) is a simple algebraic formula, its fails to apply in large systems because of its numerical cost: if n denotes the dimension of the vector representation of χ, then the computational complexity of Eq. (3) scales between n^2 and n^3 (Strassen, 1969). In term of integration cost, the KF requires $2n$ integrations of the model Eq. (1).

Hence, approximations for the KF are needed. For instance, in the Ensemble Kalman filter (EnKF), the forecast error-covariance matrix is approximated by its ensemble estimation.

$$\widehat{\mathbf{P}^f}(t) = \frac{1}{N}\sum_k e_k e_k^{\mathrm{T}}, \tag{5}$$

with $e_k = \chi_k(t) - \widehat{\chi}(t)$ where $\widehat{\chi}(t) = \frac{1}{N}\sum_k \chi_k(t)$ denotes the empirical mean and $(\chi_k(t))_{k\in[1,N]}$ is an ensemble of N forecasts (Evensen, 2009). This time, the numerical complexity scales with the number of ensemble members N and the size of the problem n: the numerical cost of an ensemble of forecast is the cost of N integrations of the model Eq. (1).

Note that the normalization by N in Eq. (5) leads to a bias that decreases as $1/N$. In EnKF framework, the normalization by $N-1$ is preferred, however since we latter consider estimation from very large ensemble size, the corrections of the estimators are not considered here, and we only consider empirical mean estimations $\frac{1}{N}\sum_k(\cdots)$ as in Eq. (5).

The next section presents another approximation for the error-covariance matrices.

2.2 Parametric formulation for the Kalman filter forecast step based on VLATcov models

In the parametric approach, a covariance model is introduced, $\mathbf{P}(\mathcal{P})$ where $\mathcal{P}$ denotes the set of parameters of the covariance model, so to approximate the error covariance matrices. For instance, the forecast-error covariance matrix $\mathbf{P}^f$, is approximated as $\mathbf{P}(\mathcal{P}^f) \approx \mathbf{P}^f$, where $\mathcal{P}^f$ is a particular set of values for the parameters. The parametric Kalman filter (PKF) dynamics remains to mimic the dynamics of Eq. (4) relying on the dynamics of the parameters $\mathcal{P}^f$,

$$\frac{d\mathcal{P}^f}{dt} = \mathcal{G}(\mathcal{P}^f), \tag{6}$$

where $\mathcal{G}$ has to be determined from the particular dynamics of Eq. (1), so that at any time t, $\mathbf{P}(\mathcal{P}^f(t))$ approximates $\mathbf{P}^f(t)$ *i.e.* $\mathbf{P}(\mathcal{P}^f(t)) \approx \mathbf{P}^f(t)$. As for the EnKF, the numerical complexity of the PKF prediction Eq. (6) scales as number of parameters and the dimension of the problem n: the numerical cost of the PKF represent the cost of few numerical integrations of the dynamics Eq. (1), depending on the number of parameters needed for the covariance approximation.

Thereafter, since we deal with the forecast step of the PKF, the upper-script ^{f} is dropped in the notation that concerns the forecast-error statistics.

This contribution will focus on the particular class of covariance model, so-called VLATcov models, parameterized from two fields, defined below: the variance field, V, and the local anisotropy tensor of the correlation functions, $\mathbf{g}$ or $\mathbf{s}$. Hence, the set of parameters is given by the couple $\mathcal{P} = (V, \mathbf{g})$ or $\mathcal{P} = (V, \mathbf{s})$, so that a VLATcov model writes as $\mathbf{P}(V, \mathbf{g})$ or $\mathbf{P}(V, \mathbf{s})$. For an error field e, the variance field is defined as

$$V = \mathbb{E}\left[e^2\right], \tag{7}$$

and is used to introduce the normalized error $\varepsilon = \frac{e}{\sqrt{V}}$. When the error field is a differential random field, that is assumed from now, the correlation function $\rho(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[\varepsilon(\mathbf{x})\varepsilon(\mathbf{y})\right]$ is flat for $\mathbf{y} = \mathbf{x}$. Then, the local anisotropy at $\mathbf{x}$ is defined as the local metric tensor $\mathbf{g}(\mathbf{x})$ (also denoted by $\mathbf{g}_\mathbf{x}$) which appears in the second-order Taylor's expansion

$$\rho(\mathbf{x}, \mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}_\mathbf{x}}, \tag{8}$$

where $||\delta\mathbf{x}||^2_{\mathbf{g}_\mathbf{x}} = \delta\mathbf{x}^{\mathrm{T}}\mathbf{g}_\mathbf{x}\delta\mathbf{x}$ denotes the norm associated with the metric tensor $\mathbf{g}_\mathbf{x}$ that is the symmetric definite positive matrice $[\mathbf{g}_\mathbf{x}]_{ij} = -\partial^2_{x^ix^j}\rho_\mathbf{x}$ where $\rho_\mathbf{x}(\delta\mathbf{x})$ stands for the local correlation function. In a 1D domain of coordinate x, the metric tensor field is the scalar field $\mathbf{g} = [g_{xx}]$.

In practice, the geometry of the local metric tensor is contravariant: the direction of largest correlation anisotropy corresponds to the principal axes of smallest eigenvalue for the metric tensor. Thus, it is useful to introduce the local aspect tensor (Derber et al., 2003)

$$\mathbf{s}(\mathbf{x}) = \left(\mathbf{g}(\mathbf{x})\right)^{-1}, \tag{9}$$

where the superscript $(\cdot)^{-1}$ denotes the matrix inverse, and whose the geometry goes as the correlation.

What makes the local metric tensor attractive is that this tensor is related to the normalized error by (see *e.g.* Pannekoucke (2021))

$$[\mathbf{g_x}]_{ij} = \mathbb{E}\left[\partial_{x^i}\varepsilon\partial_{x^j}\varepsilon\right]. \tag{10}$$

Hence, the variance Eq. (7) and the anisotropy Eq. (10) can be computed from an ensemble estimation: the variance field is estimated by

$$\widehat{V} = \frac{1}{N}\sum_k \left(e_k(t)\right)^2, \tag{11}$$

with $e_k(t) = \chi_k(t) - \widehat{\chi}(t)$, from which derivatives of the normalized error $\varepsilon_k = \frac{1}{\sqrt{V}}\left(\chi_k(t) - \widehat{\chi}(t)\right)$ leads to the estimation of the upper triangular components of the metric

$$\widehat{g}_{ij} = \frac{1}{N}\sum_k \partial_{x^i}\varepsilon_k\partial_{x^j}\varepsilon_k, \tag{12}$$

for $i \leq j$ (since $g_{ji} = g_{ij}$). While the PKF approach does not relies on any ensembles, the ensemble estimations Eq. (11) and Eq. (12) can be used to set the initial conditions for the parameters to ignite the assimilation cycles, or to validate the PKF from the diagnosis of an EnKF.

An example VLATcov model is given by the heterogeneous Gaussian-like covariance model (Paciorek & Schervish, 2006)

$$\mathbf{P}(V,\mathbf{s})(\mathbf{x},\mathbf{y}) = \sqrt{V_\mathbf{x}V_\mathbf{y}}\frac{|\mathbf{s_x}|^{1/4}|\mathbf{s_y}|^{1/4}}{|\frac{1}{2}(\mathbf{s_x}+\mathbf{s_y})|^{1/2}}\exp\left(-\frac{1}{2}||\mathbf{x}-\mathbf{y}||^2_{[\frac{1}{2}(\mathbf{s_x}+\mathbf{s_y})]^{-1}}\right) \tag{13}$$

where $|\cdot|$ denotes the matrix determinant.

When VLATcov models are used for the parametric approach, the dynamics of the parameters Eq. (6) is deduced from the time derivative of Eq. (7) and Eq. (10), and the dynamics of the error Eq. (2). For instance, the dynamics of V is deduced from

$$\partial_t V = 2\mathbb{E}\left[e\partial_t e\right], \tag{14a}$$

where replacing the trend of the error Eq. (2), will leads to the dynamics of V

$$\partial_t V = 2\mathbb{E}\left[e\mathcal{M}(e,\partial e)\right]. \tag{14b}$$

This expression can be simplified *e.g.* by considering the commutation between the expectation and partial derivatives (Pannekoucke & Arbogast, 2021).

In terms of numerical cost, the PKF based on the VLATcov model scales as the number of independent components in $\mathbf{g}$ (the number of coefficients in the upper triangle) plus one for the variance field: in a univariate over a 1D (3D) domain, this represents 2 (7) times the cost of one model forecast (which scales itself with the dimension n).

Note that the computation of dynamical equations for V and $\mathbf{g}$ (or $\mathbf{s}$) can be performed using a computed algebra system. To do so, the open source Python toolbox SymPKF[1] has been introduced (Pannekoucke & Arbogast, 2021), which computes the dynamics of the parameters and renders a numerical code to facilitate the numerical exploration of the PKF approach. Another way to simplify the computation of the parameters dynamics is to identify the contribution of each physical process present in Eq. (1) following a splitting strategy (Pannekoucke et al., 2018; Pannekoucke & Arbogast, 2021). Thereafter, the dynamics of the VLATcov parameters is computed by using SymPKF and the

[1] `https://github.com/opannekoucke/sympkf`

interested reader is referred to the Jupyter notebooks that are provided as a supplementary material to this contribution [2].

The PKF based on the VLATcov model is illustrated in the next sections for two dynamics which give an explicit form for $\mathcal{M}$ in Eq. (1).

2.3 Illustration of the PKF for simple dynamics

The transport and the diffusion equations are considered so to detail the dynamics of the variance and the anisotropy for the PKF applied for VLATcov models. Both dynamics play over a 1D periodical domain of coordinate x, so that the dynamics is an evolution equation without boundary conditions.

2.3.1 PKF prediction applied on a transport equation

The transport equation of a scalar field $c(t,x)$ by a stationary velocity field $u(x)$ writes as

$$\partial_t c + u\partial_x c = 0. \tag{15}$$

In this example, and by identification with Eq. (1), c stands for χ while $\mathcal{M}(c,\partial c) = -u\partial_x c$. This kind of equation appears for instance in the prediction of the concentration of a chemical specie as in chemical transport models.

