



**HAL**  
open science

# Boundary Conditions for the Parametric Kalman Filter Forecast

M. Sabathier, O. Pannekoucke, V. Maget, N. Dahmen

► **To cite this version:**

M. Sabathier, O. Pannekoucke, V. Maget, N. Dahmen. Boundary Conditions for the Parametric Kalman Filter Forecast. *Journal of Advances in Modeling Earth Systems*, 2023, 15 (10), 10.1029/2022MS003462 . hal-04273356v2

**HAL Id: hal-04273356**

**<https://hal.science/hal-04273356v2>**

Submitted on 8 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



## RESEARCH ARTICLE

10.1029/2022MS003462

# Boundary Conditions for the Parametric Kalman Filter Forecast

M. Sabathier<sup>1</sup> , O. Pannekoucke<sup>2,3,4</sup> , V. Maget<sup>1</sup> , and N. Dahmen<sup>1</sup> 
<sup>1</sup>ONERA, Toulouse, France, <sup>2</sup>CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France, <sup>3</sup>CERFACS, Toulouse, France, <sup>4</sup>INPT-ENM, Toulouse, France

## Special Section:

Data assimilation for Earth system models

## Key Points:

- The parametric Kalman filter (PKF) is studied in a 1D limited domain for Dirichlet and Neumann boundary conditions (BC)
- An auto-correlation time-scale of the BC perturbation for ensemble forecast is proposed for the anisotropy of the ensemble Kalman filter to match the PKF BCs
- Dirichlet (Neumann) BC on the physical dynamics implies Dirichlet (Neumann) BC on the variance and Dirichlet condition on the anisotropy

## Correspondence to:

V. Maget,  
vincent.maget@onera.fr

## Citation:

Sabathier, M., Pannekoucke, O., Maget, V., & Dahmen, N. (2023). Boundary conditions for the parametric Kalman filter forecast. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003462. <https://doi.org/10.1029/2022MS003462>

Received 24 OCT 2022

Accepted 27 SEP 2023

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

**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 covariances 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 model, 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 equation 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, the parametric Kalman filter, that opens the way to data assimilation in real applications where the boundaries are important, such as in radiation belts predictions, air quality, or wild-land fire; or the exploration of the coupling of the uncertainty in the atmosphere-ocean coupled system.

## 1. Introduction

Uncertainty prediction is a challenging topic, important in probabilistic forecasting as well as in data assimilation (DA) for example, in variational DA to improve the specification of the background term (Buehner, 2005; Lorenc, 2003). One of the main theoretical backbone in DA is given by the Kalman filter equations that applies for linear dynamics, but that cannot be used for large systems because 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).

The EnKF is widely used and has shown to perform well for many applications in geosciences for example, 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 induce 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 (BC) 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.

In weather forecasting over local-area domains, the importance of the boundary has been identified from sensitivity studies showing that errors may be traced back to the boundaries Errico et al. (1993). Hence, inaccurate BC may result in large forecast errors (Vukicevic & Errico, 1990; Vukicevic & Paegle, 1989; Warner et al., 1989). However, in preliminary studies on the potential of EnKF applied to local-area models (LAM), no perturbations of the BC were considered (Dowell et al., 2004; Snyder & Zhang, 2003), leading to a loss of variance as lead time increases (Nutter, Stensrud, & Xue, 2004; Nutter, Xue, & Stensrud, 2004). Note that, this loss of variance also occurs for the Kalman filter prediction step for example, in a transport equation over 1D limited-area domain, when the boundary condition is not well formalized in the dynamics, the inflow of variance is null at the inlet of the domain (see Appendix A). This issue can be solved from the formalism of the Kalman filter as used in optimal control, by considering that the boundary is an uncertain command as done in for example, heat conduction estimation (Scarpa & Milano, 1995). In the community of the EnKF for LAMs, the loss of variance is avoided by perturbing the BC (Nutter, Xue, & Stensrud, 2004), for example, by modeling the spatial and temporal covariance relationship of the boundary condition based on the use of multivariate covariance model or by using BC which derives from an ensemble on a larger domain (Torn et al., 2006), even if for the latter approach, the consistency across multiple domains is difficult to handle (Houtekamer & Zhang, 2016, Section 6.a). Note that in variational DA for LAMs, it is possible to include the uncertainty on the boundary 4DVar through the adjunction of a cost function term that constraints the increment or its tendency at the boundary (Gustafsson, 2012), which is a way to avoid the loss of variance above mentioned for the EnKF and for the KF.

In radiation belts where there is no larger model and thus no ensemble from which perturbation on the boundary can be introduced, we are interested by sampling boundary perturbations from a spatio-temporal covariance model. However, because the solar activity is non-stationary, we need to develop perturbations of the BC that are heterogeneous and non-stationary that is, with varying time auto-correlation, in place to the auto-regressive model often used but for which the time auto-correlation is constant (Torn et al., 2006). Moreover, we want to reduce the numerical cost of using an EnKF while keeping the uncertainty dynamics it provides. This motivation can also be encountered in the development of the assimilation scheme for operational model in air quality, that deals with hundreds of chemical species, and for which the numerical cost of an integration is large. For instance, in the operational air quality model MOCAGE developed at Météo-France, the assimilation relies on a 3D-FGAT approach (Amraoui et al., 2020; Massart et al., 2010), without ensemble, and it would be interesting to introduce flow-dependency of the background term while accounting for the transport of the boundary uncertainty during the assimilation at regional scale.

Recently another approximation for the KF, different from the EnKF, has been introduced, 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 times the integration of the transport (Pannekoucke, 2021a; Pannekoucke et al., 2018). In addition, the PKF provides an understanding of the uncertainties dynamics, which is less obvious to determine from an ensemble estimation alone. For instance, when considering a transport, the PKF equations for the error statistics consist in: the advection of the error variance without source term, meaning that the magnitude of the error variance is conserved; and the advection of the anisotropy plus a source term that corresponds to the deformation of the anisotropy due to the local shear (Pannekoucke, 2021a). As another example, when considering now a conservative dynamics or a non-linear Burgers dynamics (in the tangent-linear approximation for the error statistics), a source term appears in the PKF equation for the error variance which is related to high gradient area that can be interpreted as a production term similar to the one encountered in turbulence (Pannekoucke et al., 2018). These examples of physical interpretations of the dynamics of the error statistics, as it is provided by the PKF equations, are less easy to deduce from the ensemble estimation while the effects of these processes can be observed. Note that data-driven modeling leveraged on deep-learning, especially physics-informed neural network (Raissi et al., 2019), can be considered to determine the dynamics of the error statistics (Pannekoucke & Fablet, 2020). On the other hand, it is not necessarily easy to interpret all terms in the PKF equations for example, when considering a diffusion equation, for which the PKF shows a coupling between the error variance and the anisotropy, it is hard to physically explain terms in square of the gradient of the anisotropy. In this way, the PKF can contribute to improve the understanding of other topics of interest in DA for example, in the characterization of the dynamics of the model-error covariance due to the spatial and temporal discretization of a transport

equation (Ménard et al., 2021; Pannekoucke et al., 2021). The PKF is interesting to provide a flow dependency of the background error covariance matrix in variational DA for example, by providing a proxy for the diffusion tensor of the day in the covariance model based on the diffusion equation (Pannekoucke et al., 2016); or by providing the variance of the day in presence of model error thanks to the PKF dynamics of the variance, as in the assimilation of GOSAT methane observations using the hemispherical community multiscale air quality model (Voshtani et al., 2022). While the PKF has been mainly developed in univariate statistics, an exploration has begun to develop a multivariate statistics formulation (Perrot et al., 2023). At a more theoretical level, the parametric formulation has been considered to describe the covariance dynamics of white noise error in a conservative equation (Gilpin et al., 2022). Hence, the PKF approach seems promising to tackle the issues encountered in the development of the DA for radiation belts and air quality.

Until now, the PKF has been explored mainly on periodic 1D or 2D domains, where it has been shown to reproduce interesting features of the uncertainty dynamics in linear problem for example, for the transport (Pannekoucke, 2021a), as well as for non-linear dynamics at the second order for example, for the non-linear advection-diffusion equation (Pannekoucke et al., 2018). Note that the non-linear extension of the PKF relies on a tangent-linear evolution of error, with a feedback of the uncertainty on the dynamics of the mean through the error-variance (i.e., a fluctuation-mean interaction), leading to a Gaussian second-order filter (Jazwinski, 1970, Section 9.3). However, to go ahead toward real applications, and especially applications in bounded domains, appropriate specification of BC of the error statistics is needed for the PKF dynamics. To do so, we propose to explore the specification of the BC for the PKF when Dirichlet and Neumann conditions are considered in the physical dynamics. This exploration focuses on two dynamics of interest for our applications: the transport equation for example, for air quality or weather prediction; and the diffusion equation for example, 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 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 BC for the PKF needs an ensemble of forecasts. To do so, an intermediate Section 4 will detail how to specify BC in an EnKF experiment that produces desired error statistics. This is an important contribution of the paper in order to validate the specification of the BC 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 contained introduction to the PKF, 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 presented, 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  $\Omega$  of dimension  $d$  and coordinate system  $\mathbf{x} = (x^i)_{i \in [1,d]}$ , whose dynamics is given by

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

where  $\mathcal{M}$  stands for a function of the state  $\chi$  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, with the convention that order zero denotes the field  $\chi$  itself. Thereafter, for the sake of simplicity,  $\mathcal{M}$  is assumed linear but the formalism extends to the non-linear framework (Pannekoucke & Arbogast, 2021; Pannekoucke et al., 2018). Note that  $\chi$  can be either continuous or discrete (the discretized version of the continuous field): the discrete case leads to matrix algebra relations for example,  $\mathcal{M}$  is replaced by its matrix formulation  $\mathbf{M}$ .

Hence, for a time integration over a time window  $[0, T]$ , the dynamics Equation 1 is written as

$$\chi(T) = \mathbf{M}_{T \rightarrow 0} \chi(0), \quad (2)$$

where  $\mathbf{M}_{T \leftarrow 0}$  denotes the propagator associated with the time integration of Equation 1 over  $[0, T]$ . Compared to the classical formulation of the Kalman filter equations, we do not consider here the model error because we won't try to characterize it here, while it can lead to spurious time evolution of ensemble error statistics compared to the true dynamics, as it has been shown using the PKF in previous contributions (Ménard et al., 2021; Pannekoucke, 2021a; Pannekoucke et al., 2021; Perrot et al., 2023).

In real applications, the spatio-temporal heterogeneity of the observation network, as well as the model error, imply that  $\chi$  is not known exactly. The true state of the system is denoted by  $\chi^t$ . The analysis state, that is the estimation of the true state knowing the observations until a given time, is denoted by  $\chi^a$ . 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  $\mathbf{P}^a = \mathbb{E}[e^a(e^a)^T]$ , where  $\mathbb{E}[\cdot]$  stands for the expectation operator and where the upper script  $(\cdot)^T$  stands for the transpose operator (later the adjoint operator for matrices). Hence, the distribution of the initial condition is a Gaussian of mean  $\chi^a$  and covariance matrix  $\mathbf{P}^a$ , denoted by  $\mathcal{N}(\chi^a, \mathbf{P}^a)$ . 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  $T$ . 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  $\mathbf{P}^f(T) = \mathbb{E}[e^f(e^f)^T](T)$ , whose dynamics is written as

$$\partial_t e^f = \mathcal{M}(\partial e^f). \quad (3)$$

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

$$\mathbf{P}^f(T) = \mathbf{M}_{T \leftarrow 0}\mathbf{P}^a(\mathbf{M}_{T \leftarrow 0})^T. \quad (4)$$

Equation 4 corresponds to the Kalman filter propagator of the error covariance matrix, whose particular dynamics is given by

$$\frac{d\mathbf{P}^f}{dt} = \mathbf{M}\mathbf{P}^f + \mathbf{P}^f\mathbf{M}^T, \quad (5)$$

integrated over the period  $[0, T]$ , starting from the initial condition  $\mathbf{P}^f(0) = \mathbf{P}^a$ . Hence, the distribution of the forecast at  $T$  is a Gaussian of mean  $\chi^f(T)$  and covariance matrix  $\mathbf{P}^f(T)$ , that is  $\mathcal{N}(\chi^f, \mathbf{P}^f(T))$ .

When observations,  $y^o$ , are available at  $T$ , so that  $y^o = \mathbf{H}\chi^t(T) + e^o$  where  $\mathbf{H}$  denotes the linear observation operator that maps a state (here the true state  $\chi^t(T)$ ) into the observation space, and  $e^o$  denotes the observational error, modeled as a Gaussian random vector of zero mean and covariance matrix  $\mathbf{R}$  and assumed uncorrelated from the forecast error  $e^f(T)$ ; then the Kalman analysis equations at time  $T$  are written as

$$\chi^a(T) = \chi^f(T) + \mathbf{K}(y^o - \mathbf{H}\chi^f(T)), \quad (6a)$$

$$\mathbf{P}^a(T) = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f(T), \quad (6b)$$

where  $\mathbf{K} = \mathbf{P}^f(T)\mathbf{H}^T(\mathbf{H}\mathbf{P}^f(T)\mathbf{H}^T + \mathbf{R})^{-1}$  is the Kalman gain matrix. Equations 6a and 6b characterizes the Gaussian distribution of the analysis uncertainty, where Equation 6a is the equation of the mean, and Equation 6b is the equation of the analysis-error covariance matrix at time  $T$ . Equation 4 together with Equations 6a and 6b are the Kalman filter equations, that can be cycled providing the analysis and forecast assimilation cycles, where Equation 4 (Equations 6a and 6b) is the KF forecast (analysis) step.

Note that this contribution being focused on the uncertainties dynamics, with the corresponding Kalman filter forecast of the covariance matrix Equation 4, the uncertainty related to the observations (present when considering cycled assimilations) appears only through the statistics of the analysis error, and more precisely through the covariance matrix of the analysis error  $\mathbf{P}^a = \mathbf{P}^f(0)$ . Moreover, since we consider the model as perfect, the uncertainty considered here only comes from the initial condition and the BC.

Note also that the derivation of the Kalman filter forecast step Equation 4 follows the usual one encountered in the weather community, and can lead to drawback when applied to LAMs without care for example, a loss of variance at the inlet of the domain. We detail this issue considering a transport in a 1D limited-area domain in Appendix A. However, it results that when BC are uncertain, it is necessary to reformulate the problem. That can be done in the KF framework, by considering that the BCs are an uncertain command following the terminology used in optimal control (see e.g., Scarpa & Milano, 1995).