The computation of the PKF dynamics for Eq. (15) using SymPKF leads to the system

$$\partial_t c = -u\partial_x c, \tag{16a}$$
$$\partial_t V_c = -u\partial_x V_c, \tag{16b}$$
$$\partial_t s_{c,xx} = -u\partial_x s_{c,xx} + 2s_{c,xx}\partial_x u, \tag{16c}$$

where the anisotropy is represented by the aspect tensor $\mathbf{s} = s_{c,xx}$ in 1D domain. The PKF dynamics Eq. (16) is a system of three uncoupled partial derivative equation similar to the one first found by Cohn (1993). This system represents the dynamics of the mean state $\mathbb{E}[c]$, Eq. (16a), where the expectation operator has been removed for the sake of simplicity ; the transport of the variance, Eq. (16b) ; and the transport of the anisotropy Eq. (16b), where an additional a source term of anisotropy appears, that is due to the shear by the flow. Compared with an ensemble approach, the PKF approach opens to an understanding of the dynamics and the physics of the uncertainty.

Note that the lower script notation $_c$ for V_c and $_{c,xx}$ for $s_{c,xx}$ corresponds to the notation automatically rendered by SymPKF when processing the dynamics Eq. (15) at a symbolic level. This labelling for the parameters has been introduced when multiple fields are present *e.g.* in multivariate dynamics. While this contribution only address univariate dynamics, the notation is kept here so to facilitate the comparison with the output of SymPKF and also because another important dynamics is discussed: the diffusion equation, which is now presented.

2.3.2 PKF prediction applied on a diffusion equation

The diffusion equation of a scalar field $f(t,x)$ and of diffusion coefficient $D(x)$,

$$\partial_t f = \partial_x \left(D\partial_x f\right), \tag{17}$$

is now considered. This kind of equation appears for instance in the prediction of electron density f of the Earth radiation belts and results from a Hamiltonian formalism

[2] `https://github.com/opannekoucke/pkf-boundary`

applied on a typical Boltzmann equation, where a Fokker-Planck operator is introduced to evaluate physical interactions responsible for changing particles trapping state (Dahmen et al., 2020). In the radiation belts, the typical spatial coordinates system x in Eq. (17) stands in this case for a combined spatial and physical quantities *e.g.* the energy of the electrons. The diffusion equation is also important in the modeling of atmospheric boundary layer where it represents the effect of the turbulence (Stull, 1988). In this example, and by identification with Eq. (1), f stands for χ while $\mathcal{M}(f, \partial f) = \partial_x (D\partial_x f)$.

The computation of the PKF dynamics for Eq. (17) can be performed using SymPKF. However, because of the second order derivative, the dynamical system makes appear an unknown term $\mathbb{E}\left[\varepsilon_f \partial_x^4 \varepsilon_f\right]$, not determined from f, V_f and $s_{f,xx}$ (see Appendix A). An analytical closure has been proposed for 1D domains which states as (Pannekoucke et al., 2018)

$$\mathbb{E}\left[\varepsilon_f \partial_x^4 \varepsilon_f\right] = 3g_{f,xx}^2 - 2\partial_x^2 g_{f,xx} \tag{18a}$$

when written in metric tensor or

$$\mathbb{E}\left[\varepsilon_f \partial_x^4 \varepsilon_f\right] = \frac{2\partial_x^2 s_{f,xx}}{s_{f,xx}^2} + \frac{3}{s_{f,xx}^2} - \frac{4\left(\partial_x s_{f,xx}\right)^2}{s_{f,xx}^3} \tag{18b}$$

in aspect tensor, which leads to the PKF dynamics

$$\partial_t f = D\partial_x^2 f + \partial_x D \partial_x f, \tag{19a}$$

$$\partial_t V_f = -\frac{2DV_f}{s_{f,xx}} + D\partial_x^2 V_f - \frac{D\left(\partial_x V_f\right)^2}{2V_f} + \partial_x D\partial_x V_f, \tag{19b}$$

$$\begin{aligned}\partial_t s_{f,xx} = D\partial_x^2 s_{f,xx} + 4D \\ - \frac{2D\left(\partial_x s_{f,xx}\right)^2}{s_{f,xx}} - \frac{2Ds_{f,xx}\partial_x^2 V_f}{V_f} + \\ \frac{D\partial_x V_f \partial_x s_{f,xx}}{V_f} + \frac{2Ds_{f,xx}\left(\partial_x V_f\right)^2}{V_f^2} - 2s_{f,xx}\partial_x^2 D + \\ 2\partial_x D\partial_x s_{f,xx} - \frac{2s_{f,xx}\partial_x D\partial_x V_f}{V_f},\end{aligned} \tag{19c}$$

where in this dynamical systems, the expected value $\mathbb{E}[f]$ in Eq. (19a) is replaced by f for the sake of simplicity. The dynamics Eq. (19) makes appear the effect of the transport due the heterogeneity of the diffusion coefficient which implies a flow of velocity $-\partial_x D$, and leads to the same PKF transport dynamics Eq. (16) as discussed for Eq. (15) in the particular case where $u = -\partial_x D$. The other terms in Eq. (19) are related to the second-order derivative term $D\partial_x^2 f$, which couples the dynamics of the variance and of the anisotropy.

In term of metric, the closed Eq. (19) reads as

$$\partial_t f = D\partial_x^2 f + \partial_x D\partial_x f, \tag{20a}$$

$$\partial_t V_f = -2DV_f g_{f,xx} + D\partial_x^2 V_f - \frac{D\left(\partial_x V_f\right)^2}{2V_f} + \partial_x D\partial_x V_f, \tag{20b}$$

$$\begin{aligned}\partial_t g_{f,xx} = -4Dg_{f,xx}^2 + D\partial_x^2 g_{f,xx} + \\ \frac{2Dg_{f,xx}\partial_x^2 V_f}{V_f} + \frac{D\partial_x V_f \partial_x g_{f,xx}}{V_f} - \frac{2Dg_{f,xx}\left(\partial_x V_f\right)^2}{V_f^2} + \\ 2g_{f,xx}\partial_x^2 D + 2\partial_x D\partial_x g_{f,xx} + \frac{2g_{f,xx}\partial_x D\partial_x V_f}{V_f}.\end{aligned} \tag{20c}$$

Until now, PKF dynamics for the heterogeneous diffusion equation has been evaluated on periodic domain only, while bounded domains are often needed, *e.g.* in radiation belts predictions where the energy of electrons are limited, or in atmospheric boundary layer where the ground is a limit of the domain. The next section addresses how to specify the boundary conditions for the PKF dynamics.

3 Specification of the PKF boundary conditions

This section challenges the specification of the boundary conditions for the PKF by considering two usual kind of conditions: the Dirichlet and the Neumann conditions. We consider the particular case of the semi-bounded 1D domain $[0, \infty)$, and focus on the boundary $x = 0$. Then we extend to boundary conditions of an arbitrary domain Ω of frontier $\partial\Omega$.

3.1 Dirichlet BCs

A Dirichlet condition at the boundary consists in specifying the value of the fields at $x = 0$, that is $\chi(t, x = 0) = \chi_0(t)$.

This conditions is used for the dynamics of the mean in the PKF, but it remains to specify the variance and the anisotropy for the boundary conditions.

Therefore the Dirichlet condition implies that the error field must also verifies a Dirichlet condition *i.e.* $e(t, x = 0) = e_0(t)$. The expectation of the error field at $x = 0$ is zero by definition, and of variance $V_0(t) = \mathbb{E}\left[e_0(t)^2\right]$. Hence, the variance field must also verify a Dirichlet condition *i.e.* $V(t, x = 0) = V_0(t)$.

So for a 1D bounded domain, the Dirichlet condition on the dynamics implies to specify a Dirichlet condition on the variance and on the anisotropy. This result extends for an arbitrary domain Ω where this time, the boundary conditions for the variance and the anisotropy are Dirichlet conditions on the frontier $\partial\Omega$.

In case where the bounded domain is nested within a larger domain where uncertainty is known from a PKF dynamics, then the variance and the anisotropy at the boundary can be set from the variance and the anisotropy known in the larger domain. When the uncertainty at large scale is featured from an ensemble of forecasts, the statistics at the boundary should be set as the statistics estimated from the ensemble of large scale forecasts at the boundary points *e.g.* for VLATcov models, the variance and the anisotropy can be estimated from the ensemble of large scale forecasts from Eq. (11) and Eq. (12) respectively.

Hence, Dirichlet condition in case of nested models easily extends in 2D and 3D domains where it remains to specify the variance and the anisotropy of the local area model from the variance and the anisotropy of the coupling model.

3.2 Neumann BCs

Neumann conditions at the boundaries write as null fluxes *i.e.* $\partial_x\chi(t, x = 0) = 0$. This implies that the error field must also verifies a Neumann condition *i.e.* $\partial_x e(t, x = 0) = 0$. Again, we are looking for the boundary conditions for the variance and the anisotropy.

The condition on the variance is deduced from the Taylor expansion of the error at the vicinity of $x = 0$ as follows. The expectation of the square of the second order expansion of the error

$$e(t, \delta x) = e(t, 0) + \frac{1}{2}\partial_x^2 e(t, 0)\delta x^2 + \mathcal{O}(\delta x^3),$$

leads to the local expansion of the variance $V(t, x) = \mathbb{E}\left[e^2\right](t, x)$,

$$V(t, \delta x) = V(t, 0) + \mathbb{E}\left[e\partial_x^2 e\right](t, 0)\delta x^2 + \mathcal{O}\left(\delta x^4\right).$$

As the local Taylor expansion of the variance field at $x = 0$, this implies that the first order derivative is null, *i.e.*

$$\partial_x V(t, 0) = 0, \tag{21a}$$

which means that the condition in variance at the boundary $x = 0$ follows a Neumann condition.

For the anisotropy, the Neumann condition on the variance,Eq. (21a), implies that the metric tensor $g(t,x) = \mathbb{E}\left[(\partial_x\varepsilon)^2\right](t,x)$ simplifies as $g(t,0) = \frac{1}{V(t,0)}\mathbb{E}\left[(\partial_x e(t,0))^2\right]$. Then the Neumann condition on e, $\partial_x e(t,0) = 0$ *i.e.* implies that the condition for the metric is a Dirichlet condition,

$$g(t, x = 0) = 0. \tag{21b}$$

Note that the later Dirichlet condition for the metric translates as the singular condition in aspect tensor, $s(t,0) = +\infty$, which makes appears that forecast step of the PKF written in aspect tensor is not well defined, and that is is preferable to consider the PKF as written in variance/metric tensors.