## 2.2. Ensemble Approximation of the Kalman Filter Forecast Step

While the KF forecast step Equation 4 is a simple algebraic formula, it fails to apply in large systems because of its numerical cost: if  $n$  denotes the dimension of the vector representation of  $\chi$ , then the computational complexity of Equation 4 scales between  $n^2$  and  $n^3$  (Strassen, 1969). In terms of integration cost, the KF requires  $2n$  integrations of the model Equation 1.

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

$$\widehat{\mathbf{P}}^f(t) = \frac{1}{N} \sum_k e_k e_k^T, \quad (7)$$

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 Equation 1.

Note that the normalization by  $N$  in Equation 7 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  $\frac{1}{N} \sum_k(\dots)$  as in Equation 7.

In the EnKF forecast step applied for LAMs, the uncertainty on the boundary condition can be introduced by adding perturbations at the boundary for each member of the ensemble. This can be done by considering a modeling of the multivariate covariance at the boundary or by considering boundary condition derived from an ensemble on a larger domain (Torn et al., 2006).

While this contribution focuses on the forecast step, we can mention that if ensemble computation of the KF analysis step Equations 6a and 6b can be implemented following different algorithm for example, the EnKF version based on perturbation of observation introduced by Burgers et al. (1998), the square root filter of Tippett et al. (2003) or the ensemble transform Kalman filter (ETKF) of Bishop et al. (2001) and its variational implementation (Bocquet, 2011; Harlim & Hunt, 2007); the KF forecast step is mainly the same whatever the variant of the EnKF considered. Hence, for this study dedicated to the exploration of the specification of the BC for the PKF forecast step, the particular implementation the EnKF has not impact and no assimilation experiment is conducted here.

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

## 2.3. 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 PKF dynamics aims to mimic the dynamics of Equation 5 relying on the dynamics of the parameters  $\mathcal{P}^f$ ,

$$\frac{d\mathcal{P}^f}{dt} = \mathcal{G}(\mathcal{P}^f), \quad (8)$$

where  $\mathcal{G}$  has to be determined from the particular dynamics of Equation 1, so that at any time  $t$ ,  $\mathbf{P}(\mathcal{P}^f(t))$  approximates  $\mathbf{P}^f(t)$  that is,  $\mathbf{P}(\mathcal{P}^f(t)) \approx \mathbf{P}^f(t)$ . As for the EnKF, the numerical complexity of the PKF prediction Equation 8 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 Equation 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 is written as  $\mathbf{P}(V, \mathbf{g})$  or  $\mathbf{P}(V, \mathbf{s})$ . For an error field  $e$ , the variance field is defined as

$$V = \mathbb{E}[e^2], \quad (9)$$

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  $\rho(\mathbf{x}, \mathbf{y}) = \mathbb{E}[\varepsilon(\mathbf{x})\varepsilon(\mathbf{y})]$  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  $\mathbf{g}_x$ ) which appears in the second-order Taylor's expansion

$$\rho(\mathbf{x}, \mathbf{x} + \delta\mathbf{x}) \approx 1 - \frac{1}{2} \|\delta\mathbf{x}\|_{\mathbf{g}_x}^2, \quad (10)$$

where  $\|\delta\mathbf{x}\|_{\mathbf{g}_x}^2 = \delta\mathbf{x}^T \mathbf{g}_x \delta\mathbf{x}$  denotes the norm associated with the metric tensor  $\mathbf{g}_x$  that is the symmetric definite positive matrix  $[\mathbf{g}_x]_{ij} = -\partial_{x^i x^j}^2 \rho_x$  where  $\rho_x(\delta\mathbf{x})$  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)

$$\mathbf{s}(\mathbf{x}) = (\mathbf{g}(\mathbf{x}))^{-1}, \quad (11)$$

where the superscript  $(\cdot)^{-1}$  denotes the matrix inverse, and whose the geometry goes as the correlation. In a 1D domain of coordinate  $x$ , the aspect tensor field is the scalar field  $\mathbf{s} = [s_{xx}]$ . Note that, in a 1D domain, the square root of  $s_{xx}$  is homogeneous to a length, leading to the so-called length scale  $l(x) = \sqrt{s_{xx}(x)}$ , which is often introduced in diagnoses.

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

$$[\mathbf{g}_x]_{ij} = \mathbb{E}[\partial_{x^i} \varepsilon \partial_{x^j} \varepsilon]. \quad (12)$$

Hence, the variance Equation 9 and the anisotropy Equation 12 can be computed from an ensemble estimation: the variance field is estimated by

$$\hat{V} = \frac{1}{N} \sum_k (e_k(t))^2, \quad (13)$$

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

$$\hat{g}_{ij} = \frac{1}{N} \sum_k \partial_{x^i} \varepsilon_k \partial_{x^j} \varepsilon_k, \quad (14)$$

for  $i \leq j$  (since  $g_{ji} = g_{ij}$ ). Note that in a 1D domain, the ensemble estimation of the length-scale field is computed as  $\hat{l} = \frac{1}{\sqrt{\hat{g}_{xx}}}$ . While the PKF approach does not relies on any ensembles, the ensemble estimations Equations 13 and 14 can be used to set the initial conditions for the parameters to initiate 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)

$$\mathbf{P}(V, \mathbf{s})(\mathbf{x}, \mathbf{y}) = \sqrt{V_x V_y} \frac{|\mathbf{s}_x|^{1/4} |\mathbf{s}_y|^{1/4}}{|\frac{1}{2}(\mathbf{s}_x + \mathbf{s}_y)|^{1/2}} \exp\left(-\frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_{[\frac{1}{2}(\mathbf{s}_x + \mathbf{s}_y)]^{-1}}^2\right) \quad (15)$$

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

When VLATcov models are used for the parametric approach, the dynamics of the parameters Equation 8 is deduced from the time derivative of Equations 9 and 12, and the dynamics of the error Equation 3. Hence, in

univariate statistics, the PKF equations for VLATcov models consists in three equations. The first equation corresponds to the prediction of the mean, that is Equation 1 in the linear case. The second equation is the dynamics of the variance,  $V$ , which is deduced from

$$\partial_t V = 2\mathbb{E}[e\partial_t e], \quad (16a)$$

where replacing the trend of the error Equation 3, will leads to the dynamics of  $V$

$$\partial_t V = 2\mathbb{E}[e\mathcal{M}(\partial e)]. \quad (16b)$$

The third equation is the dynamics of the anisotropy,  $g$ , whose components evolve as

$$\partial_t g_{ij} = \mathbb{E}[\partial_t(\partial_{x^i}\varepsilon\partial_{x^j}\varepsilon)], \quad (17)$$

and from which we can deduce the dynamics of the aspect tensor  $\mathbf{s} = \mathbf{g}^{-1}$  with  $\partial_t \mathbf{s} = -\mathbf{s}(\partial_t \mathbf{g})\mathbf{s}$ . The dynamics of the variance Equation 16b and of the anisotropy Equation 17 can be simplified for example, by considering the commutation between the expectation and partial derivatives (Pannekoucke & Arbogast, 2021).

Note that for non-linear dynamics, the PKF forecast step for VLATcov models stands as (see Pannekoucke & Arbogast, 2021 for details)

$$\partial_t \mathbb{E}[\chi] = \mathcal{M}(\partial \mathbb{E}[\chi]) + \mathcal{M}''(\partial \mathbb{E}[\chi])(\mathbb{E}[\partial e \otimes \partial e]), \quad (18a)$$

$$\partial_t V = 2\mathbb{E}[e\partial_t e], \quad (18b)$$

$$\partial_t g_{ij} = \mathbb{E}[\partial_t(\partial_{x^i}\varepsilon\partial_{x^j}\varepsilon)], \quad (18c)$$

where  $\mathcal{M}''$  is a linear operator that refers to the Hessian that is computed with respect to the mean state  $\mathbb{E}[\chi]$ ;  $\partial e \otimes \partial e$  denotes the tensor product of the partial derivatives with respect to the spatial coordinates, that is, terms such as  $\partial^k e \partial^m e$  for any positive integers ( $k, m$ ); and where the trends that appear in the right-hand side of Equations 18b and 18c should be formulated from the dynamics for example, using  $\partial_t e = \mathcal{M}'(\partial \mathbb{E}[\chi])(\partial e)$  for the trend of  $e$  where  $\mathcal{M}'$  is a linear operator that refers to the tangent-linear model that is computed with respect to the mean state  $\mathbb{E}[\chi]$ .

The computation of dynamical equations (Equations 18a–18c) for the mean, the variance  $V$  and the anisotropy  $\mathbf{g}$  (or  $\mathbf{s}$ ) can be performed using a computed algebra system. To do so, the open source Python toolbox SympKF has been introduced (Pannekoucke, 2021b; 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 Equation 1 following a splitting strategy (Pannekoucke & Arbogast, 2021; Pannekoucke et al., 2018). Thereafter, the dynamics of the VLATcov parameters is computed by using SympKF and the interested reader is referred to the Jupyter notebooks that are provided as a supplementary material to this contribution (Sabathier et al., 2022).

Compared to the ensemble implementation of the KF (e.g., the EnKF or the ETKF), the contribution of the PKF is that it replaces the computational cost for the ensemble prediction step by the numerical cost of the time integration of the system Equations 18a–18c. More precisely, in univariate statistics, 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  $n$ ). This numerical cost of the PKF is then competitive to the ensemble method where it often needs dozen of members. However in multivariate statistics, the numerical cost of the PKF scales as the square of the number of prognostic fields, which is a strong limitation. For example, a multivariate assimilation in air quality should consider hundreds of chemical species (Perrot et al., 2023), but in practice, only a few species are assimilated, without correction for other unobserved chemical species (the forecast error of the observed and the unobserved chemical species are then assumed as decorrelated), which makes the PKF interesting for these applications (in CAMS regional air quality production 2.40 (CAMS, 2022), the univariate 3DVar system of MOCAGE is used for the separated assimilation of ozone, nitrogen dioxide, sulfur dioxide, and fine particulate matter PM2.5 and PM10, following a configuration similar to the one used for the monitoring atmospheric composition and climate: interim implementation forecast system detailed by Marécal et al. (2015)).



While it is not the purpose of the present work, we can mention that the analysis step in the PKF consist in a sequential update of the parameters during the assimilation of observations that is, the update of the variance and of the anisotropy for the VLATcov model after the assimilation of each observation. For instance, at the leading order, the update of the aspect tensor field due to the assimilation of an in situ observation is written as

$$\mathbf{s}^a = \frac{V^a}{V^f} \mathbf{s}^f, \quad (19)$$

where  $V^a$  is the analysis-error variance field that results from the assimilation of the observation (see update Equation 3 for the variance and Equation 12 for the aspect tensor in Pannekoucke (2021a) for details). In the PKF, a sequential processing of batches of observation can be consider to parallelize the assimilation, similarly to the procedure employed in the EnKF analysis step (Houtekamer & Mitchell, 2001).

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 Equation 1.

## 2.4. Illustration of the PKF Forecast Step 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 BC. The PKF for the transport has been already detailed in previous contributions (see e.g., Pannekoucke & Arbogast, 2021), and is recapped here for self consistency. The PKF for the heterogeneous diffusion equation is original, and extends the derivation of the PKF applied to the homogeneous diffusion equation (Pannekoucke et al., 2018).

### 2.4.1. PKF Prediction Applied on a Transport Equation

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

$$\partial_t c + u \partial_x c = 0. \quad (20)$$

In this example, and by identification with Equation 1,  $c$  stands for  $\chi$  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 Equation 20 using SymPKF leads to the system

$$\partial_t c = -u \partial_x c, \quad (21a)$$

$$\partial_t V_c = -u \partial_x V_c, \quad (21b)$$

$$\partial_t s_{c,xx} = -u \partial_x s_{c,xx} + 2s_{c,xx} \partial_x u, \quad (21c)$$

where the anisotropy is represented by the aspect tensor  $\mathbf{s} = s_{c,xx}$  in 1D domain. The PKF dynamics Equations 21a–21c 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  $\mathbb{E}[c]$ , Equation 21a, where the expectation operator has been removed for the sake of simplicity; the transport of the variance, Equation 21b; and the transport of the anisotropy Equation 21c, where an additional source term of anisotropy appears, that is due to the shear by the flow. Compared with an ensemble approach, the PKF approach allows for an understanding of the dynamics and the physics of the uncertainty.

This example shows that the PKF does not need solving the full state dimension covariance matrix, but only the dynamics of the parameters used in the modeling of the covariance matrix: the variance and the anisotropy in the VLATcov model considered here which evolves from Equations 21b and 21c respectively. However, the parallelization of the PKF is less obvious than for the parallel computation of an ensemble of forecast, since it requires to parallelize the coupled system Equations 21a–21c.

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 Equation 20 at a symbolic level. This labeling for the parameters has been introduced when multiple fields are present for example, 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.

#### 2.4.2. PKF Prediction Applied on a Diffusion Equation

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

$$\partial_t f = \partial_x(D\partial_x f), \quad (22)$$

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 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 Equation 22 stands in this case for a combined spatial and physical quantities for example, 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 Equation 1,  $f$  stands for  $\chi$  while  $\mathcal{M}(f, \partial f) = \partial_x(D\partial_x f)$ .

The computation of the PKF dynamics for Equation 22 can be performed using SymPKF. However, because of the second order derivative, the dynamical system makes appear an unknown term  $\mathbb{E}[\varepsilon_f \partial_x^4 \varepsilon_f]$ , not determined from  $f$ ,  $V_f$  and  $s_{f,xx}$  (see Appendix B). An analytical closure has been proposed for 1D domains which states as (Pannekoucke et al., 2018)

$$\mathbb{E}[\varepsilon_f \partial_x^4 \varepsilon_f] = 3g_{f,xx}^2 - 2\partial_x^2 g_{f,xx} \quad (23a)$$

when written in metric tensor or

$$\mathbb{E}[\varepsilon_f \partial_x^4 \varepsilon_f] = \frac{2\partial_x^2 s_{f,xx}}{s_{f,xx}^2} + \frac{3}{s_{f,xx}^2} - \frac{4(\partial_x s_{f,xx})^2}{s_{f,xx}^3} \quad (23b)$$

in aspect tensor, which leads to the PKF dynamics

$$\partial_t f = D\partial_x^2 f + \partial_x D\partial_x f, \quad (24a)$$

$$\partial_t V_f = -\frac{2DV_f}{s_{f,xx}} + D\partial_x^2 V_f - \frac{D(\partial_x V_f)^2}{2V_f} + \partial_x D\partial_x V_f, \quad (24b)$$

$$\begin{aligned} \partial_t s_{f,xx} = & D\partial_x^2 s_{f,xx} + 4D \\ & - \frac{2D(\partial_x s_{f,xx})^2}{s_{f,xx}} - \frac{2Ds_{f,xx}\partial_x^2 V_f}{V_f} \\ & + \frac{D\partial_x V_f \partial_x s_{f,xx}}{V_f} + \frac{2Ds_{f,xx}(\partial_x V_f)^2}{V_f^2} - 2s_{f,xx}\partial_x^2 D \\ & + 2\partial_x D\partial_x s_{f,xx} - \frac{2s_{f,xx}\partial_x D\partial_x V_f}{V_f}, \end{aligned} \quad (24c)$$

where in this dynamical system, the expected value  $\mathbb{E}[f]$  in Equation 24a is replaced by  $f$  for the sake of simplicity. The dynamics Equations 24a–24c show 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 Equations 21a–21c as discussed for Equation 20 in the particular case where  $u = -\partial_x D$ . The other terms in Equations 24a–24c 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 Equations 24a–24c read as

$$\partial_t f = D\partial_x^2 f + \partial_x D\partial_x f, \quad (25a)$$

$$\partial_t V_f = -2DV_f g_{f,xx} + D\partial_x^2 V_f - \frac{D(\partial_x V_f)^2}{2V_f} + \partial_x D\partial_x V_f, \quad (25b)$$

$$\begin{aligned} \partial_t g_{f,xx} = & -4Dg_{f,xx}^2 + D\partial_x^2 g_{f,xx} \\ & + \frac{2Dg_{f,xx}\partial_x^2 V_f}{V_f} + \frac{D\partial_x V_f \partial_x g_{f,xx}}{V_f} - \frac{2Dg_{f,xx}(\partial_x V_f)^2}{V_f^2} \\ & + 2g_{f,xx}\partial_x^2 D + 2\partial_x D\partial_x g_{f,xx} + \frac{2g_{f,xx}\partial_x D\partial_x V_f}{V_f}. \end{aligned} \quad (25c)$$

Compared with the use of an ensemble estimation for the error statistics, the dynamics Equations 25a–25c appear rather complicated, especially the dynamics of the metric Equation 25c. However, this coupled system is simple to solve numerically for example, considering a finite difference scheme and automatic code generation. Moreover, the numerical cost of Equations 25a–25c is of the order of three times the cost the original heterogeneous diffusion Equation 22 while it provides an accurate estimation of the error statistics (see numerical experiment in Section 5.3), where a large ensemble is needed to reach this accuracy and avoid the sampling noise whose magnitude, for an ensemble of  $N_e$  members, scales in  $1/\sqrt{N_e}$  from the central limit theorem (CLT).

Until now, PKF dynamics for the heterogeneous diffusion equation has been evaluated on periodic domain only, while bounded domains are often needed, for example, 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 BC for the PKF dynamics.

### 3. Specification of the Boundary Conditions for the PKF Forecast Step

This section tackles the specification of the BC 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 BC of an arbitrary domain  $\Omega$  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 BC for the variance and the anisotropy.

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

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  $\Omega$  where this time, the BC 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 for example, for VLATcov models, the variance and the anisotropy can be estimated from the ensemble of large scale forecasts from Equations 13 and 14 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 are written as null fluxes that is,  $\partial_x \chi(t, x = 0) = 0$ . This implies that the error field must also verifies a Neumann condition that is,  $\partial_x e(t, x = 0) = 0$ . Again, we are looking for the BC 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  $V(t, x) = \mathbb{E}[e^2](t, x)$ ,

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

As the local Taylor expansion of the variance field at  $x = 0$  has no first-order term in  $\delta x$ , this implies that the first order derivative is null, that is,

$$\partial_x V(t, 0) = 0, \tag{26a}$$

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, Equation 26a, implies that the metric tensor  $g(t, x) = \mathbb{E}[(\partial_x \varepsilon)^2](t, x)$  simplifies as  $g(t, 0) = \frac{1}{V(t, 0)} \mathbb{E}[(\partial_x e(t, 0))^2]$ . Then the Neumann condition on  $e$ ,  $\partial_x e(t, 0) = 0$  that is, implies that the condition for the metric is a Dirichlet condition,

$$g(t, x = 0) = 0. \tag{26b}$$

Note that the Dirichlet condition for the metric Equation 26b is equivalent to an infinite in aspect tensor at the boundary that is,  $s(t, 0) = +\infty$ . Therefore, the PKF formulated in aspect tensor is singular near the boundaries where the error field verifies a Neumann condition. Hence, for a Neumann condition on the error, it is preferable to consider the PKF formulation in variance/metric tensor rather than in variance/aspect tensor (while there is no preference on unbounded domains e.g., in periodic domains or flat spaces).

Hence, the Neumann condition on the error for a 1D domain leads, for the PKF dynamics, to a Neumann condition in variance and a Dirichlet condition in metric. This extends to a 2D or 3D domain  $\Omega$  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 BC 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. Methodology for Validating the PKF From an EnKF Simulations, and Specification of the BCs for EnKF

### 4.1. Methodology for the Validation of the PKF From an Ensemble Method

In the previous sections we specified the PKF equations as well as the BC it needs depending on the boundary condition for the error at the boundary. Note that for linear dynamics as for the transport and the diffusion that are considered here, the dynamics of the error statistics is independent from the dynamics of the mean, so hereafter we will not be considering the dynamics of the mean but rather focusing on the error statistics. For instance, the PKF dynamics for the error statistics for the transport with a positive wind on a bounded 1D domain and Dirichlet condition in  $x = 0$ , Equations 21b and 21c are written as

$$\partial_t V_c = -u \partial_x V_c, \tag{27a}$$

$$V_c(0, x) = V_c^0(x), \tag{27b}$$

$$V_c(t, 0) = b_{V_c}(t), \tag{27c}$$

$$\partial_t s_{c,xx} = -u \partial_x s_{c,xx} + 2s_{c,xx} \partial_x u, \tag{27d}$$

$$s_{c,xx}(0, x) = s_{c,xx}^0(x), \quad (27e)$$

$$s_{c,xx}(t, 0) = b_{s_{c,xx}}(t), \quad (27f)$$

where  $b_{V_c}$  and  $b_{s_{c,xx}}$  stands respectively for the error variance and anisotropy time functions prescribed at the inlet of the domain in  $x = 0$ , and where  $V_c^0$  and  $s_{c,xx}^0$  are the initial condition fields.

Now we are interested to verify that the resulting PKF dynamics, such as Equations 27a–27f, is really able to reproduce the dynamics of the error statistics as an ensemble method would provide. But then, it is necessary to create an ensemble of forecast with an appropriate specification of the boundary error. Again, for the transport with Dirichlet condition, it remains to perform an ensemble of integration of the dynamics

$$\partial_t e_c = -u \partial_x e_c, \quad (28a)$$

$$e_c(0, x) = e_c^0(x), \quad (28b)$$

$$e_c(t, 0) = b_{e_c}(t), \quad (28c)$$

where  $b_{e_c}$  and  $e_c^0$  stands respectively for the error time function prescribed at the inlet of the domain in  $x = 0$ , and the initial error field.

In real application where the inflow at the inlet is related to physical structures for example, a mid-latitude vortex entering in a LAM, the error is continuous and differentiable in time and space meaning this imposes a constraint of coherence between the error field inside the domain and its boundary condition at the inlet. Then, the resulting error statistics should also be continuous and differentiable in time and space: this imposes a constraint of coherence between the variance (or the anisotropy) field inside the domain and its boundary condition at the inlet.

Note that when a covariance model is used in EnKF to perturb BC, an assumption of stationarity of the error statistics is often encountered, for example, by considering a constant time correlation in an auto-regressive time process (Torn et al., 2006) (similar to the one introduced in Appendix A). However for physical phenomenons entering in a LAM, the correlation time scale is expected to evolve accordingly to the spatial anisotropy, because the spatial anisotropy is governed by the dynamics and the observational network (Bouttier, 1994; Cohn, 1993; Pannekoucke, 2021a). It would therefore be interesting to illustrate how the PKF performs for heterogeneous and non-stationary error statistics at the frontier.

Thus, we consider the following validation framework where the error statistics are given by a prescribed initial condition together with its BC which leads to smooth evolution of the error statistics, without any constraint of homogeneity and stationarity of the statistics. For instance, when considering the transport, this means that we consider the situation where the error statistics are given by a prescribed initial condition ( $V_c^0(x), s_{c,xx}^0(x)$ ) together with its BC ( $b_{V_c}(t), b_{s_{c,xx}}(t)$ ) which leads to smooth fields ( $V_c(t, x), s_{c,xx}(t, x)$ ). But, for the ensemble use for the validation, this implies to populate an ensemble of  $N_e$  initial conditions ( $e_{c,k}(0, x)_{k \in [1, N_e]}$ ) together with their BC ( $b_{e_c,k}(t)_{k \in [1, N_e]}$ ) such that at a given time  $t$  and position  $x$ , the resulting error variance (anisotropy) is  $V_c(t, x)$  ( $s_{c,xx}(t, x)$ ). The problem that arises is how to specify the auto-correlation time scale of the boundary perturbation in order to obtain the desired error statistics ( $V_c(t, x), s_{c,xx}(t, x)$ ).

Note that ensemble forecasting for the diffusion equation with Neumann BC corresponds to integrating an ensemble of initial conditions that verifies the Neumann conditions at the boundaries, and without additional perturbation of the error field at the boundaries along the time integration (the metric at the boundary is constant equal to zero). It results that the specification of the auto-correlation time scale to build perturbations of the error field at the boundaries only concerns the 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.

#### 4.2. Specification of the Auto-Correlation Time-Scale of the BC Perturbations for Ensemble of Forecast

The problem faced here is that, with BC being time-series, it has an associated 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 Equation 12, 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  $V_\eta(t) = \mathbb{E}[\eta(t)^2]$ , is defined by

$$\mathbf{g}_{tt}(t) = \mathbb{E} \left[ \partial_t \left( \frac{\eta(t)}{V_\eta(t)} \right) \partial_t \left( \frac{\eta(t)}{V_\eta(t)} \right) \right]. \quad (29)$$

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

$$\mathbf{g}_{tt,x=0}(t) = \mathbb{E}[\partial_t \varepsilon(t, \mathbf{x} = 0) \partial_t \varepsilon(t, \mathbf{x} = 0)], \quad (30a)$$

where  $\varepsilon = e/\sqrt{V}$  is the normalized error associated with the spatial error  $e$ . While Equation 30a 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 Equation 3 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 C)

$$g_{tt} \underset{x=0}{=} \frac{1}{V} \mathbb{E} \left[ \left( \mathcal{M}(\varepsilon \sqrt{V}, \partial(\varepsilon \sqrt{V})) \right)^2 \right] - \frac{1}{4V^2} (\partial_t V)^2, \quad (30b)$$

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 30b may be complicated, and can contain unknown terms such as  $\mathbb{E}[\varepsilon_f \partial_x^4 \varepsilon_f]$  encountered for the heterogeneous diffusion dynamics in Section 2.4.2. The next two sub-sections will detail the link between the temporal and the spatial metrics for the transport and for the diffusion.

### 4.3. 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 20 is now considered.

Following the theoretical derivation of the temporal metric Equation 30b, SymPKF is used to derive the relation between the temporal anisotropy tensor of the error and the spatial anisotropy tensor of the error which yields

$$g_{c,tt} \underset{x=0}{=} u^2 g_{c,xx} + \frac{u^2 (\partial_x V_c)^2}{4V_c^2} + \frac{u \partial_t V_c \partial_x V_c}{2V_c^2} + \frac{(\partial_t V_c)^2}{4V_c^2}. \quad (31)$$

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 ( $\partial_t V_c = 0$ ), Equation 31 reads as

$$g_{c,tt} \underset{x=0}{=} u^2 g_{c,xx}, \quad (32)$$

which is physically interpretable since Equation 32, 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 Equation 32 as an approximation of the true time-scale even when assumptions leading to Equation 32 are not verified.

Note that Equation 32 can be obtained when considering that the dynamics of the variance Equation 21b applies at the boundary, leading to replace the trend of the variance by  $\partial_t V_c = -u \partial_x V_c$  in Equation 31 so to obtain Equation 32.

To conclude this paragraph, the ensemble forecasting under Dirichlet BC 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 Equation 32.



We proceed in the same way for the diffusion equation.

#### 4.4. Dirichlet BCs for Ensemble Forecasting of the Diffusion Equation

To continue going toward more and more realistic modeling, the heterogeneous diffusion Equation 22 is now considered to compute the spatio-temporal link Equation 30b in the diffusion case.

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

$$g_{f,\mu}(t, x) \approx 3D(x)^2 g_{f,xx}(t, x). \quad (33)$$

Note that Equation 33 is an equality when the variance and the diffusion fields are homogeneous, and when the variance is stationary at the boundary.

Hence, as for the transport, the ensemble forecasting under Dirichlet BC, 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 Equation 33.

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 20 is considered with Dirichlet boundary condition at  $x = 0$  and free boundary at  $x = \Lambda$ . For the second setting, the heterogeneous diffusion Equation 22 is considered with Dirichlet BC at both boundaries  $x = 0$  and  $x = \Lambda$ . For the third setting, the same diffusion equation is considered but this time with Neumann BC at  $x = 0$  and  $x = \Lambda$ .