Hence, the Neumann condition for a 1D domain translates for the PKF as a Neumann condition in variance and a Dirichlet condition in metric. This extends to a 2D or 3D domain Ω where this time the Neumann condition for the variance states as a null flux along the normal direction of the frontier $\partial\Omega$ of the domain. The Dirichlet condition for the metric reads equivalently $g(t,\mathbf{x}) = 0$ for $\mathbf{x} \in \partial\Omega$, but a weaker condition could be introduced where the tangential components of the metric at the boundary are not zero (not addressed here).

Now that the boundary conditions for the PKF have been theoretically specified for the Dirichlet and the Neumann conditions, a numerical validation as well as a comparison with the usual EnKF approach is introduced. But to do so, it is necessary to specify an appropriate setting for the boundary of the EnKF, as discussed in the next section.

4 Specification of the BCs for EnKF simulations

The numerical validation of the PKF prediction, applied for bounded domains, is performed by considering an ensemble of forecasts approach. Compared to the PKF simulation, an ensemble approach relies on the computation of an ensemble of forecasts which requires an appropriate specification of the initial and boundary perturbation. In the realm of data assimilation applied to limited area models, the perturbation at the boundary often comes from an ensemble of global forecast *e.g.* forecast computed on the sphere for weather forecasting.

In the validation framework considered here, an appropriate specification of the boundary perturbation is needed so to compare to the PKF. This constraint of validation implies to introduce a way to specify the variance and the time-scale of the perturbation at the boundary in 1D domain in such a way that the perturbation of the initial condition (in space) are smoothly connected to the perturbation at the boundaries (in time). Indeed, to be representative to radiation belts dynamics, the boundary conditions have to be strongly dynamic over time. This is a constrain we take great care to analyze in order to test the PKF robustness to such an environment.

Generating a set of perturbations with such properties is achieved by sampling a multivariate normal distribution. The covariance matrix associated with the multivariate distribution contains the variance and the length-scales for the perturbation of the initial condition as well as for the perturbation of the boundary conditions.

Note that ensemble forecasting under Neumann boundary conditions, corresponds to an initial value problem where each member is integrated from an initial condition that verifies the Neumann conditions. Hence, the main difficulty encountered is how to specify the auto-correlation time scale of the perturbation at the boundaries for Dirichlet conditions.

In what follows, the specification of the time auto-correlation is first presented for an arbitrary evolution equation, then it is applied for the transport and for the diffusion equation.

4.1 Specification of the auto-correlation time-scale of the BC perturbations for ensemble of forecast

The problem faced here is that, with boundary conditions being time-series, the scale in the covariance matrix used to generate the set of perturbations is a time-scale. However, the metric tensor g_{xx} is related to the spatial length-scale of the perturbation as denoted by the index $_{xx}$. In order to specify the boundary condition of the metric tensor field, we need to find an equation linking, on the boundaries, the spatial metric tensor g_{xx} with the time-scale used to generate the perturbation.

Similarly to the spatial metric tensor Eq. (10), the temporal metric tensor g_{tt} that characterize the auto-correlation of a smooth centered random field, $\eta(t)$, depending on the time, and of variance $V_\eta(t) = \mathbb{E}\left[\eta(t)^2\right]$, is defined by

$$\mathbf{g}_{tt}(t) = \mathbb{E}\left[\partial_t\left(\frac{\eta(t)}{V_\eta(t)}\right)\partial_t\left(\frac{\eta(t)}{V_\eta(t)}\right)\right]. \tag{22}$$

This temporal metric tensor is directly related with the time-scale of the perturbation. In 1D $g_{tt} = \frac{1}{L_t^2}$ with L_t the auto-correlation time-scale.

Without loss of generality, the boundary $x = 0$ is considered, and the goal is to characterize the temporal metric tensor $g_{tt}(t, x = 0)$. If $\eta(t)$ denotes the random error at $x = 0$, then by continuity, the error and the random forcing verify $e(t, x = 0) = \eta(t)$. Then, it results that the variances verify $V_\eta(t) = V(t, x = 0)$, and the temporal metric tensor reads as

$$\mathbf{g}_{tt,x=0}(t) = \mathbb{E}\left[\partial_t\varepsilon(t, \mathbf{x} = 0)\partial_t\varepsilon(t, \mathbf{x} = 0)\right], \tag{23a}$$

where $\varepsilon = e/\sqrt{V}$ is the normalized error associated with the spatial error e. While Eq. (23a) only holds at the boundary $x = 0$, the spatio-temporal smoothness of e implies a link between the temporal metric at the boundary and the spatial metric within the domain, which results from the dynamics of the error Eq. (2) at $x = 0$: $\partial_t e(t, x = 0) = \mathcal{M}(e, \partial e)(t, x = 0)$. In particular, the temporal metric reads as (see Appendix B)

$$g_{tt} \underset{x=0}{=} \frac{1}{V}\mathbb{E}\left[\left(\mathcal{M}(\varepsilon\sqrt{V}, \partial(\varepsilon\sqrt{V}))\right)^2\right] - \frac{1}{4V^2}\left(\partial_t V\right)^2, \tag{23b}$$

where the terms $\mathbb{E}\left[\left(\mathcal{M}(\varepsilon\sqrt{V}, \partial(\varepsilon\sqrt{V}))\right)^2\right]$ and $\partial_t V$ can make appear the spatial metric field g_{xx} at $x = 0$.

One pitfall is that equation Eq. (23b) may be complicated, and can contain unknown terms such as $\mathbb{E}\left[\varepsilon_f\partial_x^4\varepsilon_f\right]$ encountered for the heterogeneous diffusion dynamics in section 2.3.2. The next two sub-sections will detail the link between the temporal and the spatial metrics for the transport and for the diffusion.

4.2 Dirichlet BC for ensemble forecasting of the positive velocity transport equation

To illustrate the relation between the temporal and the spatial metric tensor, the transport equation Eq. (15) is now considered.

Following the theoretical derivation of the temporal metric Eq. (23b), the computation using SymPKF leads to the relation between the auto-correlation time-scale of the boundary perturbation and the spatial error anisotropy tensor that reads as

$$g_{c,tt} \underset{x=0}{=} u^2 g_{c,xx} + \frac{u^2 \left(\partial_x V_c\right)^2}{4V_c^2} + \frac{u\partial_t V_c \partial_x V_c}{2V_c^2} + \frac{\left(\partial_t V_c\right)^2}{4V_c^2}. \tag{24}$$

This spatio-temporal consistency for the temporal and spatial statistics is difficult to interpret physically without approximations. However, under the assumptions of local homogeneity ($\partial_x V_c = 0$) and of stationarity for the variance $\partial_t V_c = 0$), Eq. (24) reads as

$$g_{c,tt} \underset{x=0}{=} u^2 g_{c,xx}, \tag{25}$$

which is physically interpretable since Eq. (25), written in time-scale and length-scale, reads as $L_t = \frac{L_x}{u}$: the usual rule relating time and space in a transport. Later, the numerical investigation will consider Eq. (25) as an approximation of the true time-scale even when assumptions leading to Eq. (25) are not verified.

Note that Eq. (25) can be obtained when considering that the dynamics of the variance Eq. (16b) applies at the boundary, leading to replace the trend of the variance by $\partial_t V_c = -u\partial_x V_c$ in Eq. (24) so to obtain Eq. (25).

To conclude this paragraph, the ensemble forecasting under Dirichlet boundary conditions and applied to the transport equation, remains to populate an ensemble of boundary perturbations with a prescribed temporal variance and an auto-correlation time scale given by Eq. (25).

We proceed in the same way for the diffusion equation.

4.3 Dirichlet BCs for ensemble forecasting of the diffusion equation

To continue going towards more and more realistic modeling, the heterogeneous diffusion equation Eq. (17) is now considered to compute the spatio-temporal link Eq. (23b) in the diffusion case.

Form a derivation detailed in Appendix C, the auto-correlation time scale of boundary perturbation can be related to the spatial error correlation length-scale by the proxy

$$g_{f,tt}(t,x) \approx 3D(x)^2 g_{f,xx}(t,x). \tag{26}$$

Note that Eq. (26) is an equality when the variance and the diffusion fields are homogeneous, and when the variance is stationary at the boundary.

To conclude this paragraph, the ensemble forecasting under Dirichlet boundary conditions, and applied to the diffusion equation, remains to populate an ensemble of boundary perturbations with a prescribed temporal variance and an auto-correlation time scale given by Eq. (26).

We are now ready to validate the PKF approach from an ensemble validation designed to produce desired error statistics.

5 Numerical investigation

The goal of the numerical investigation is to validate the PKF on a bounded domain as well as the equations developed in Section 4, by comparing the PKF dynamics with an ensemble simulation.

5.1 Setting for the numerical experiments

For this investigation three different settings are considered. All experiments take place on a 1D bounded domain $x \in [0, \Lambda]$. For the first one, the transport equation

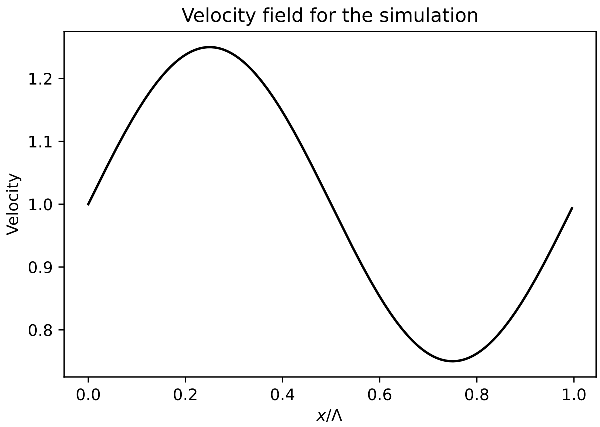

Figure 1: Heterogeneous velocity field considered for the numerical simulation of the transport dynamics.

Eq. (15) is considered with Dirichlet boundary condition at $x = 0$ and free boundary at $x = \Lambda$. For the second setting, the heterogeneous diffusion equation Eq. (17) is considered with Dirichlet boundary conditions at both boundaries $x = 0$ and $x = \Lambda$. For the third setting, the same diffusion equation is considered but this time with Neumann boundary conditions at $x = 0$ and $x = \Lambda$.

The transport and the diffusion being linear, the dynamics of the mean is the same for the PKF and for the EnKF. Hence, without loss of generality, to focus on the validation of the error statistics, the mean state is not considered in the following (the reader can consider the mean state as constant). Then, the ensemble of forecast is equivalent to the forecasts of an ensemble of perturbations $(e_k)_{k\in[1,N_e]}$, with appropriate boundary conditions.