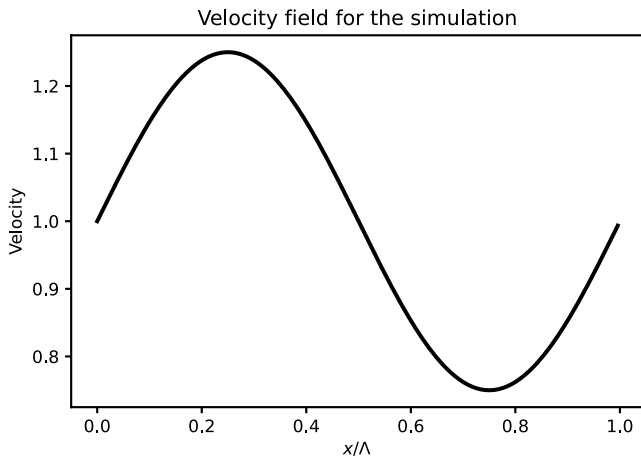
The transport and the diffusion being linear, the dynamics of the mean is the same for the PKF and for the EnKF. Hence, without loss of generality, to focus on the validation of the error statistics, the mean state is not considered in the following (the reader can consider the mean state as constant).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

Each time the variance, Equations 21b and 24b, and anisotropy tensor, Equations 21c and 24c, produced by the PKF dynamics are compared with the variance and anisotropy tensor diagnosed from an ensemble of  $N_e = 6,400$  forecasts. The size of the ensemble considered here, 6,400, is quite unusual compared with operational ensembles that often count dozens of members; but the reason is that we want to avoid sampling noise to be able to verify that the PKF dynamics of the error statistics reproduces the EnKF, the latter being considered as the reference here. Hence, if a difference appears between the EnKF and the PKF, then it is less likely to be related to sampling noise except when the magnitude of the difference falls under the confident interval deduced from the CLT.

The domain is discretized in  $n = 241$  grid points and the spatial derivative operator  $\partial_x$  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 20 is considered. The velocity wind for the simulation is set as the heterogeneous stationary field  $u(x) = 1 + \frac{1}{4} \sin\left(\frac{2\pi}{\Lambda} x\right)$  shown in Figure 1. This specific wind was considered



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

because it reproduces an average current oriented along increasing  $x$ , with an acceleration (deceleration) that mimics that encountered in a real flow.

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  $dt \approx 4 \cdot 10^{-3}$ . 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 that is, 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 the time window  $[0, t_{end}]$  brought back to a virtual physical extension of the domain by multiplying with  $-u(0)$ .

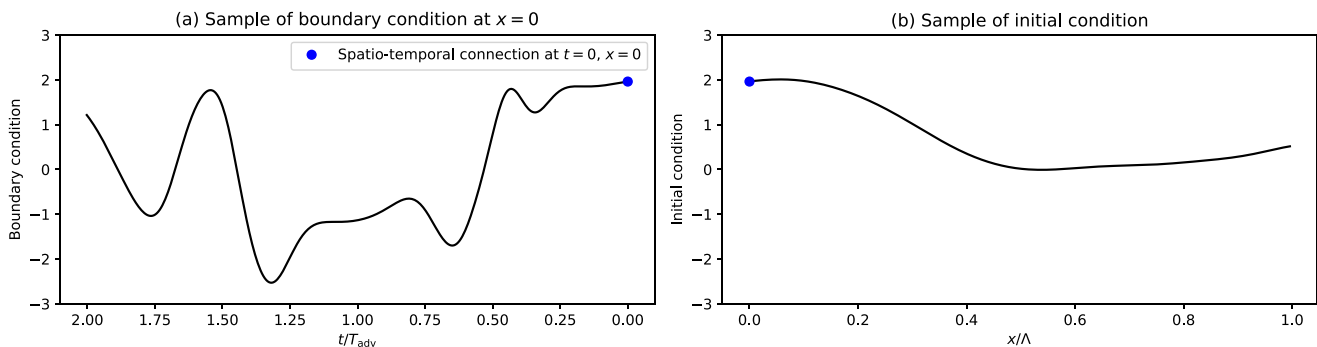
Then on this extended domain a variance field,  $V_0$  and a length-scale field  $L_0$  are defined which will be used to generate the perturbations. For this experiment the fields  $V_0$  and  $L_0$ , that constitute the PKF initial and BC, are set as follows. The initial variance is set homogeneous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

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

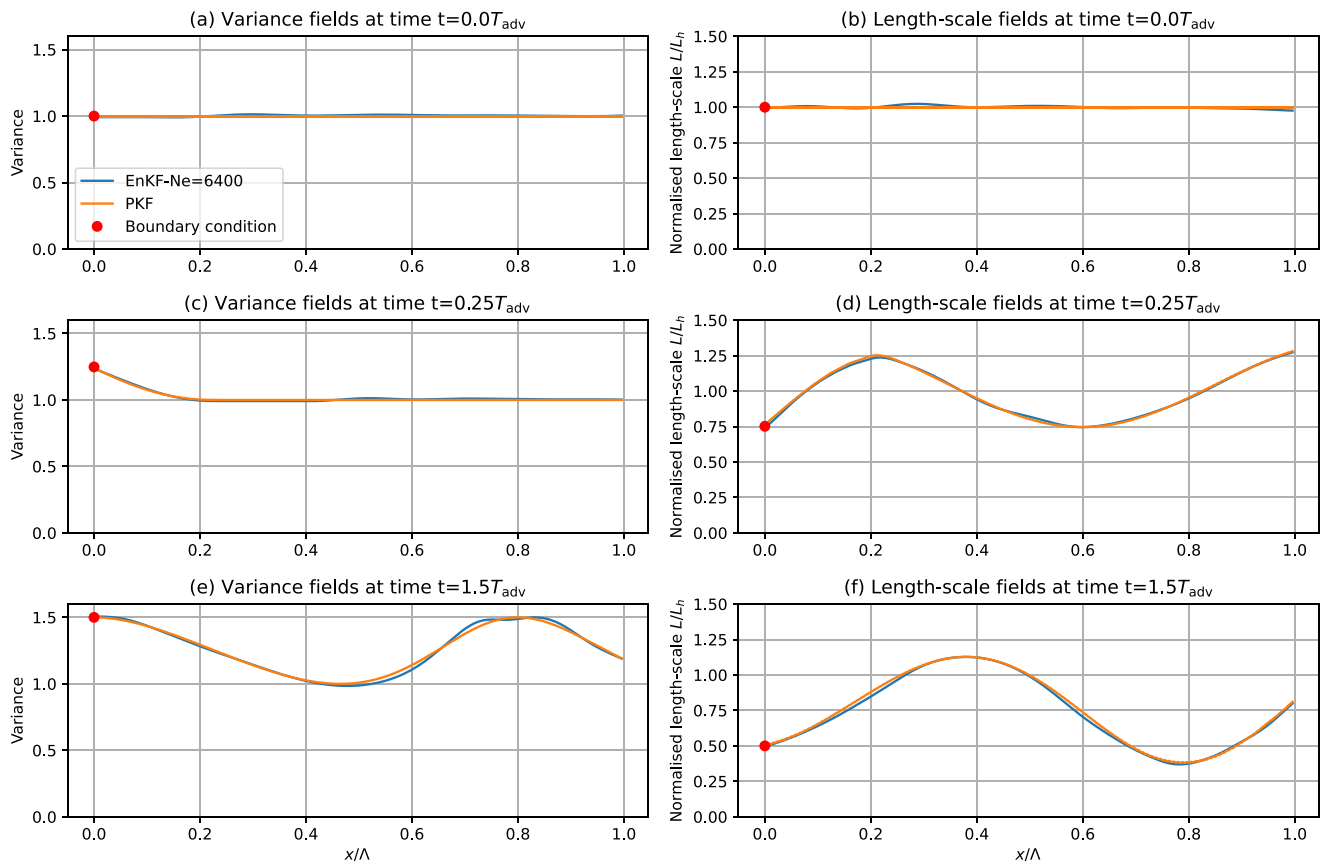
Using Equation 15 and the relation between the length-scale and the anisotropy tensor in 1D,  $s_0 = L_0^2$ , the covariance matrix,  $\mathbf{P}_0 = \mathbf{P}(V_0, s_0)$ , is defined from which the spatio-temporal perturbations are sampled for each  $k$  as  $e_k = \mathbf{P}_0^{1/2} \zeta_k$ , where  $\zeta_k$  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  $\mathbf{P}_0$ , that is,  $\mathbf{P}_0 = \mathbf{P}_0^{1/2} \left(\mathbf{P}_0^{1/2}\right)^T$ . 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 Figure 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  $x = 0$ .

Figure 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



**Figure 2.** Sample of a generated perturbation split into an initial condition and a boundary condition that are smoothly connected. Note that the time axis in panel (a) has been inverted so that the global image, formed by merging panels (a) and (b), reveals the extended domain  $[-u(0)t_{end}, \Lambda]$  that we considered for the simultaneous construction of the boundary and initial condition (see text for details).



**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  $t = 0$ ,  $t = 0.25T_{adv}$  and  $t = 1.5T_{adv}$ .

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 Equation 21c. 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.

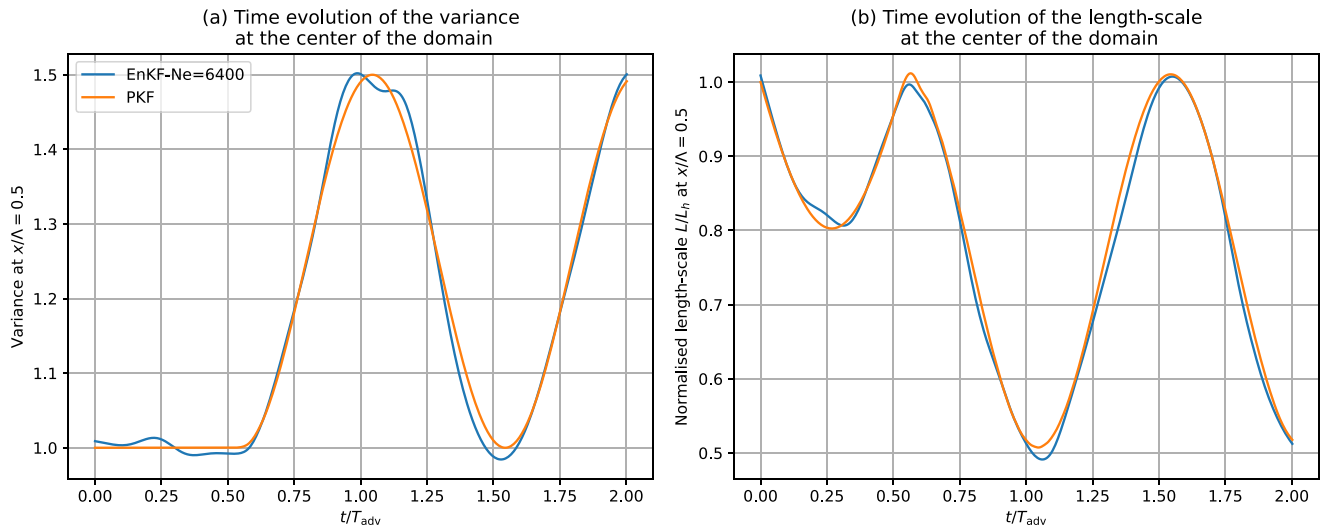
In order to strengthen these results, we show in Figure 4 the evolution through time of the fields for the middle point of the domain  $x = 0.5\Lambda$ . As expected, the variance in panel (a) remains constant until the information from the boundary condition arrives, where oscillations start, following 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  $N_e = 6,400$ .

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 Equation 32, 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 BC applied for a diffusion equation.



**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.

### 5.3. Application to the Diffusion Equation

In this experiment setting, the heterogeneous diffusion Equation 22 is considered. The temporal discretization scheme used for the ensemble simulation is a backward Euler scheme (implicit Euler method) with a fixed time-step  $dt_{BE} \approx 2 \cdot 10^{-4}$ . For the PKF dynamics we used a Runge–Kutta scheme of order 4 with a fixed time-step  $dt_{RK4} \approx 5 \cdot 10^{-6}$ . The simulation is performed from time  $t = 0$  to  $t_{end} = 1.2T_{diff}$  with  $T_{diff} = \frac{\Lambda^2}{4D_{max}}$  the time scale of the diffusion of a half-domain.

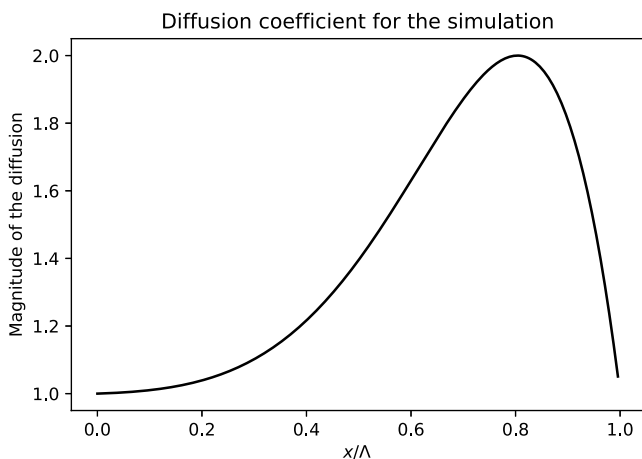
The diffusion coefficient for the simulation is set as the heterogeneous stationary field  $D(x) = 1 + \frac{A}{A_{max}}$  with  $A(x) = \sin(\pi x) (1 + x)^8$  where  $A_{max} = \text{Max}_x A(x)$ , and is shown in Figure 5. This diffusion field reproduces the kind of diffusion encountered in the dynamics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

#### 5.3.1. Dirichlet Boundary Conditions

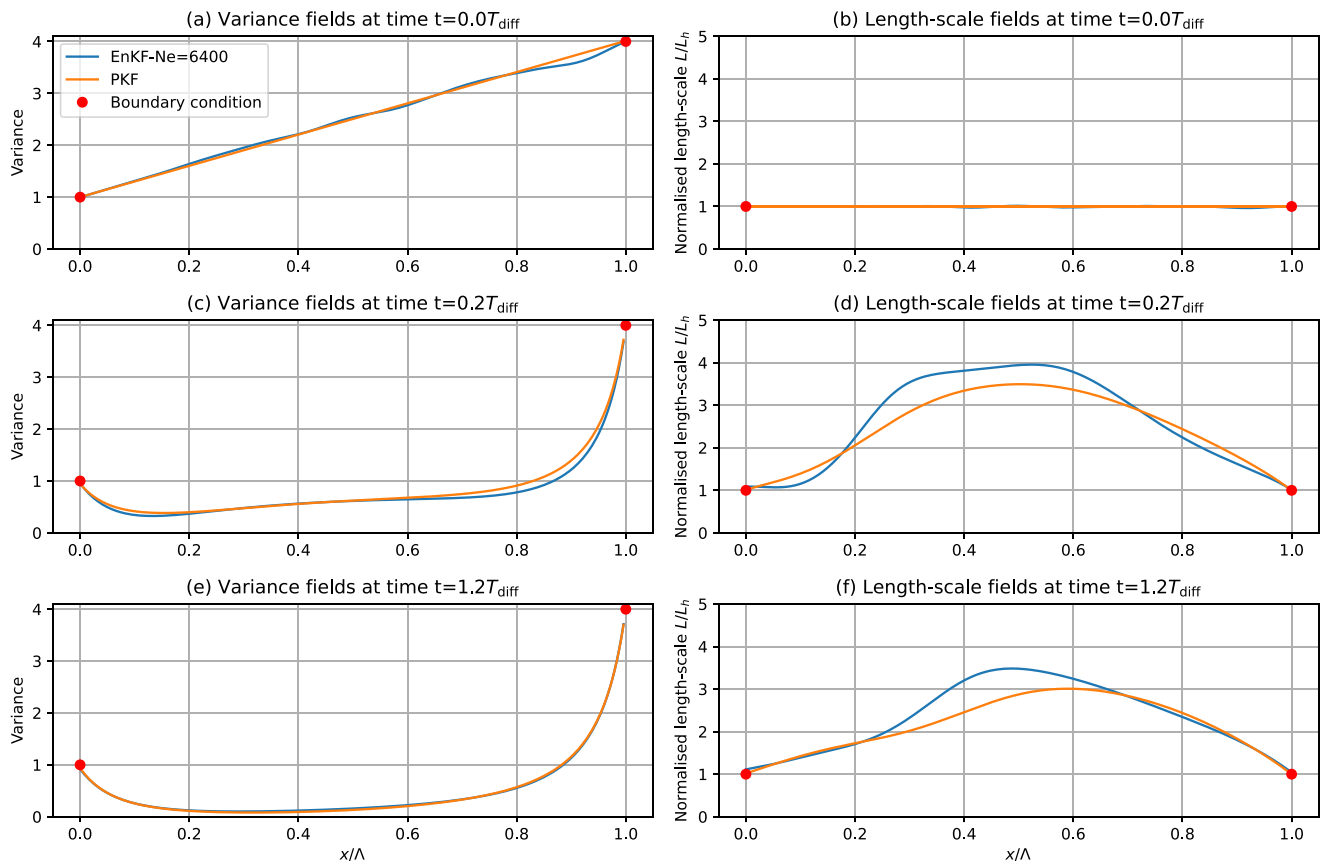
To generate a coherent set of perturbations for the ensemble simulation, the same technique described in Section 5.2 is considered except that both boundaries at  $x = 0$  and  $x = \Lambda$  are subject to Dirichlet conditions. The extended domain considered is  $[-\sqrt{D(0)t_{end}}, 0] \cup [0, \Lambda]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

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  $L_0(t = 0, x) = L_h = 0.1\Lambda$ . For the left boundary condition at  $x = 0$ , the variance and the length-scale are stationary and set equal to 1 and  $L_h$  respectively that is,  $V_0(t, x = 0) = 1$  and  $L_0(t, x = \Lambda) = L_h$ . For the right boundary condition at  $x = \Lambda$ , the variance and the length-scale are stationary and set equal to 4 and  $L_h$  respectively that is,  $V_0(t, x = \Lambda) = 4$  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 Figure 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 Figure 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  $t = 0.2T_{diff}$ .



**Figure 5.** Heterogeneous diffusion coefficient generated for the experiment.



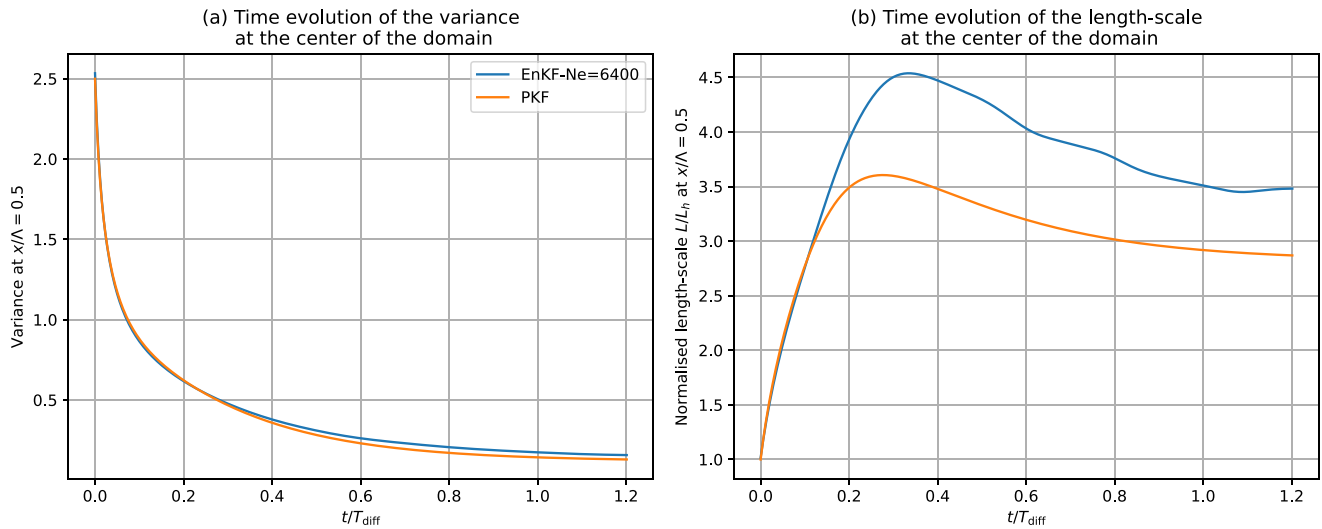
**Figure 6.** Comparison of the forecast-error variance (left column) and normalized length-scale (right column) fields dynamics for the heterogeneous diffusion equation on a 1D bounded domain with Dirichlet boundary conditions, and shown at times  $t = 0$ ,  $t = 0.2T_{diff}$ , and  $t = 1.5T_{diff}$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

Due to physical diffusion, far from the boundaries for example, at the center of the domain, the magnitude of the error is expected to decrease over time with an attenuation of the variance in  $\frac{1}{\sqrt{t}}$  (see Pannekoucke et al., 2016, Equation 25 or Pannekoucke et al., 2021, Equation 48a), while the length-scale should increase in  $\frac{1}{\sqrt{t}}$  (see Pannekoucke et al., 2016, Equation 24 or Pannekoucke et al., 2021, Equation 48b); and at the boundaries the uncertainty should remain as specified by the Dirichlet conditions. This is precisely the behavior 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 of around 20% between the length-scale computed by the PKF and the one estimated from the ensemble. This gap can be due to the closure Equations 23a and 23b but it has a limited impact on the variance field (panel c) which suggests 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 Figure 7, the variance and length-scale are close to the permanent regime at  $t = 1.2T_{diff}$  showing that the PKF performed well even over a significant time period.

Note that we consider that the PKF performs well here, even with a difference of 20% on the length-scale near the middle of the domain (see Figures 6f, 6d, and 7b), because the length-scale predicted by the PKF has the appropriate order of magnitude, and that the variance is really well reproduced (see Figures 6c, 6d, and 7a). From our experience, an error in variance is more damaging than an error in span length during assimilation cycle (see Pannekoucke, 2021a, Figure 9 or Perrot et al., 2023, Figure 7). However, the question to know if whether or not the PKF performs well should be addressed by comparing the results of the two schemes in the context of DA



**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.

with synthetic observations, in order to see how the difference of length-scale impacts the root mean squared error of the analyzed state, and see if this difference is admissible or not, which is beyond the scope of the present contribution.

To conclude, this experiment has confirmed the specification of the Dirichlet BC 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 dynamics as diagnosed from the EnKF but at a lower cost corresponding the price of two time integrations compared to the 6,400 integrations needed for the ensemble. Another result is that the simulations also validate the theoretical derivation of the time-scale setting Equation 33 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 DA. In particular the square-root covariance  $\mathbf{P}_0^{1/2}$  resulting from the integration of the pseudo-diffusion equation reads as the linear operator

$$\mathbf{P}_0^{1/2} = \mathbf{\Sigma}\mathbf{W}\mathbf{L}, \tag{34}$$

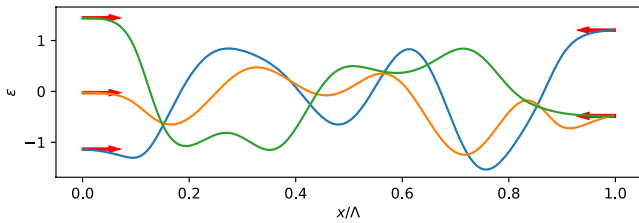
where  $\mathbf{L} = e^{\frac{1}{2}\kappa\partial_x^2}$  is the propagator associated with the diffusion equation

$$\partial_\tau u = \kappa\partial_x^2 u \tag{35}$$

of pseudo-time  $\tau$ , integrated from  $\tau = 0$  to  $\tau = \frac{1}{2}$ , and using Neumann conditions at the boundaries (Mirouze & Weaver, 2010);  $\mathbf{W}$  is a diagonal normalization so that  $\mathbf{W}\mathbf{L}(\mathbf{W}\mathbf{L})^T$  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  $\kappa$  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 Equation 34,  $e_k = \mathbf{P}_0^{1/2}\zeta_k$ , where  $\zeta_k$  is a sample of centered Gaussian random vector. Figure 8 shows some samples of the normalized error resulting from





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

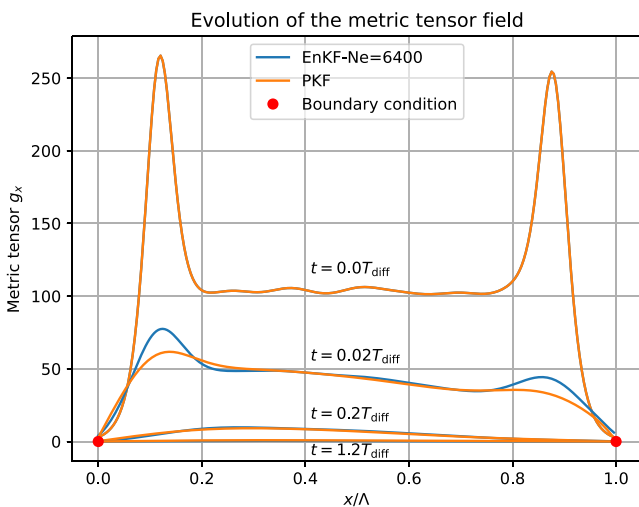
Neumann BCs, the initial condition for the PKF is the variance field of linear profile shown in Figure 10a and the experiment metric field diagnosed from the ensemble of initial conditions shown in Figure 9 for  $t = 0$  (orange line, superimposed to the blue line of the EnKF diagnosis).

The PKF dynamics is computed and the results are compared with the ensemble of forecasts of the heterogeneous diffusion Equation 22 and Neumann conditions on both sides. The results are shown in Figure 9 for the metric, and in Figure 10 for the variance and the length-scale (computed from the inverse of the metric). The results are shown for times of interest selected from the time evolution reproduced in Figure 11 where a relaxation toward a stationary state of uncertainty appears.

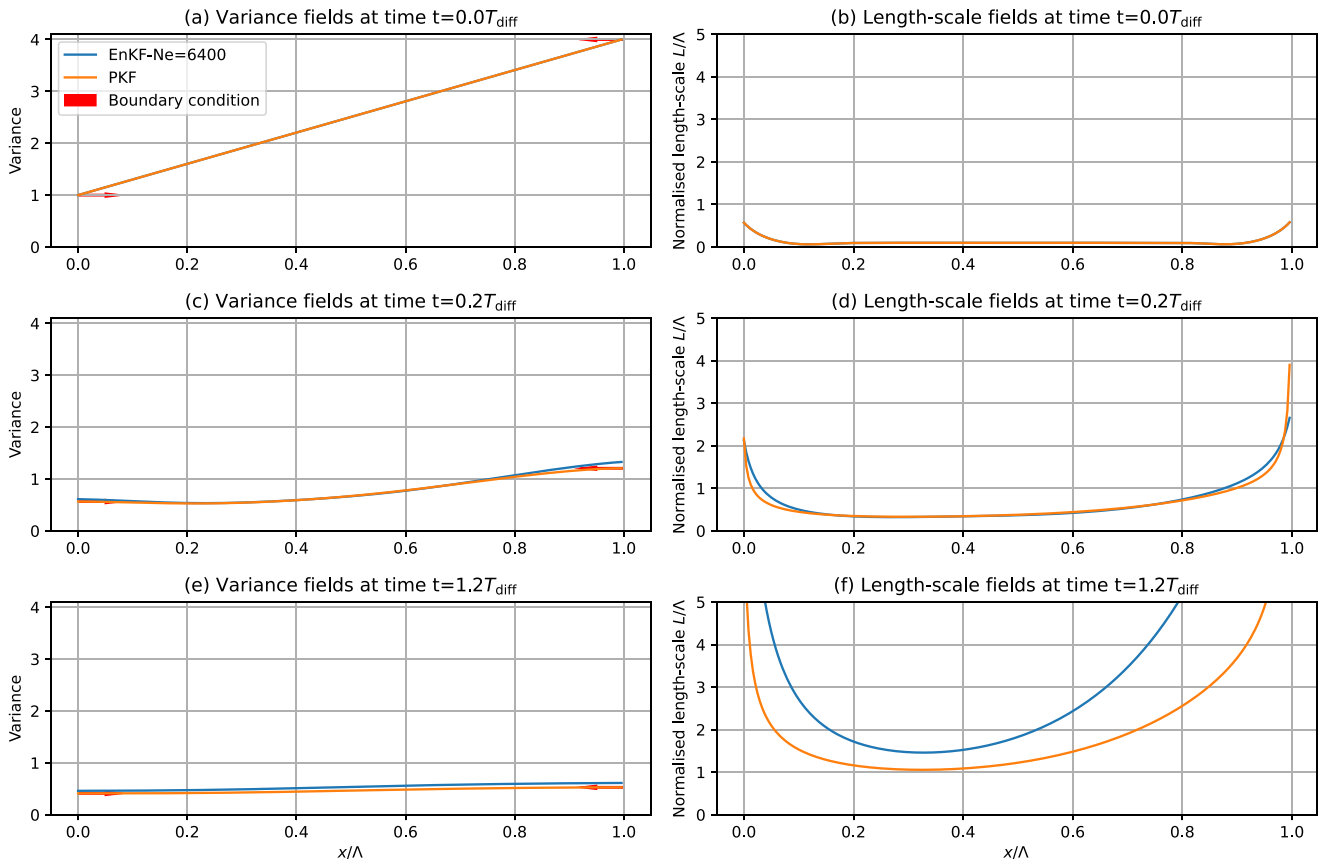
As expected for a diffusion, the variance decreases with respect to time, while the length-scale increases. Note that for Neumann condition, the variance at the boundary also decreases while it was constant in the Dirichlet condition. For the ensemble estimation, the length-scales at the boundaries (blue 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 Equation 14, while the metric remains null at the boundaries during the simulation (see Figure 9 the red dots).