Each time the variance, Eq. (16b) and Eq. (19b), and anisotropy tensor, Eq. (16c) and Eq. (19c), produced by the PKF dynamics are compared with the variance and anisotropy tensor diagnosed from an ensemble of $N_e = 6400$ forecasts.

The domain is discretized in $n = 241$ grid points and the spatial derivative operator ∂_x is discretized with a centered finite difference scheme leading to a second order of consistency. The temporal discretization scheme varies with each experiment and is detailed in each sections.

5.2 Application to the transport equation

In this experiment setting, the transport equation Eq. (15) is considered. The velocity wind for the simulation is set as the heterogeneous stationary field $u(x) = 1 + \frac{1}{4}\sin(\frac{2\pi}{\Lambda}x)$ shown in Fig. 1.

The temporal discretization scheme used for the ensemble simulation as well as the PKF dynamics is a Runge-Kutta scheme of order 4 with a fixed time-step $dt \approx 4.10^{-3}$. The simulation is conducted from time $t = 0$ until $t_{end} = 2T_{adv}$ with the advection time scale $T_{adv} = \frac{\Lambda}{u_{max}}$.

In order to generate a coherent set of perturbations for the ensemble simulation *i.e.* an initial condition and a boundary condition that are smoothly connected, an extended domain $[-u(0)t_{end}, \Lambda]$ is created from the union of the physical domain $[0, \Lambda]$ and

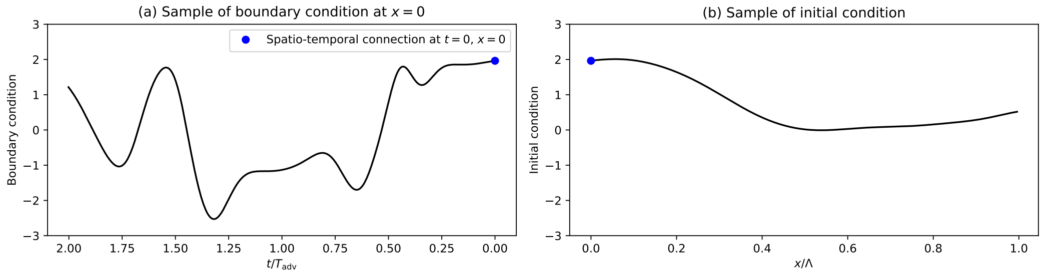

Figure 2: Sample of a generated perturbation split into an initial condition and a boundary condition that are smoothly connected.

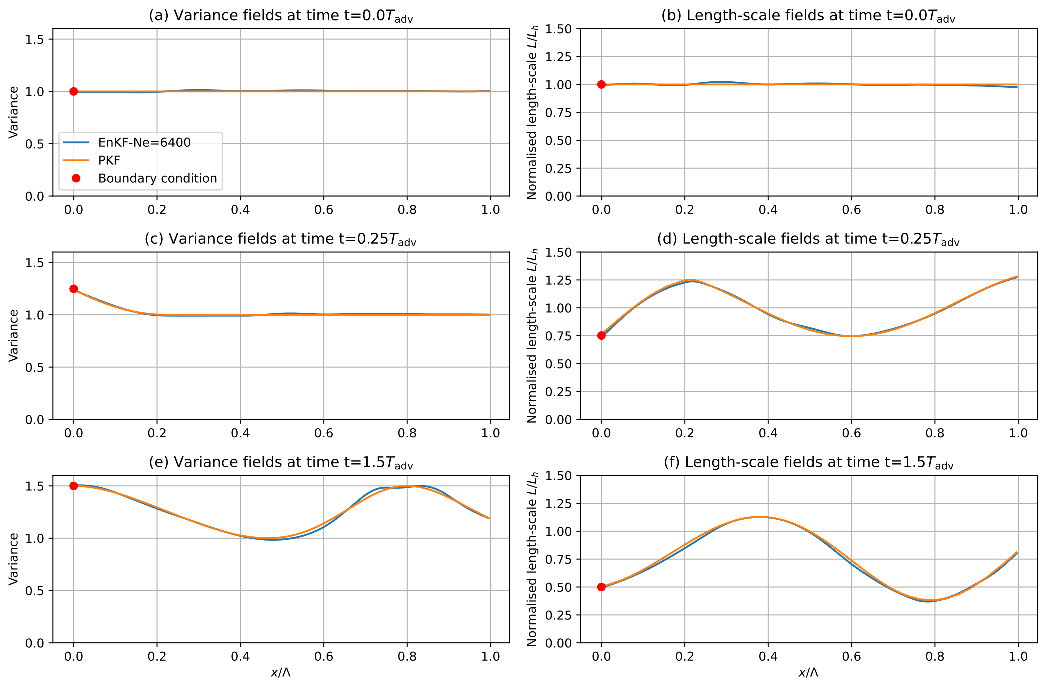

Figure 3: Comparison of the forecast-error variance (left column) and normalized length-scale (right column) fields dynamics for the heterogeneous advection equation on a 1D bounded domain with Dirichlet boundary conditions at $x = 0$ and open boundary condition at $x = \Lambda$. The results are shown for times $t = 0$, $t = 0.25T_{\text{adv}}$ and $t = 1.5T_{\text{adv}}$.

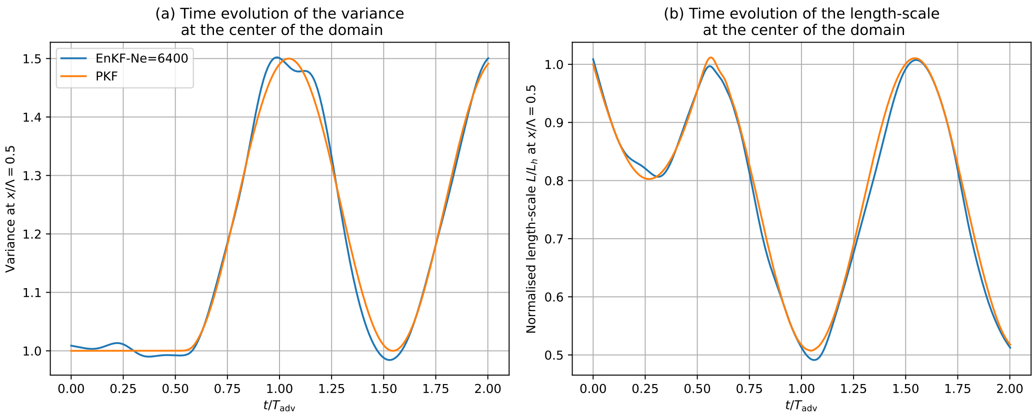

Figure 4: Time evolution of the forecast-error variance (a) and normalized length-scale (b) at $x = 0.5\Lambda$, for the advection equation with Dirichlet boundary conditions.

the time window $[0, t_{end}]$ brought back to a virtual physical extension of the domain by multiplying with $-u(0)$.

Then on this extended domain a variance field, V_0 and a length-scale field L_0 are defined which will be used to generate the perturbations. For this experiment the fields V_0 and L_0, that constitute the PKF initial and boundary conditions, are set as follows. The initial variance is set homogeneous and equal to 1 over the physical domain $V_0(t = 0, x) = 1$ and the boundary variance is set to the periodical function $V_0(t, x = 0) = \frac{5}{4} - \frac{1}{4}cos(\frac{2\pi}{T_{adv}}t)$. Like for the variance, the initial length-scale is set homogeneous and equal to 10% of the domain length $L_0(t = 0, x) = L_h = 0.1\Lambda$ and the boundary length-scale is set to the periodical function $L_0(t, x = 0) = 0.1\Lambda(\frac{3}{4} + \frac{1}{4}cos(\frac{2\pi}{T_{adv}}t))$.

This setting for the variance and the length-scale is chosen so to represent a typical behaviour encountered in numerical weather forecasting, where large scale are more predictable than small scales, which is also the case in radiation belts dynamics forecasting.

Using Eq. (13) and the relation between the length-scale and the anisotropy tensor in 1D, $s_0 = L_0^2$, the covariance matrix, $\mathbf{P}_0 = \mathbf{P}(V_0, s_0)$, is defined from which the spatio-temporal perturbations are sampled for each k as $e_k = \mathbf{P}_0^{1/2}\zeta_k$, where ζ_k is a sample of a centered and normalized Gaussian random vector, and where $\mathbf{P}_0^{1/2}$ stands for the square root matrix of $\mathbf{P}_0$, *i.e.* $\mathbf{P}_0 = \mathbf{P}_0^{1/2}\left(\mathbf{P}_0^{1/2}\right)^{\mathrm{T}}$. The square root $\mathbf{P}_0^{1/2}$ has been computed from the singular value decomposition of the matrix $\mathbf{P}_0$.

An example of a perturbation sample is presented in Fig. 2 where the temporal evolution $e(t, x = 0)$ is shown in panel (a) while the initial condition within the domain, $e(t = 0, x)$ is given in panel (b). Note that the time axis in panel (a) has been inverted so to facilitate the understanding. The blue dots corresponds to the value of the sampled error field e at $t = 0$ and $x = 0$.

The figure Fig. 3 shows both variance and length-scale fields that are computed from the PKF and the ensemble simulations and compared at three different timestamps. The first panels (a) and (b) respectively show the variance and the length-scale normalized by L_h at initial time. As prescribed, the initial variance is homogeneous and equal to 1. The initial length-scale is also homogeneous and equal to L_h. The panels (c) and (d) present the evolution at time $t = 0.25T_{adv}$. As expected the variance (the length-scale) increases (decreases) according to the specified boundary condition close to $x = 0$ (red dot). The length-scale is also modified over the whole domain, this is the effect of the velocity gradient of the term $2s_{c,xx}\partial_x u$ in Eq. (16c). Since the velocity field is positive, the variance and the length-scale are transported to the right of the domain. Finally, panels (e) and (f) show the fields at time $t = 1.5t_{adv}$. The information injected by the boundary condition at $x = 0$ has reached the other side of the domain unscathed for the variance and length-scale fields.