Compared with the EnKF diagnosis, the PKF performs well by reproducing the same behavior 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  $\Lambda$ , as diagnosed from the ensemble is subject to the limitation of the numerical computation of Equation 14 for large correlations that can present a positive bias (Pannekoucke et al., 2008), but the magnitude of this positive bias should be low in the present experiment where a large ensemble of 6,400 members is used. Moreover, the large length-scale of the EnKF can also be influenced by model error (Pannekoucke et al., 2021), while the PKF can be less sensitive to numerical error (Perrot et al., 2023, Appendix A). 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. Note that these comments rely on the length-scale but this diagnosis is computed from the inverse of the metric tensor and that therefore, the discrepancies between the EnKF and the PKF (in terms of length-scales) will be exacerbated as the metric tensor is getting closer to 0. In terms of metric tensor, Figure 9, the agreement between EnKF and PKF seems to be much better. In addition, the very large value of the length-scale encountered here, whose amplitude can be more than five times the domain extension, should be viewed with caution: the length-scale is only a diagnosis of the local spatial extension of the correlation functions based on the curvature diagnosis (small separation distance  $\delta x$  in Equation 10) and may not be representative of the global extension (Ménétrier et al., 2015). This limitation can be seen in Figure 12 which compares some of the correlation functions diagnosed from the ensemble with the Gaussian correlation of same length-scale at the end of the time integration  $t = 1.2T_{diff}$ . It appears that the Gaussian approximation can over-estimate the correlation when the length-scale is very large for example, in Figure 12a where the length-scale is larger than  $5\Lambda$ . But at the same time, the Gaussian correlation can under-estimate the true correlation



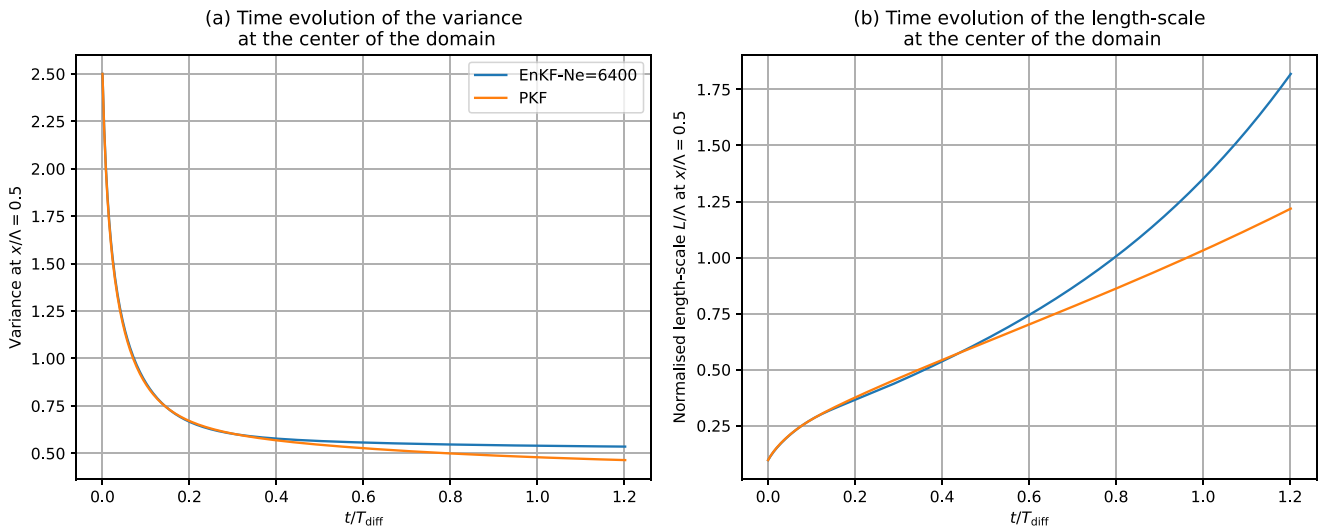
**Figure 9.** Forecast-error metric field for the heterogeneous diffusion equation on a 1D bounded domain with Neumann boundary conditions, shown at times  $t = 0$ ,  $t = 0.02T_{diff}$ ,  $t = 0.2T_{diff}$ , and  $t = 1.2T_{diff}$ .



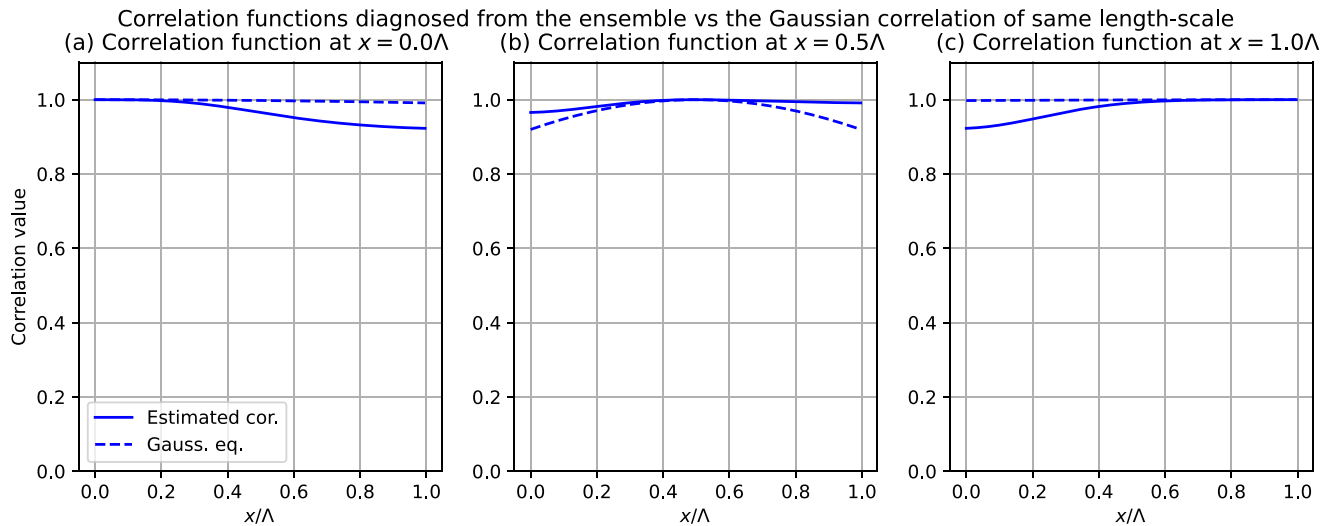
**Figure 10.** Comparison of the forecast-error variance (left column) and length-scale (right column) fields dynamics for the heterogeneous diffusion equation on a 1D bounded domain with Neumann boundary conditions, and shown at times  $t = 0$ ,  $t = 0.2T_{diff}$ , and  $t = 1.2T_{diff}$ .

at large distance as shown in Figure 12b for a length-scale around  $\Lambda$ . Hence, for correlation modeling, it may be useful to attenuate excessively large length-scale values when using the PKF.

To conclude, this experiment has confirmed the specification of the Neumann BC 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



**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.



**Figure 12.** Correlation function diagnosed from the ensemble (solid line) and compared with the Gaussian correlation of same length-scale (dashed line), for the points  $x = 0$ ,  $x = 0.5\Lambda$ , and  $x = \Lambda$ , and at  $t = 1.2T_{diff}$

uncertainty dynamics as diagnosed from the EnKF but at a lower cost corresponding to the price of two time integrations compared to the 6,400 integrations needed for the ensemble.

This ends the validation of the specification of the BC 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 PKF, that is a recent approximation of the Kalman filter proposed for application in large systems. The parametric approach investigated here consists 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 the prediction of the error statistics at a lower numerical cost than the one needed for the full covariance propagation described in the Kalman filter forecast step.

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 in air quality at regional scale or in radiation belts “weather.”

Two kind of boundaries have been considered, the Dirichlet and the Neumann conditions depending on the dynamics. We obtained that a Dirichlet boundary condition for the dynamics implies to specify a Dirichlet conditions for the variance and the anisotropy in the PKF dynamics, no matter how the PKF dynamics is specified in metric or in aspect tensor. Similarly, a Neumann condition implies to specify a Neumann condition for the variance but a Dirichlet condition for the metric. However, for Neuman conditions, the formulation of the PKF in metric is more adapted than the one in aspect tensor, which would require infinite value at the boundary.

Two dynamics of interest for weather forecasting, air quality or radiation belt dynamics have been considered: the transport and the diffusion equation. For these dynamics, the PKF predictions have been validated numerically in a 1D limited area domain, from the comparison to an ensemble estimation for example, by considering heterogeneous and non-stationary uncertainty at the boundary for Dirichlet conditions. For Dirichlet conditions, the ensemble of forecast has been designed by adding, to each member, a perturbation of the boundary with a varying correlation time-scale, for which we derived analytical formula that depend on the dynamics and introduced to obtain a prescribed spatial anisotropy near the boundary. For both the advection and the diffusion, the PKF has been shown able to reproduce the uncertainty dynamics diagnosed from the ensemble of forecast. Note that this result is not obvious, even when considering a linear-Gaussian setting: the PKF based on VLATcov model

only approximate the full covariance dynamics from the dynamics of the variance and the anisotropy, that is a crude approximation of the full covariance dynamics. Hence, it is an encouraging result for the application in air quality (for the transport equation) and in radiation belts monitoring (for the diffusion). For other applications, these results will have to be confirmed by considering non-linear dynamics of interest, and multivariate statistics extension of the PKF, as it is required for numerical weather prediction; which needs complementary exploration of the PKF beyond the scope of the present contribution and are challenging topics. Note also that the numerical testbed indirectly validates the auto-correlation time scale formula, and constitutes a contribution to the ensemble methods (while these formula are not needed for the PKF).

The next step will be to study the BC conditions for domains of larger dimensions, where we expect some changes for example, 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 DA or observation targeting applied for LAM, that could be another topic to investigate with the PKF.

Beyond these challenging topics, we can mention that the results in 1D should already find important applications for example, in the dynamics of uncertainty in the boundary layer for air quality, or wild-land fire predictions; or in the exploration of the coupling of uncertainty in the coupled atmosphere-ocean system during the alternating integration of the two fluids.

## Appendix A: Accounting for the Uncertainty at the Boundary Condition in a KF

This section illustrates the KF equation in presence of a boundary, applied to a transport. The KF forecast step is first introduced where the boundary condition appears as a command. Then it is applied when the boundary is perfectly known or uncertain. An illustration of uncertain boundary is proposed when the error is a stationary auto-regressive noise of order 1 (AR(1)), for which the formalism can be reworded without command but considering an extended domain with an equivalent model error.

### A1. KF Forecast Step for a Transport in a Limited-Area 1D Domain

The transport  $\partial_t c + U \partial_x c = 0$ , of a concentration field  $c(t, x)$  at a constant positive velocity  $U$ , along a 1D limited-area domain  $[0, \Lambda]$  of coordinate  $x$ , with a Dirichlet condition  $c(t, x = 0) = b(t)$  at the inlet and open condition at the outlet  $x = 1$ , discretized on a regular grid in time  $(t_q)_{q \in \mathbb{N}}$  and space  $(x_k)_{k \in [0, n-1]}$  (i.e.,  $x_0 = 0$ ) with a time step  $\delta t$  and a grid space  $\delta x$  defined by  $\delta t = \frac{\delta x}{U}$ , and when the boundary condition is constant over a time step, the dynamics reads as

$$c_{q+1} = \mathbf{M}c_q + \mathbf{F}b_q, \quad (\text{A1})$$

where here,  $c_{q+1}$  denote the vector resulting from the discretization of  $c(t_q, x)$ ,  $\mathbf{F} = [1, 0, \dots, 0]^T$  and

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \dots & 0 & 1 & 0 \end{bmatrix}, \quad (\text{A2})$$

so that the value of the concentration  $c(t_{q+1}, x_k)$  at a grid point  $x_k$  and at a time  $t_{q+1}$  is equals to  $c(t_q, x_{k-1})$  providing that  $0 < k < n - 1$ ; at the boundaries the first component of  $c_{q+1}$  is  $b_q$  (i.e.,  $c(t_{q+1}, 0) = b_q$ ) while the value of  $c(t_q, x_{n-1})$ . After a long period of time integration, when  $t_q > \Lambda/U$ ,  $c(t_{q+1}, x_k) = b_{q-k}$  for any  $k \in [0, n - 1]$ . Note that, in optimal control,  $b_q$  in Equation A1 plays the role of a command.

From Equation A1, it follows that the dynamics of the forecast error is written as

$$e_{c, q+1}^f = \mathbf{M}e_{c, q}^f + \mathbf{F}e_{b, q}, \quad (\text{A3})$$

where  $e_c^f(e_b)$  denotes the forecast error on  $c$  (the error on  $b$ ). As it is common in KF theory, the errors are assumed centered (i.e.,  $\mathbb{E}[e_c^f] = \mathbb{E}[e_b] = 0$  for any time).