In order to strengthen these results, we show in Fig. 4 the evolution through time of the fields for the middle point in the domain $x = 0.5\Lambda$. As expected, the variance in panel (a) remains constant until the information from the boundary condition arrives, where oscillations start, following the prescribed sine shape of the boundary condition shifted in time. In panel (b), the length-scale follows the same kind of dynamic except that the length-scale varies from $t = 0$ to $t = 0.5T_{adv}$, a variation that is not due to the boundary condition but to the heterogeneity of the wind field. Note that ensemble estimation of the variance and of the length-scale are subject to some sampling noise even with the large ensemble size $N_e = 6400$.

Overall, this simulation shows no numerical artifact and the PKF and EnKF forecasts overlap perfectly. Moreover, the continuous and differentiable error statistics of the EnKF statistics shows that the generated duets of errors for the initial condition and boundary condition have been appropriately specified.

These results validate that the specification of the PKF boundaries proposed in Section 3.1 is correct when applying Dirichlet condition in a transport dynamics. Moreover it also validates the specification of the perturbations Eq. (25), introduced in Section 4, for the ensemble validation to build prescribed error statistics.

Note that, this example has also shown the ability of the PKF to apply for open boundary condition.

Now, we validate the PKF boundary conditions applied for a diffusion equation.

5.3 Application to the diffusion equation

In this experiment setting, the heterogeneous diffusion Eq. (17) is considered. The temporal discretization scheme used for the ensemble simulation is a backward Euler scheme (implicit Euler method) with a fixed time-step $dt_{BE} \approx 2.10^{-4}$. For the PKF dynamics we used a Runge-Kutta scheme of order 4 with a fixed time-step $dt_{RK4} \approx 5.10^{-6}$. The simulation is performed from time $t = 0$ to $t_{end} = 1.2T_{\text{diff}}$ with $T_{\text{diff}} = \frac{\Lambda^2}{4D_{max}}$ the time scale of the diffusion of a half-domain.

The diffusion coefficient for the simulation is set as the heterogeneous stationary field $D(x) = 1 + \frac{A}{A_{max}}$ with $A(x) = sin(\pi x)(1 + x)^8$ where $A_{\max} = Max_x A(x)$, and is shown in Fig. 5. This diffusion field reproduces the kind of diffusion encountered in the dynamics of radiation belts in order to evaluate the ability of PKF to solve this problem.

5.3.1 Dirichlet boundary conditions

To generate a coherent set of perturbations for the ensemble simulation, the same technique described in Section 5.2 is considered except that both boundaries at $x = 0$ and $x = \Lambda$ are subject to Dirichlet conditions. The extended domain considered is $[-\sqrt{D(0)t_{end}}, 0] \cup [0, \Lambda] \cup [\Lambda, \Lambda + \sqrt{D(\Lambda)t_{end}}]$.

This time, the parameters considered for the simulation and ensemble generation are as follows, the initial variance is set to the linear function $V_0(t = 0, x) = 1 + \frac{3}{\Lambda}x$ and the initial length-scale is set homogeneous and equal to 10% of the domain length

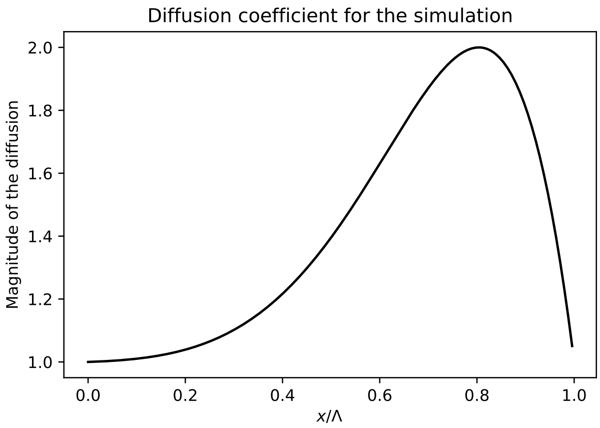

Figure 5: Heterogeneous diffusion coefficient generated for the experiment

$L_0(t=0,x) = L_h = 0.1\Lambda$. For the left boundary condition at $x=0$, the variance and the length-scale are stationary set equal to 1 and L_h respectively *i.e.* $V_0(t,x=0) = 1$ and $L_0(t,x=\Lambda) = L_h$. For the right boundary condition at $x=\Lambda$, the variance and the length-scale are stationary set equal to 4 and L_h respectively *i.e.* $V_0(t,x=\Lambda) = 4$ and $L_0(t,x=\Lambda) = L_h$. From this specification, an ensemble of perturbations has been populated following the same procedure, $e_k = \mathbf{P}_0^{1/2}\zeta_k$, as detailed in Section 5.2. The resulting perturbations are similar to the ones shown in Fig. 2 for the advection, except that there is a right extension of the domain in addition of the left extension for the advection (not shown).

The comparison between the PKF and EnKF predictions at different time steps are shown in Fig. 6. The first panels (a) and (b) are coherent with the specification of the initial condition for both the EnKF and the PKF. Panels (c) and (d) show the evolution of the variance and length-scale at $t = 0.2T_{\text{diff}}$.

Due to physical diffusion, far from the boundaries *e.g.* at the center of the domain, the magnitude of the error is expected to decrease over time with an attenuation of the variance, while the length-scale should increase; and at the boundaries the uncertainty should remained as specified by the Dirichlet conditions. This is precisely the behaviour observed for both the EnKF and the PKF, at the center of the domain and at the boundaries where the Dirichlet condition imposes fixed values for the variance and the length-scale on both sides of the domain.

However, panel (d) shows a noticeable gap between the length-scale computed by the PKF and the one estimated from the ensemble. This gap can be due to the closure Eq. (18) but it has a limited impact on the variance field (panel c) which makes appear that the PKF prediction of the variance is an accurate proxy for the EnKF estimation.

On the last panels (e) and (f), the variance and the length-scale settle down and the values predicted by the PKF are close to the values computed from the ensemble except for the error observed between the length-scale fields in the middle of the domain. As seen in Fig. 7, the variance and length-scale are close to the permanent regime at $t = 1.2T_{\text{diff}}$ showing that the PKF well performed even over a significant time period.

To conclude, this experiment has confirmed the specification of the Dirichlet boundary conditions of Section 3.1 for the PKF applied to a heterogeneous diffusion equation. It has shown the ability of the PKF to accurately approximate the uncertainty dynam-

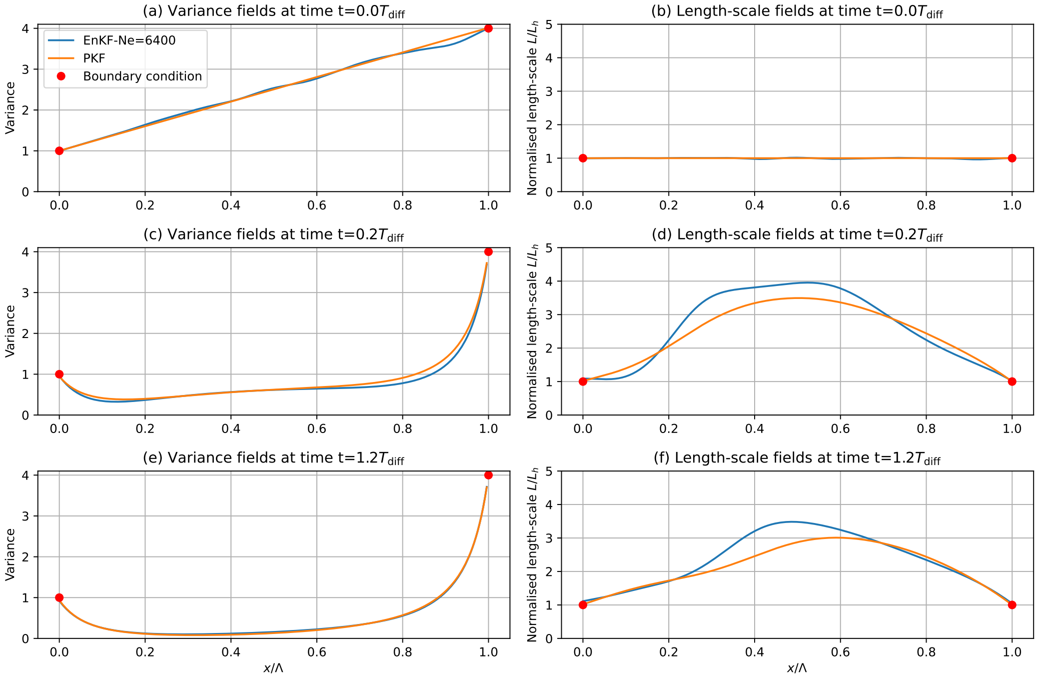

Figure 6: Comparison of the forecast-error variance (left column) and normalized length-scale (right column) fields dynamics for the heterogeneous diffusion equation on a 1D bounded domain with Dirichlet boundary conditions, and shown at times $t = 0$, $t = 0.2T_{\text{diff}}$ and $t = 1.5T_{\text{diff}}$.

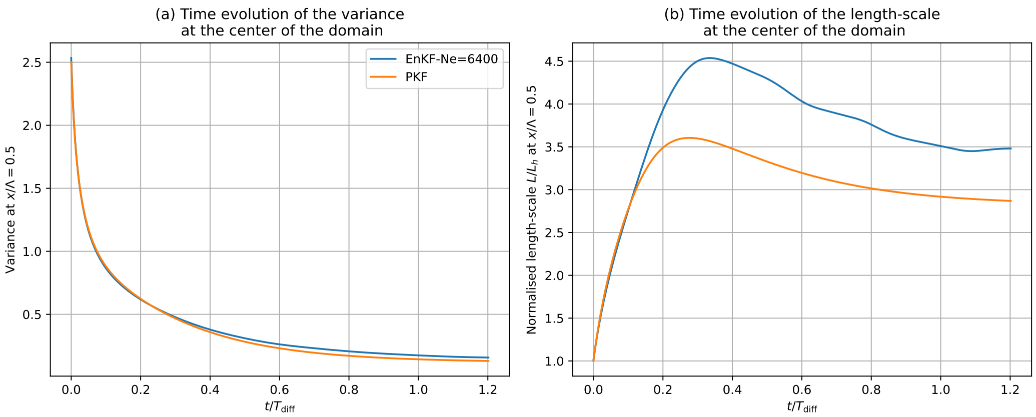

Figure 7: Time evolution of the forecast-error variance (a) and normalized length-scale (b) at $x = 0.5\Lambda$, for the diffusion equation with Dirichlet boundary conditions.