### A2. Perfectly Known Boundary Condition

In the case where the boundary condition at the inlet is perfectly known (i.e.,  $e_{bq} = 0$  for any  $q$ ), but the initial condition is uncertain, characterized by Gaussian error distribution of covariance  $\mathbf{P}^a$ ; then the Gaussian uncertainty, characterized by the covariance  $\mathbf{P}_q^f$  at  $t_q$ , evolves following the Kalman filter forecast step Equation 4 that is

$$\mathbf{P}_{q+1}^f = \mathbf{M}\mathbf{P}_q^f\mathbf{M}^T. \quad (\text{A4})$$

Hence, after one time step ( $q = 1$ ), the forecast-error covariance matrix reads as  $\mathbf{P}_1^f = \mathbf{M}\mathbf{P}_0^f\mathbf{M}^T$  ( $\mathbf{P}_0^f = \mathbf{P}^a$ ), that is a matrix that looks like

$$\mathbf{P}_1^f = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & & & & \\ \vdots & & (\mathbf{P}_0^f)^- & & \\ 0 & & & & \end{bmatrix} \quad (\text{A5})$$

where  $(\mathbf{P}_0^f)^-$  denotes square matrix of dimension  $n - 1$  whose coefficients are those of the submatrix  $\mathbf{P}_{[0:n-2,0:n-2]}^a$  of  $\mathbf{P}^a$  since  $\mathbf{P}_0^f = \mathbf{P}^a$ . It appears that the variance at the boundary is null, which is coherent when the boundary condition (not necessary constant in time) is perfectly known. After,  $q > 0$  iterations of Equation A4,  $\mathbf{P}_{q+1}^f = (\mathbf{M})^q\mathbf{P}_0^f(\mathbf{M}^T)^q$ , the  $q$ th first rows and lines of  $\mathbf{P}_q^f$  are null. Since  $\mathbf{M}$  is nilpotent of order  $n$  (i.e.,  $\mathbf{M}^{n-1} \neq 0$  and  $\mathbf{M}^n = 0$ ), after  $n$  iterations, the forecast-error covariance matrix is null, meaning that there is no more error within the domain (independently of the initial uncertainty of the initial condition).

While in this section we considered a perfectly known boundary, this is not the situation encountered in real applications for example, in forecasting at regional scale where uncertainty at large scale should introduce an inflow of uncertainty at the entrance of the domain.

Hence, the inflow of ‘‘certainty’’ that results from the common covariance propagation Equation A4 of the Kalman filter forecast step, but used without care, explains the loss of variance observed in ensemble forecasting when no perturbation of the boundary is used to represent the uncertainty (Nutter, Stensrud, & Xue, 2004; Nutter, Xue, & Stensrud, 2004).

The next section provides the correct formulation of the covariance dynamics in presence of uncertainty at the boundary.

### A3. Uncertain Boundary Condition

Now, when considering that the boundary condition is uncertain, the forecast-error covariance matrix evolves as

$$\mathbf{P}_q^f = \mathbf{M}\mathbf{P}_q^f\mathbf{M}^T + \mathbf{F}\mathbb{E}\left[e_{bq}(e_{bq})^T\right]\mathbf{F}^T + \mathbf{M}\mathbb{E}\left[e_{c,q}^f(e_{bq})^T\right]\mathbf{F}^T + \left(\mathbf{M}\mathbb{E}\left[e_{c,q}^f(e_{bq})^T\right]\mathbf{F}^T\right)^T, \quad (\text{A6})$$

and reads as

$$\mathbf{P}_q^f = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & (\mathbf{P}_q^f)^- & \\ 0 & & & \end{bmatrix} + \begin{bmatrix} V_{bq} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & (0) & \\ 0 & & & \end{bmatrix} + \begin{bmatrix} 0 & 0 & \dots & 0 \\ * & & & \\ \vdots & & (0) & \\ * & & & \end{bmatrix} + \begin{bmatrix} 0 & * & \dots & * \\ 0 & & & \\ \vdots & & (0) & \\ 0 & & & \end{bmatrix}, \quad (\text{A7})$$

where  $V_{bq} = \mathbb{E}\left[e_{bq}(e_{bq})^T\right]$  is the variance of the uncertainty at the boundary, and where stars, \*, correspond to a priori non zero coefficients which represents the correlations between the uncertainty at the boundary with the forecast-error in the remaining part of the domain  $(0, \Lambda]$ . Hence, the first term in Equation A10,  $\mathbf{M}\mathbf{P}_q^f\mathbf{M}^T$ ,

is similar to the matrix shown in Equation A5. The other terms in Equation A10 will contribute to fill the first row and column of the matrix: the second term of Equation A10,  $\mathbf{F}\mathbb{E}\left[e_{bq}(e_{bq})^T\right]\mathbf{F}^T = V_{bq}\mathbf{F}\mathbf{F}^T$ , where  $\mathbf{F}\mathbf{F}^T$  is the null matrix except for the first coefficient that is one, fills the first coefficient of  $\mathbf{P}_{q+1}^f$ . Similarly, the third and the fourth terms of Equation A10 will fill the first row and the first column.

#### A4. Example With a Stationary AR(1) Boundary Error

As an instructive example, we consider the case where the error is a stationary AR(1) random process,  $e_{bq+1} = ae_{bq} + \sqrt{1-a^2}\sqrt{V_b}\zeta_q$ , where  $V_b$  denotes the time independent variance of  $e_{bq}$  (independent of  $q$ );  $a = \exp(-\delta t/\tau)$  where  $\tau$  is the correlation time scale; and  $\zeta = (\zeta_q)_{q \in \mathcal{N}}$  is a process of independent and independent random variable, where for each  $q$ ,  $\zeta_q$  is normal law that is,  $\zeta_q \sim \mathcal{N}(0, 1)$ . The time correlation for this processes is the exponential  $cor(e_{bq}, e_{bq+r}) = \exp(-r\delta t/\tau)$ , that also reads as  $cor(e_{bq}, e_{bq-r}) = \exp(-r\delta t/\tau)$  because of the stationarity of the process (homogeneity in time).

Since after a long period,  $c(t_{q+1}, x_k) = b_{q-k}$ , it follows that  $e_c(t_{q+1}, x_k) = e_{bq-k}$  so that the  $k$ th component of the vector  $\mathbb{E}\left[e_c^f(t_{q+1}, x_k)\right]$  is  $cov(e_c^f(t_{q+1}, x_k), e_{bq}) = V_b cor(e_{bq}, e_{bq-k}) = V_b cor(e_{bq+k}, e_{bq}) = V_b cor(e_{bq+k}, e_{bq}) = V_b \exp(-k\delta t/\tau)$ . With  $e_c^f(t_{q+1}, 0) = e_{bq}$ ,  $e_c^f(t_{q+1}, x_k) = e_c^f(t_{q+1}, x_{k-1})$ , and  $\delta t = \delta x/U$ , then  $cov(e_c^f(t_{q+1}, x_0), e_c^f(t_{q+1}, x_k)) = V_b \exp(-k\delta x/(U\tau))$ . Hence, after a relaxation period, the error covariance matrix is stationary with a spatial correlation function given as an exponential correlation of length-scale  $l = U\tau$ .

#### A5. Extension of the Space Domain for a Boundary Error Modeled by an AR(1)

When the boundary error is an auto-regressive process of order 1, it is possible to rewords the dynamics Equation A3, by extending the physical domain with an additional one point corresponding to the boundary condition that will enter to the domain in the future. Focusing on error dynamics, this reads as follow:

Considering the extended vector  $\tilde{e}_q = [e_{bq}, e_q^f]^T$  of dimension  $n + 1$ , then Equation A3 is written as

$$\tilde{e}_{q+1} = \tilde{\mathbf{M}}\tilde{e}_q + \tilde{\mathbf{F}}\zeta_q, \quad (\text{A8})$$

where

$$\tilde{\mathbf{M}} = \begin{bmatrix} a & 0 \\ \mathbf{F} & \mathbf{M} \end{bmatrix}, \quad (\text{A9})$$

and  $\tilde{\mathbf{F}} = [\sqrt{1-a^2}V_b, 0, \dots, 0]^T$ . With this formulation, Equation A8 is similar to the classic formulation of the KF equation in presence of model error (that would be here the second term  $\tilde{\mathbf{F}}\zeta_q$ ), leading to the error covariance dynamics

$$\tilde{\mathbf{P}}_{q+1}^f = \tilde{\mathbf{M}}\tilde{\mathbf{P}}_q^f\tilde{\mathbf{M}}^T + \tilde{\mathbf{Q}}, \quad (\text{A10})$$

where  $\tilde{\mathbf{Q}} = \tilde{\mathbf{F}}\tilde{\mathbf{F}}^T$ .

This extension of the domain also applies for auto-regressive process of order  $m$  where we need to add  $m$  points.

## Appendix B: Closure of the PKF Dynamics for the Diffusion Equation

The computation of the PKF dynamics for the diffusion Equation 22, with SymPKF, leads to the dynamical system

$$\partial_t f = D\partial_x^2 f + \partial_x D\partial_x f, \quad (\text{B1a})$$

$$\partial_t V_f = -\frac{2DV_f}{s_{f,xx}} + D\partial_x^2 V_f - \frac{D(\partial_x V_f)^2}{2V_f} + \partial_x D\partial_x V_f, \quad (\text{B1b})$$



$$\begin{aligned}
 \partial_t s_{f,xx} &= 2Ds_{f,xx}^2 \mathbb{E}(\varepsilon_f \partial_x^4 \varepsilon_f) - 3D\partial_x^2 s_{f,xx} - 2D \\
 &+ \frac{6D(\partial_x s_{f,xx})^2}{s_{f,xx}} - \frac{2Ds_{f,xx} \partial_x^2 V_f}{V_f} + \frac{D\partial_x V_f \partial_x s_{f,xx}}{V_f} \\
 &+ \frac{2Ds_{f,xx} (\partial_x V_f)^2}{V_f^2} - 2s_{f,xx} \frac{d^2}{dx^2} D \\
 &+ 2\partial_x D \partial_x s_{f,xx} - \frac{2s_{f,xx} \partial_x D \partial_x V_f}{V_f}
 \end{aligned} \tag{B1c}$$

where this time the term  $\mathbb{E}[\varepsilon_f \partial_x^4 \varepsilon_f]$  is not determined from  $f$ ,  $V_f$  and  $s_{f,xx}$ . This dynamics can be closed considering the closure Equation 23b.

### Appendix C: Specification of the Temporal Metric Tensor for Evolution Equations

This section details the link between the temporal metric Equation 30a,  $g_{tt} = \mathbb{E}[\partial_t \varepsilon \partial_t \varepsilon]$ , and the dynamics of the error. Since the trend of the normalized error reads as

$$\partial_t \varepsilon = \frac{1}{\sqrt{V}} \partial_t e - \frac{1}{2V^{3/2}} e \partial_t V, \tag{C1}$$

then the temporal metric tensor is written as

$$g_{tt} = \frac{1}{V} \mathbb{E}[(\partial_t e)^2] - \frac{1}{V^2} \mathbb{E}[e \partial_t e] \partial_t V + \frac{1}{4V^3} \mathbb{E}[e^2] (\partial_t V)^2. \tag{C2}$$

However, we recognize the expression of the variance  $V = \mathbb{E}[e^2]$  and its trend, Equation 16a, so that the temporal metric simplifies as

$$g_{tt} = \frac{1}{V} \mathbb{E}[(\partial_t e)^2] - \frac{1}{4V^2} (\partial_t V)^2. \tag{C3a}$$

Introducing the trend of the error Equation 3 and by definition of  $\varepsilon = e/\sqrt{V}$ , the temporal metric reads as

$$g_{tt} = \frac{1}{V} \mathbb{E} \left[ \left( \mathcal{M}(\varepsilon \sqrt{V}, \partial_t(\varepsilon \sqrt{V})) \right)^2 \right] - \frac{1}{4V^2} (\partial_t V)^2. \tag{C3b}$$

### Appendix D: 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

$$\begin{aligned}
 g_{f,tt} &= D^2 \mathbb{E}(\varepsilon_f \partial_x^4 \varepsilon_f) + 2D^2 \partial_x^2 g_{f,xx} - \frac{D^2 g_{f,xx} \partial_x^2 V_f}{V_f} \\
 &+ \frac{D^2 \partial_x V_f \partial_x g_{f,xx}}{V_f} + \frac{3D^2 g_{f,xx} (\partial_x V_f)^2}{2V_f^2} + \frac{D^2 (\partial_x^2 V_f)^2}{4V_f^2} \\
 &- \frac{D^2 (\partial_x V_f)^2 \partial_x^2 V_f}{4V_f^3} + \frac{D^2 (\partial_x V_f)^4}{16V_f^4} + D\partial_x D \partial_x g_{f,xx} \\
 &+ \frac{Dg_{f,xx} \partial_x D \partial_x V_f}{V_f} + \frac{Dg_{f,xx} \partial_t V_f}{V_f} + \frac{D\partial_x D \partial_x V_f \partial_x^2 V_f}{2V_f^2} \\
 &- \frac{D\partial_t V_f \partial_x^2 V_f}{2V_f^2} - \frac{D\partial_x D (\partial_x V_f)^3}{4V_f^3} + \frac{D\partial_t V_f (\partial_x V_f)^2}{4V_f^3} \\
 &+ g_{f,xx} (\partial_x D)^2 + \frac{(\partial_x D)^2 (\partial_x V_f)^2}{4V_f^2} - \frac{\partial_x D \partial_t V_f \partial_x V_f}{2V_f^2} \\
 &+ \frac{(\partial_t V_f)^2}{4V_f^2}.
 \end{aligned} \tag{D1}$$

Considering the analytical closure Equation 23a for the unclosed term  $\mathbb{E}(\epsilon_f \partial_x^3 \epsilon_f)$ , the correspondence is written as

$$\begin{aligned}
 g_{f,tt} = & 3D^2 g_{f,xx}^2 - \frac{D^2 g_{f,xx} \partial_x^2 V_f}{V_f} + \frac{D^2 \partial_x V_f \partial_x g_{f,xx}}{V_f} \\
 & + \frac{3D^2 g_{f,xx} (\partial_x V_f)^2}{2V_f^2} + \frac{D^2 (\partial_x^2 V_f)^2}{4V_f^2} - \frac{D^2 (\partial_x V_f)^2 \partial_x^2 V_f}{4V_f^3} \\
 & + \frac{D^2 (\partial_x V_f)^4}{16V_f^4} + D \partial_x D \partial_x g_{f,xx} + \frac{D g_{f,xx} \partial_x D \partial_x V_f}{V_f} \\
 & + \frac{D g_{f,xx} \partial_t V_f}{V_f} + \frac{D \partial_x D \partial_x V_f \partial_x^2 V_f}{2V_f^2} \\
 & - \frac{D \partial_t V_f \partial_x^2 V_f}{2V_f^2} - \frac{D \partial_x D (\partial_x V_f)^3}{4V_f^3} + \frac{D \partial_t V_f (\partial_x V_f)^2}{4V_f^3} \\
 & + g_{f,xx} (\partial_x D)^2 + \frac{(\partial_x D)^2 (\partial_x V_f)^2}{4V_f^2} - \frac{\partial_x D \partial_t V_f \partial_x V_f}{2V_f^2} + \frac{(\partial_t V_f)^2}{4V_f^2}.
 \end{aligned} \tag{D2}$$