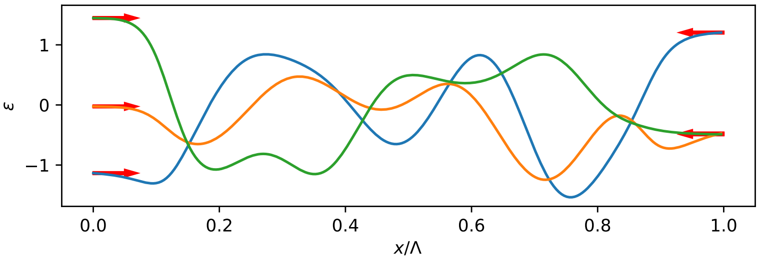

Figure 8: Samples of random error generated as initial condition that verify the Neumann condition at the boundaries $x = 0$ and $x = \Lambda$.

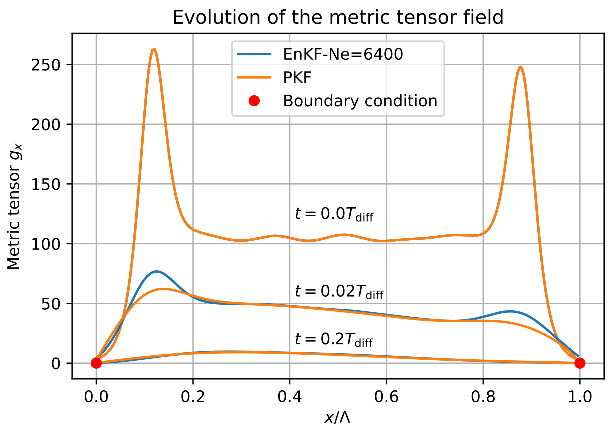

Figure 9: Forecast-error metric field for the heterogeneous diffusion equation on a 1D bounded domain with Neumann boundary conditions, shown at times $t = 0$, $t = 0.2T_{\text{diff}}$ and $t = 1.2T_{\text{diff}}$.

ics as diagnosed from the EnKF but at a lower cost corresponding the price of two time integration compared with the 6400 integrations needed for the ensemble. Another result is that the simulations also validate the theoretical derivation of the time-scale setting Eq. (26) needed to obtain a specific length-scale at the boundaries.

We end the numerical validation by considering the Neumann conditions applied to the heterogeneous diffusion equation.

5.3.2 Neumann boundary conditions

As above mentioned in Section 4, compared with the Dirichlet, the Neumann conditions are simulated in an ensemble of forecasts, as an initial condition problem without perturbation at the boundaries. The problem is then to produce an ensemble of initial conditions that verify the Neumann conditions.

To do so, a covariance model based on a homogeneous pseudo-diffusion equation has been considered (Weaver & Courtier, 2001). The terminology *pseudo* means that the diffusion is not physical but only a tricky way to create large covariance model as used in variational data assimilation. In particular the square-root covariance $\mathbf{P}_0^{1/2}$ resulting

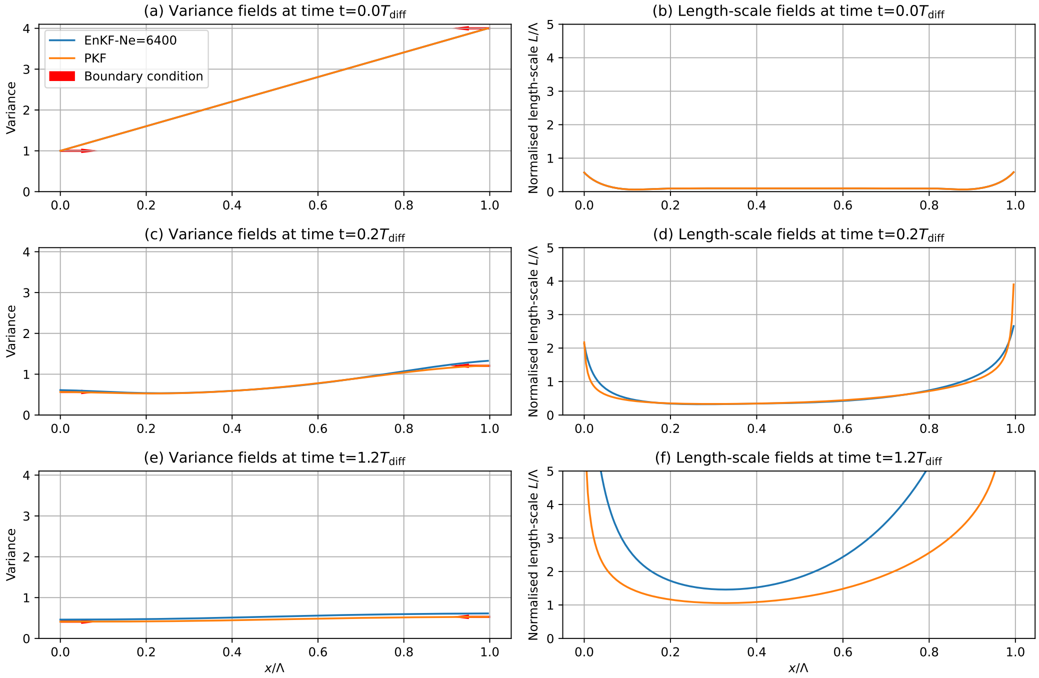

Figure 10: Comparison of the forecast-error variance (left column) and length-scale (right column) fields dynamics for the heterogeneous diffusion equation on a 1D bounded domain with Neumann boundary conditions, and shown at times $t = 0$, $t = 0.2T_{\mathrm{diff}}$ and $t = 1.2T_{\mathrm{diff}}$.

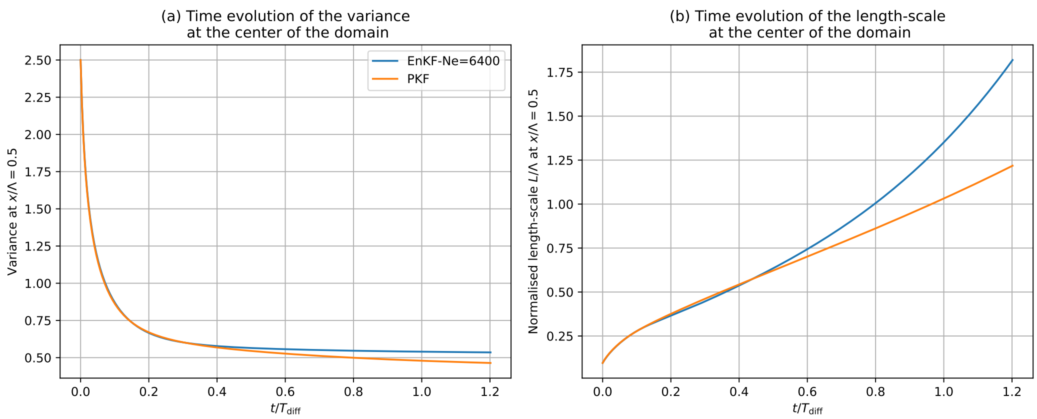

Figure 11: Time evolution of the forecast-error variance (a) and length-scale (b) at $x = 0.5\Lambda$, for the diffusion equation with Neumann boundary conditions.

from the integration of the pseudo-diffusion equation reads as the linear operator

$$\mathbf{P}_0^{1/2} = \mathbf{\Sigma}\mathbf{W}\mathbf{L}, \tag{27}$$

where $\mathbf{L} = e^{\frac{1}{2}\kappa\partial_x^2}$ is the propagator associated with the diffusion equation

$$\partial_\tau u = \kappa\partial_x^2 u \tag{28}$$

of pseudo-time τ, integrated from $\tau = 0$ to $\tau = \frac{1}{2}$, and using Neumann conditions at the boundaries (Mirouze & Weaver, 2010); $\mathbf{W}$ is a diagonal normalisation so that $\mathbf{WL}(\mathbf{WL})^{\mathrm{T}}$ is a correlation operator ; and $\mathbf{\Sigma}$ is a diagonal matrix of standard deviations, so that the spatial variance field is the linear profile with $V_0(t = 0, x = 0) = 1$ and $V_0(t = 0, x = \Lambda) = 4$. Note that the pseudo-diffusion coefficient κ is related to the length-scale l of the correlation functions as to $\kappa = l^2/2$ (Pannekoucke & Massart, 2008). For the numerical application, $l_h = 0.1\Lambda$.

Again, an ensemble of initial conditions are populated from the square-root Eq. (27), $e_k = \mathbf{P}_0^{1/2}\zeta_k$, where ζ_k is a sample of centered Gaussian random vector. Fig. 8 shows some samples of the normalized error resulting from Eq. (27) *i.e.* $\varepsilon_k = \mathbf{WL}\zeta_k$. As it is expected, the normalized error are plate at the boundaries (red arrows pointing toward the interior of the domain). The resulting anisotropy diagnosed from the ensemble of initial condition $t = 0$ leads to the metric field shown in Fig. 9 (in blue but super-imposed by the orange line). As expected, far from the boundary, the metric is homogeneous equal to $g = 1/l_h^2$ *i.e.* near $x = 0.5\Lambda$, but oscillates near the boundaries to reach a value of zero at the boundaries. The oscillations is due to the constraint of symmetry of the covariance matrix (Pannekoucke et al., 2018).

As discussed in Section 3.2, for Neumann conditions the PKF dynamics is solved following its metric formulation, which is given by Eq. (20) for the physical diffusion equation Eq. (17). For the numerical validation of the Neumann BCs, the initial condition for the PKF is the variance field of linear profile shown in Fig. 10-(a) and the experiment metric field diagnosed from the ensemble of initial conditions shown in Fig. 9 for $t = 0$ (orange line, superimposed to the blue line of the EnKF diagnosis).

The PKF dynamics is computed and the results are compared with the ensemble of forecasts of the heterogeneous diffusion equation Eq. (17) and Neumann conditions on both sides. The results are shown in Fig. 9 for the variance and the length-scale (computed from the inverse of the metric), and in Fig. 9 for the metric. The results are shown for times of interest selected from the time evolution reproduced in Fig. 11 where a relaxation toward a stationary state of uncertainty appears.

As expected for a diffusion, the variance decreases along the time, while the length-scale increases. Note that for Neumann condition, the variance at the boundary also decreases while it was constant in the Dirichlet condition. For the ensemble estimation, the length-scales at the boundaries (blue lines in panel (b-d-f)) are large but finite where it is expected to be infinite: this is due to the numerical estimation of the length-scale deduced from Eq. (12), while the metric remains zero at the boundaries during the simulation (see Fig. 9 the red dots).