The latter expression being quite complex, simplifications are introduced. First the variance field is assumed locally homogeneous at the boundary that is,  $\partial_x V_f(t, x = 0) = 0$ , so that Equation D2 simplifies as

$$\begin{aligned}
 g_{f,tt} = & 3D^2 g_{f,xx}^2 + D \partial_x D \partial_x g_{f,xx} + \frac{D g_{f,xx} \partial_t V_f}{V_f} \\
 & + g_{f,xx} (\partial_x D)^2 + \frac{(\partial_t V_f)^2}{4V_f^2}.
 \end{aligned} \tag{D3}$$

Then, if the variance is moreover assumed stationary, then Equation D3 becomes

$$g_{f,tt} = 3D^2 g_{f,xx}^2 + D \partial_x D \partial_x g_{f,xx} + g_{f,xx} (\partial_x D)^2. \tag{D4}$$

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

$$g_{f,tt} = 3D^2 g_{f,xx}. \tag{D5}$$

While Equation D5 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 Equation D2 can be obtained when first considering the dynamics of the variance given by Equation 24b, leading to replace the trend of the variance by  $\partial_t V_f = -2D V_f g_{f,xx} + D \partial_x^2 V_f - \frac{D (\partial_x V_f)^2}{2V_f} + \partial_x D \partial_x V_f$ , so that Equation D2 simplifies as

$$\begin{aligned}
 g_{f,tt} = & 2D^2 g_{f,xx}^2 + \frac{D^2 \partial_x V_f \partial_x g_{f,xx}}{V_f} \\
 & + \frac{D^2 g_{f,xx} (\partial_x V_f)^2}{V_f^2} + D \partial_x D \partial_x g_{f,xx} \\
 & + \frac{2D g_{f,xx} \partial_x D \partial_x V_f}{V_f} + g_{f,xx} (\partial_x D)^2,
 \end{aligned} \tag{D6}$$

from which the assumption of local homogeneity at the boundary that is,  $\partial_x V_f(t, x = 0) = 0$ , leads to

$$g_{f,tt} = 2D^2 g_{f,xx}^2 + D \partial_x D \partial_x g_{f,xx} + g_{f,xx} (\partial_x D)^2. \tag{D7}$$

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

$$g_{f,tt} = 2D^2 g_{f,xx}, \quad (D8)$$

which is a different result from Equation D5.

It is not clear whether the appropriate consistency should be given by Equations D5 or D8 that is, if it is right to replace the trend of the variance Equation 24b in the consistency relation Equation D2.

From numerical experiment, it appears that setting the time auto-correlation of boundary perturbation with Equation D5 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.

## Data Availability Statement

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](https://doi.org/10.5281/zenodo.7193985) and developed openly at <https://github.com/opannekoucke/pkf-boundary> (Sabathier et al., 2022).

## References

- Amraoui, L. E., Sič, B., Piacentini, A., Marécal, V., Frebourg, N., & Attié, J.-L. (2020). Aerosol data assimilation in the MOCAGE chemical transport model during the TRAQA/ChArMEx campaign: Lidar observations. *Atmospheric Measurement Techniques*, 13(9), 4645–4667. <https://doi.org/10.5194/amt-13-4645-2020>
- Bishop, C., Etherton, B., & Majumdar, S. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129(3), 420–436. [https://doi.org/10.1175/1520-0493\(2001\)129\(0420:ASWTET\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129(0420:ASWTET)2.0.CO;2)
- Bocquet, M. (2011). Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5), 735–750. <https://doi.org/10.5194/npg-18-735-2011>
- Bourdarie, S. A., & Maget, V. F. (2012). Electron radiation belt data assimilation with an ensemble Kalman filter relying on the Salammbó code. *Annales Geophysicae*, 30(6), 929–943. <https://doi.org/10.5194/angeo-30-929-2012>
- Bouttier, F. (1994). A dynamical estimation of error covariances in an assimilation system. *Monthly Weather Review*, 122(10), 2376–2390. [https://doi.org/10.1175/1520-0493\(1994\)122<2376:adeofe>2.0.co;2](https://doi.org/10.1175/1520-0493(1994)122<2376:adeofe>2.0.co;2)
- Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, 131(607), 1013–1043. <https://doi.org/10.1256/QJ.04.15>
- Burgers, G., van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6), 1719–1724. [https://doi.org/10.1175/1520-0493\(1998\)126\(1719:ASITEK\)2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126(1719:ASITEK)2.0.CO;2)
- CAMS. (2022). The CAMS European air quality ensemble forecasts welcomes two new state-of-the-art models. Retrieved from <https://atmosphere.copernicus.eu/cams-european-air-quality-ensemble-forecasts-welcomes-two-new-state-art-models>
- Cohn, S. E. (1993). Dynamics of short-term univariate forecast error covariances. *Monthly Weather Review*, 121(11), 3123–3149. [https://doi.org/10.1175/1520-0493\(1993\)121<3123:dostuf>2.0.co;2](https://doi.org/10.1175/1520-0493(1993)121<3123:dostuf>2.0.co;2)
- Dahmen, N., Rogier, F., & Maget, V. (2020). On the modelling of highly anisotropic diffusion for electron radiation belt dynamic codes. *Computer Physics Communications*, 254, 107342. <https://doi.org/10.1016/j.cpc.2020.107342>
- Derber, J., Purser, R., Wu, W.-S., Treadon, R., Pondeca, M., Parrish, D., & Kleist, D. (2003). Flow dependent Jb in a global grid-point 3D-var. In *Proceedings of the ECMWF annual seminar on recent developments in data assimilation for atmosphere and ocean* (pp. 125–134).
- Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C., & Crook, N. A. (2004). Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Monthly Weather Review*, 132(8), 1982–2005. [https://doi.org/10.1175/1520-0493\(2004\)132\(1982:watrit\)2.0.co;2](https://doi.org/10.1175/1520-0493(2004)132(1982:watrit)2.0.co;2)
- Errico, R. M., Vukićević, T., & Raeder, K. (1993). Comparison of initial and lateral boundary condition sensitivity for a limited-area model. *Tellus A: Dynamic Meteorology and Oceanography*, 45(5), 539. <https://doi.org/10.3402/tellusa.v45i5.15057>
- Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter*. Springer-Verlag Berlin Heidelberg. <https://doi.org/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. <https://doi.org/10.1137/21M1442449>
- Gustafsson, N. (2012). 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. <https://doi.org/10.3402/tellusa.v64i0.17518>
- Harlim, J., & Hunt, B. R. (2007). A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus A*, 59(2), 225–237. <https://doi.org/10.1111/j.1600-0870.2007.00225.x>
- Houtekamer, P., & Mitchell, H. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1), 123–137. [https://doi.org/10.1175/1520-0493\(2001\)129\(0123:ASEKFF\)2.0.CO;2](https://doi.org/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(12), 4489–4532. <https://doi.org/10.1175/MWR-D-15-0440.1>
- Jazwinski, A. (1970). *Stochastic processes and filtering theory*. Academic Press.
- Lorenç, A. (2003). The potential of ensemble Kalman filter for NWP – A comparison with 4D-var. *Quarterly Journal of the Royal Meteorological Society*, 129(595), 3183–3203. <https://doi.org/10.1256/qj.02.132>
- Maget, V., Sicard-Piet, A., Bourdarie, S., Lazaro, D., Turner, D. L., Daglis, I. A., & Sandberg, I. (2015). Improved outer boundary conditions for outer radiation belt data assimilation using THEMIS-SST data and the Salammbó-EnKF code. *Journal of Geophysical Research: Space Physics*, 120(7), 2015JA021001. <https://doi.org/10.1002/2015JA021001>
- Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., et al. (2015). A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geoscientific Model Development*, 8(9), 2777–2813. <https://doi.org/10.5194/gmd-8-2777-2015>

## Acknowledgments

We would like to thank the three anonymous referees and Terence O’Kane for their fruitful comments, which helped to improve the article. We 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).



- Massart, S., Pajot, B., Piacentini, A., & Pannekoucke, O. (2010). On the merits of using a 3D-FGAT assimilation scheme with an outer loop for atmospheric situations governed by transport. *Monthly Weather Review*, *138*(12), 4509–4522. <https://doi.org/10.1175/2010MWR3237.1>
- Ménard, R., Skachko, S., & Pannekoucke, O. (2021). Numerical discretization causing error variance loss and the need for inflation. *Quarterly Journal of the Royal Meteorological Society*, *147*(740), 3498–3520. <https://doi.org/10.1002/qj.4139>
- Ménétrier, B., Montmerle, T., Michel, Y., & Berre, L. (2015). Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review*, *143*(5), 1622–1643. <https://doi.org/10.1175/mwr-d-14-00157.1>
- Mirouze, I., & Weaver, A. T. (2010). Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, *136*(651), 1421–1443. <https://doi.org/10.1002/qj.643>
- Nutter, P., Stensrud, D., & Xue, M. (2004). Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Monthly Weather Review*, *132*(10), 2358–2377. [https://doi.org/10.1175/1520-0493\(2004\)132\(2358:EOCRAT\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132(2358:EOCRAT)2.0.CO;2)
- Nutter, P., Xue, M., & Stensrud, D. (2004). Application of lateral boundary condition perturbations to help restore dispersion in limited-area ensemble forecasts. *Monthly Weather Review*, *132*(10), 2378–2390. [https://doi.org/10.1175/1520-0493\(2004\)132\(2378:AOLBCP\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132(2378:AOLBCP)2.0.CO;2)
- Paciorek, C., & Schervish, M. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*(5), 483–506. <https://doi.org/10.1002/env.785>
- Pannekoucke, O. (2021a). An anisotropic formulation of the parametric Kalman filter assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *73*(1), 1–27. <https://doi.org/10.1080/16000870.2021.1926660>
- Pannekoucke, O. (2021b). opannekoucke/sympkf: A symbolic and computational toolbox for the design of parametric Kalman filter dynamics. Retrieved from <https://github.com/opannekoucke/sympkf>
- Pannekoucke, O., & Arbogast, P. (2021). SymPKF (v1.0): A symbolic and computational toolbox for the design of parametric Kalman filter dynamics. *Geoscientific Model Development*, *14*(10), 5957–5976. <https://doi.org/10.5194/gmd-14-5957-2021>
- Pannekoucke, O., Berre, L., & Desroziers, G. (2008). Background error correlation length-scale estimates and their sampling statistics. *Quarterly Journal Royal Meteorological Society*, *134*(631), 497–508. <https://doi.org/10.1002/qj.212>
- Pannekoucke, O., Bocquet, M., & Ménard, R. (2018). Parametric covariance dynamics for the nonlinear diffusive Burgers' equation. *Nonlinear Processes in Geophysics*, *2018*, 1–21. <https://doi.org/10.5194/npg-2018-10>
- Pannekoucke, O., & Fablet, R. (2020). PDE-NetGen 1.0: From symbolic partial differential equation (PDE) representations of physical processes to trainable neural network representations. *Geoscientific Model Development*, *13*(7), 3373–3382. <https://doi.org/10.5194/gmd-13-3373-2020>
- 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*(635), 1425–1438. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.3402/tellusa.v68.31547>
- Perrot, A., Pannekoucke, O., & Guidard, V. (2023). Toward a multivariate formulation of the parametric Kalman filter assimilation: Application to a simplified chemical transport model. *Nonlinear Processes in Geophysics*, *30*(2), 139–166. <https://doi.org/10.5194/npg-30-139-2023>
- Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- Sabathier, M., Pannekoucke, O., & Maget, V. (2022). opannekoucke/pkf-boundary: Boundary conditions for the parametric Kalman filter forecast [Software]. Zenodo. <https://doi.org/10.5281/zenodo.7193985>
- Scarpa, F., & Milano, G. (1995). Kalman smoothing technique applied to the inverse heat conduction problem. *Numerical Heat Transfer, Part B: Fundamentals*, *28*(1), 79–96. <https://doi.org/10.1080/10407799508928822>
- Snyder, C., & Zhang, F. (2003). Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, *131*(8), 1663–1677. <https://doi.org/10.1175/2555.1>
- Strassen, V. (1969). Gaussian elimination is not optimal. *Numerische Mathematik*, *13*(4), 354–356. <https://doi.org/10.1007/bf02165411>
- Stull, R. (1988). *An introduction to boundary layer meteorology*. Kluwer Academic.
- Tippett, M., Anderson, J., Bishop, C., Hamill, T., & Whitaker, J. (2003). Ensemble square root filters. *Monthly Weather Review*, *131*(7), 1485–1490. [https://doi.org/10.1175/1520-0493\(2003\)131\(1485:ESRF\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131(1485:ESRF)2.0.CO;2)
- Torn, R. D., Hakim, G. J., & Snyder, C. (2006). Boundary conditions for limited-area ensemble Kalman filters. *Monthly Weather Review*, *134*(9), 2490–2502. <https://doi.org/10.1175/mwr3187.1>
- Voshtani, S., Ménard, R., Walker, T. W., & Hakami, A. (2022). Assimilation of GOSAT methane in the hemispheric CMAQ; Part I: Design of the assimilation system. *Remote Sensing*, *14*(2), 371. <https://doi.org/10.3390/rs14020371>
- Vukicevic, T., & Errico, R. M. (1990). The influence of artificial and physical factors upon predictability estimates using a complex limited-area model. *Monthly Weather Review*, *118*(7), 1460–1482. [https://doi.org/10.1175/1520-0493\(1990\)118\(1460:tioaap\)2.0.co;2](https://doi.org/10.1175/1520-0493(1990)118(1460:tioaap)2.0.co;2)
- Vukicevic, T., & Paegle, J. (1989). The influence of one-way interacting lateral boundary conditions upon predictability of flow in bounded numerical models. *Monthly Weather Review*, *117*(2), 340–350. [https://doi.org/10.1175/1520-0493\(1989\)117\(0340:tioowi\)2.0.co;2](https://doi.org/10.1175/1520-0493(1989)117(0340:tioowi)2.0.co;2)
- Warner, T. T., Key, L. E., & Lario, A. M. (1989). Sensitivity of mesoscale-model forecast skill to some initial-data characteristics, data density, data position, analysis procedure and measurement error. *Monthly Weather Review*, *117*(6), 1281–1310. [https://doi.org/10.1175/1520-0493\(1989\)117\(1281:somfst\)2.0.co;2](https://doi.org/10.1175/1520-0493(1989)117(1281:somfst)2.0.co;2)
- 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*(575), 1815–1846. <https://doi.org/10.1002/qj.49712757518>