Compared with the EnKF diagnosis, the PKF perform well by reproducing the same behaviour of the uncertainty dynamics as for the EnKF, except that the length-scale predicted from the PKF underestimates the length-scale diagnosed from the ensemble. However, the very large length-scale values, larger than the domain size Λ, as diagnosed from the ensemble is subject to the limitation of the numerical computation of Eq. (12) for large correlations that can present a positive bias (Pannekoucke et al., 2008). Moreover, the large length-scale of the EnKF can also be influenced by model error (Pannekoucke et al., 2021). Because of these limitations, it is not certain that the EnKF reproduces the true dynamics of the uncertainty for these extreme values of the length-scales, while it is considered as the reference. Hence, the discrepancy between the PKF and the EnKF reference, may not be due to a defect of the PKF that could be better than the ensemble estimation here.

To conclude, this experiment has confirmed the specification of the Neumann boundary conditions of Section 3.2 for the PKF applied to a heterogeneous diffusion equation. It has shown the ability of the PKF to accurately approximate the uncertainty dynamics as diagnosed from the EnKF but at a lower cost corresponding the price of two time integration compared with the 6400 integrations needed for the ensemble.

This ends the validation of the specification of the boundary conditions for the PKF. The summary of the results obtained in the paper as well as the perspective of the work are given in the following section of conclusion.

6 conclusion

This work contributed to explore the parametric Kalman filter (PKF), that is a recent approximation of the Kalman filter proposed for application in large systems. The parametric approach investigated here consist to approximate the forecast-error covariance matrix by a covariance model parameterized from the variance and the anisotropy. The anisotropy can be specified in term of metric tensor or its inverse, the aspect tensor, that is the square of the length-scale in 1D domains. The PKF dynamics describes how the mean, the variance and the anisotropy evolve in time, leading to a low cost prediction of the error statistics that is the full covariance propagation in the Kalman filter or an ensemble of forecast in the ensemble Kalman filter approximation.

In this contribution, we proposed how to specify the error statistics at the boundary of a domain when considering a PKF forecast of the uncertainty. We detail here pragmatic solutions for large systems with strong variability at their domain's edge, such as atmospheric weather and radiation belts "weather".

Two kind of boundaries have been considered, the Dirichlet and the Neumann conditions depending on the dynamics. We obtained that the Dirichlet condition for the dynamics translates for the PKF dynamics as Dirichlet conditions for the variance and the metric or the aspect tensor. For Neumann conditions, the PKF conditions are Neumann for the variance and Dirichlet for the metric, and the formulation of the PKF in metric is more adapted than in aspect tensor which would required infinite boundary conditions.

The theoretical specification of the boundary conditions has been tested and validated for two important dynamics: the transport and the diffusion equation. Both dynamics are important for weather forecasting, air quality or radiation belt dynamics, which are some of the problems we are interested in.

To validate the specification of the boundary conditions and to evaluate the accuracy of the PKF to predict the dynamics of the uncertainty, a numerical test-bed has been considered in a 1D domain for the advection and the diffusion equation. An ensemble of forecast has been considered as a reference, where appropriate time-scale for the perturbation of the boundaries have been proposed in this paper for Dirichlet conditions and dependent on the dynamics. For both the advection and the diffusion, the PKF has been shown able to reproduce the uncertainty dynamics diagnosed from the ensemble of forecast. This indirectly validates the time-scale introduced to create the boundary perturbations introduced for the ensemble of forecast, and constitutes a contribution to the ensemble methods while it is not needed for the PKF.

In particular, it appears that the specification of boundary conditions in the PKF is much easier than for the EnKF, that needs perturbations of the boundary for Dirichlet conditions or plate error at the boundary for Neumann conditions. While in practice, EnKF applied for bounded domains often relies on ensemble computed on larger domain *e.g.* in weather forecasting, in some applications, no such larger simulation is available *e.g.* in radiation belts forecasting. As already demonstrated for the understanding of the model error covariance, the PKF approach provides new tools to better understand and modeled the dynamics of uncertainty that is of interest not only for the PKF itself, but also for the widely used ensemble methods.

The next step will be to study the BC conditions for domains of larger dimensions, where we expect some changes *e.g.* non-zero components of the metric tensor along the tangential direction to the boundary in Neumann conditions.

We can mention that the dynamics of the uncertainty for bounded domains can be of importance in variational data assimilation or observation targeting applied for local area models, that could be another topics to investigate with the PKF.

Beyond these challenging topic, we can mention that the results in 1D should already found important applications *e.g.* in the dynamics of uncertainty in the boundary layer for air quality, wild-land fire predictions or atmosphere-ocean coupling.

Appendix A Closure of the PKF Dynamics for the diffusion equation

The computation of the PKF dynamics for the diffusion equation Eq. (17), with SymPKF, leads to the dynamical system

$$\partial_t f = D\partial_x^2 f + \partial_x D\partial_x f, \tag{A1a}$$

$$\partial_t V_f = -\frac{2DV_f}{s_{f,xx}} + D\partial_x^2 V_f - \frac{D\left(\partial_x V_f\right)^2}{2V_f} + \partial_x D\partial_x V_f, \tag{A1b}$$

$$\begin{aligned}\partial_t s_{f,xx} = 2Ds_{f,xx}^2\mathbb{E}\left(\varepsilon_f\partial_x^4\varepsilon_f\right) - 3D\partial_x^2 s_{f,xx} - 2D + \\ \frac{6D\left(\partial_x s_{f,xx}\right)^2}{s_{f,xx}} - \frac{2Ds_{f,xx}\partial_x^2 V_f}{V_f} + \frac{D\partial_x V_f\partial_x s_{f,xx}}{V_f} + \\ \frac{2Ds_{f,xx}\left(\partial_x V_f\right)^2}{V_f^2} - 2s_{f,xx}\frac{d^2}{dx^2}D + \\ 2\partial_x D\partial_x s_{f,xx} - \frac{2s_{f,xx}\partial_x D\partial_x V_f}{V_f}\end{aligned} \tag{A1c}$$

where this time the term $\mathbb{E}\left[\varepsilon_f\partial_x^4\varepsilon_f\right]$ is not determined from f, V_f and $s_{f,xx}$. This dynamics can be closed considering the closure Eq. (18).

Appendix B Specification of the temporal metric tensor for evolution equations

This section details the link between the temporal metric Eq. (23a), $\mathbf{g}_{tt} = \mathbb{E}\left[\partial_t\varepsilon\partial_t\varepsilon\right]$, and the dynamics of the error. Since the trend of the normalized error reads as

$$\partial_t\varepsilon = \frac{1}{\sqrt{V}}\partial_t e - \frac{1}{2V^{3/2}}e\partial_t V, \tag{B1}$$

then the temporal metric tensor writes as

$$g_{tt} = \frac{1}{V}\mathbb{E}\left[\left(\partial_t e\right)^2\right] - \frac{1}{V^2}\mathbb{E}\left[e\partial_t e\right]\partial_t V + \frac{1}{4V^3}\mathbb{E}\left[e^2\right]\left(\partial_t V\right)^2. \tag{B2}$$

However, we recognize the expression of the variance $V = \mathbb{E}\left[e^2\right]$ and its trend, Eq. (14a), so that the temporal metric simplifies as

$$g_{tt} = \frac{1}{V}\mathbb{E}\left[\left(\partial_t e\right)^2\right] - \frac{1}{4V^2}\left(\partial_t V\right)^2. \tag{B3a}$$

Introducing the trend of the error Eq. (2) and by definition of $\varepsilon = e/\sqrt{V}$, the temporal metric reads as

$$g_{tt} = \frac{1}{V}\mathbb{E}\left[\Big(\mathcal{M}(\varepsilon\sqrt{V}, \partial(\varepsilon\sqrt{V}))\Big)^2\right] - \frac{1}{4V^2}\left(\partial_t V\right)^2. \tag{B3b}$$

Appendix C Time auto-correlation boundary condition for the diffusion equation

The computation of the time auto-correlation metric Leverages on SymPKF. For the diffusion equation, SymPKF leads to

$$\begin{aligned} g_{f,tt} = {} & D^2\mathbb{E}\left(\varepsilon_f\partial_x^4\varepsilon_f\right) + 2D^2\partial_x^2 g_{f,xx} - \frac{D^2 g_{f,xx}\partial_x^2 V_f}{V_f} + \\ & \frac{D^2\partial_x V_f\partial_x g_{f,xx}}{V_f} + \frac{3D^2 g_{f,xx}\left(\partial_x V_f\right)^2}{2V_f^2} + \frac{D^2\left(\partial_x^2 V_f\right)^2}{4V_f^2} \\ & - \frac{D^2\left(\partial_x V_f\right)^2\partial_x^2 V_f}{4V_f^3} + \frac{D^2\left(\partial_x V_f\right)^4}{16V_f^4} + D\partial_x D\partial_x g_{f,xx} + \\ & \frac{D g_{f,xx}\partial_x D\partial_x V_f}{V_f} + \frac{D g_{f,xx}\partial_t V_f}{V_f} + \frac{D\partial_x D\partial_x V_f\partial_x^2 V_f}{2V_f^2} \\ & - \frac{D\partial_t V_f\partial_x^2 V_f}{2V_f^2} - \frac{D\partial_x D\left(\partial_x V_f\right)^3}{4V_f^3} + \frac{D\partial_t V_f\left(\partial_x V_f\right)^2}{4V_f^3} + \\ & g_{f,xx}\left(\partial_x D\right)^2 + \frac{\left(\partial_x D\right)^2\left(\partial_x V_f\right)^2}{4V_f^2} - \frac{\partial_x D\partial_t V_f\partial_x V_f}{2V_f^2} + \\ & \frac{\left(\partial_t V_f\right)^2}{4V_f^2}. \end{aligned} \tag{C1}$$

Considering the analytical closure Eq. (18) for the unclosed term $\mathbb{E}\left(\varepsilon_f\partial_x^4\varepsilon_f\right)$, the correspondence writes as

$$\begin{aligned} g_{f,tt} = {} & 3D^2 g_{f,xx}^2 - \frac{D^2 g_{f,xx}\partial_x^2 V_f}{V_f} + \frac{D^2\partial_x V_f\partial_x g_{f,xx}}{V_f} + \\ & \frac{3D^2 g_{f,xx}\left(\partial_x V_f\right)^2}{2V_f^2} + \frac{D^2\left(\partial_x^2 V_f\right)^2}{4V_f^2} - \frac{D^2\left(\partial_x V_f\right)^2\partial_x^2 V_f}{4V_f^3} + \\ & \frac{D^2\left(\partial_x V_f\right)^4}{16V_f^4} + D\partial_x D\partial_x g_{f,xx} + \frac{D g_{f,xx}\partial_x D\partial_x V_f}{V_f} + \\ & \frac{D g_{f,xx}\partial_t V_f}{V_f} + \frac{D\partial_x D\partial_x V_f\partial_x^2 V_f}{2V_f^2} \\ & - \frac{D\partial_t V_f\partial_x^2 V_f}{2V_f^2} - \frac{D\partial_x D\left(\partial_x V_f\right)^3}{4V_f^3} + \frac{D\partial_t V_f\left(\partial_x V_f\right)^2}{4V_f^3} + \\ & g_{f,xx}\left(\partial_x D\right)^2 + \frac{\left(\partial_x D\right)^2\left(\partial_x V_f\right)^2}{4V_f^2} - \frac{\partial_x D\partial_t V_f\partial_x V_f}{2V_f^2} + \frac{\left(\partial_t V_f\right)^2}{4V_f^2}. \end{aligned} \tag{C2}$$

The latter expression being quite complex, simplifications are introduced. First the variance field is assumed locally homogeneous at the boundary *i.e.* $\partial_x V_f(t, x = 0) = 0$, so that Eq. (C2) simplifies as

$$\begin{aligned} g_{f,tt} = {} & 3D^2 g_{f,xx}^2 + D\partial_x D\partial_x g_{f,xx} + \frac{D g_{f,xx}\partial_t V_f}{V_f} + \\ & g_{f,xx}\left(\partial_x D\right)^2 + \frac{\left(\partial_t V_f\right)^2}{4V_f^2}. \end{aligned} \tag{C3}$$

Then, if the variance is moreover assumed stationary, then Eq. (C3) becomes

$$g_{f,tt} = 3D^2 g_{f,xx}^2 + D\partial_x D\partial_x g_{f,xx} + g_{f,xx}\left(\partial_x D\right)^2. \tag{C4}$$

Eventually, then the diffusion coefficient field is homogeneous, then the spatio-temporal connection between the temporal metric and the spatial metric reads

$$g_{f,tt} = 3D^2 g_{f,xx}. \tag{C5}$$

While Eq. (C5) is a particular case, this equality is considered as a proxy for setting the auto-correlation time scale of the boundary perturbation even when the variance and the diffusion fields are heterogeneous.

Note that another expression for the spatio-temporal consistency Eq. (C2) can be obtained when first considering the dynamics of the variance given by Eq. (19b), leading to replace the trend of the variance by $\partial_t V_f = -2DV_f g_{f,xx} + D\partial_x^2 V_f - \frac{D(\partial_x V_f)^2}{2V_f} + \partial_x D\partial_x V_f$, so that Eq. (C2) simplifies as

$$\begin{split} g_{f,tt} = 2D^2 g_{f,xx}^2 + \frac{D^2\partial_x V_f \partial_x g_{f,xx}}{V_f} + \\ \frac{D^2 g_{f,xx}\left(\partial_x V_f\right)^2}{V_f^2} + D\partial_x D\partial_x g_{f,xx} + \\ \frac{2Dg_{f,xx}\partial_x D\partial_x V_f}{V_f} + g_{f,xx}\left(\partial_x D\right)^2, \end{split} \tag{C6}$$

from which the assumption of local homogeneity at the boundary *i.e.* $\partial_x V_f(t, x = 0) = 0$, leads to

$$g_{f,tt} = 2D^2 g_{f,xx}^2 + D\partial_x D\partial_x g_{f,xx} + g_{f,xx}\left(\partial_x D\right)^2. \tag{C7}$$

When the diffusion field is constant, then the time auto-correlation metric is related to the space auto-correlation metric by

$$g_{f,tt} = 2D^2 g_{f,xx}, \tag{C8}$$

which is a different result from Eq. (C5).

It is not clear whether the appropriate consistency should be given by Eq. (C5) or Eq. (C8) *i.e.* if it is right to replace the trend of the variance Eq. (19b) in the consistency relation Eq. (C2).

From numerical experiment, it appears that setting the time auto-correlation of boundary perturbation with Eq. (C5) in the EnKF is in agreement with the PKF results. This suggests that taking into account the trend of the variance would lead to a kind of over-specification of the boundary condition for the diffusion equation in an EnKF approach.

Appendix D Open Research

V1.0 of the *Boundary conditions for the parametric kalman filter forecast* software used to compute and analyze the numerical experiments presented in this paper is preserved at 10.5281/zenodo.7193985 and developed openly at https://github.com/opannekoucke/pkf-boundary. Sabathier et al. (2022)

Acknowledgments
The authors would like to thank François Rogier (ONERA) for fruitful discussions. M. Sabathier PhD Thesis is supported by CNES grant 51/19168. O. Pannekoucke is supported by the French national program LEFE/INSU grant "Multivariate Parametric Kalman Filter" (MPKF).

References

Bourdarie, S. A., & Maget, V. F. (2012, jun). Electron radiation belt data assimilation with an ensemble kalman filter relying on the salammbô code. *Annales Geophysicae*, *30*(6), 929–943. doi: 10.5194/angeo-30-929-2012

Cohn, S. E. (1993). Dynamics of short-term univariate forecast error covariances. *Monthly Weather Review*, *121*, 3123-3149.

Dahmen, N., Rogier, F., & Maget, V. (2020, sep). On the modelling of highly anisotropic diffusion for electron radiation belt dynamic codes. *Computer Physics Communications*, *254*, 107342. doi: 10.1016/j.cpc.2020.107342

Derber, J., Purser, R., Wu, W.-S., Treadon, R., Pondeca, M., Parrish, D., & Kleist, D. (2003, 09). Flow dependent jb in a global grid-point 3d-var..

Evensen, G. (2009). *Data assimilation: The ensemble kalman filter*. Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-03711-5

Gilpin, S., Matsuo, T., & Cohn, S. E. (2022). Continuum covariance propagation for understanding variance loss in advective systems. *SIAM/ASA Journal on Uncertainty Quantification*, *10*(3), 886-914. doi: 10.1137/21M1442449

Gustafsson, N. (2012, mar). Control of lateral boundary conditions in four-dimensional variational data assimilation for a limited area model. *Tellus A: Dynamic Meteorology and Oceanography*, *64*(1), 17518. doi: 10.3402/tellusa.v64i0.17518

Houtekamer, P., & Mitchell, H. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*, 123–137. doi: 10.1175/1520-0493(2001)129⟨0123:ASEKFF⟩2.0.CO;2

Houtekamer, P., & Zhang, F. (2016). Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*, 4489–4532. doi: 10.1175/MWR-D-15-0440.1

Maget, V., Sicard-Piet, A., Bourdarie, S., Lazaro, D., Turner, D. L., Daglis, I. A., & Sandberg, I. (2015, July). Improved outer boundary conditions for outer radiation belt data assimilation using THEMIS-SST data and the Salammbo-EnKF code. *Journal of Geophysical Research: Space Physics*, *120*(7), 2015JA021001. Retrieved 2015-09-23, from `http://onlinelibrary.wiley.com/doi/10.1002/2015JA021001/abstract` doi: 10.1002/2015JA021001

Ménard, R., Skachko, S., & Pannekoucke, O. (2021, aug). Numerical discretization causing error variance loss and the need for inflation. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.4139

Mirouze, I., & Weaver, A. T. (2010, Jul). Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, *136*(651), 1421–1443. doi: 10.1002/qj.643

Paciorek, C., & Schervish, M. (2006, 08). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*, 483-506. doi: 10.1002/env.785

Pannekoucke, O. (2021). An anisotropic formulation of the parametric kalman filter assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *73*(1), 1-27. Retrieved from `https://doi.org/10.1080/16000870.2021.1926660` doi: 10.1080/16000870.2021.1926660

Pannekoucke, O., & Arbogast, P. (2021, oct). SymPKF (v1.0): a symbolic and computational toolbox for the design of parametric kalman filter dynamics. *Geoscientific Model Development*, *14*(10), 5957–5976. doi: 10.5194/gmd-14-5957-2021

Pannekoucke, O., Berre, L., & Desroziers, G. (2008). Background error correlation length-scale estimates and their sampling statistics. *Quaterly Journal Royal Meteorological Society*, *134*, 497–508.

Pannekoucke, O., Bocquet, M., & Ménard, R. (2018). Parametric covariance dynamics for the nonlinear diffusive burgers' equation. *Nonlinear Processes in Geophysics*, *2018*, 1–21. Retrieved from `https://www.nonlin-processes-geophys-discuss.net/npg-2018-10/` doi: https://doi.org/10.5194/npg-2018-10

Pannekoucke, O., & Massart, S. (2008). Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Quarterly Journal of the Royal Meteorological Society.*, *134*, 1425–1438. doi: 10.1002/qj.288

Pannekoucke, O., Ménard, R., El Aabaribaoune, M., & Plu, M. (2021). A methodology to obtain model-error covariances due to the discretization scheme from the parametric kalman filter perspective. *Nonlinear Processes in Geophysics*, *28*(1), 1–22. Retrieved from `https://npg.copernicus.org/articles/28/1/2021/` doi: 10.5194/npg-28-1-2021

Pannekoucke, O., Ricci, S., Barthelemy, S., Ménard, R., & Thual, O. (2016). Parametric kalman filter for chemical transport models. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 31547. Retrieved from `https://doi.org/10.3402/tellusa.v68.31547` doi: 10.3402/tellusa.v68.31547

Sabathier, M., Pannekoucke, O., & Maget, V. (2022, oct). *opannekoucke/pkf-boundary: Boundary conditions for the parametric kalman filter forecast* [Software]. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7193985` doi: 10.5281/zenodo.7193985

Strassen, V. (1969, aug). Gaussian elimination is not optimal. *Numerische Mathematik*, *13*(4), 354–356. doi: 10.1007/bf02165411

Stull, R. (1988). *An introduction to boundary layer meteorology.* Kluwer Academic.

Weaver, A., & Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation (tech. memo. ecmwf, num. 306). *Quarterly Journal of the Royal Meteorological Society*, *127*, 1815–1846. doi: 10.1002/qj.49712757518