https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

A methodology to obtain model-error covariances due to the discretization scheme from the parametric Kalman filter perspective

Olivier Pannekoucke[1], Richard Ménard[2], Mohammad El Aabaribaoune[3], and Matthieu Plu[4]

[1]INPT-ENM, CNRM UMR 3589, Météo-France/CNRS, CERFACS, Toulouse, France
[2]ARQI/Air Quality Research Division, Environment and Climate Change Canada, Dorval (Québec), Canada
[3]CNRM UMR 3589, Météo-France/CNRS, CERFACS, Toulouse, France
[4]CNRM UMR 3589, Météo-France/CNRS, Toulouse, France

Correspondence: O. Pannekoucke (olivier.pannekoucke@meteo.fr)

Abstract. This contribution adresses the characterization of the model-error covariance matrix from the new theoretical perspective provided by the parametric Kalman filter method which approximates the covariance dynamics from the parametric evolution of a covariance model. The classical approach to obtain the modified equation of a dynamics is revisited to formulate a parametric diagnosis of the model-error covariance matrix. As an illustration, the particular case of the advection equation is considered as a simple test bed. After the theoretical derivation of both the forecast-error and the predictability-error covariance matrices, a numerical simulation is proposed which demonstrates the skill of the parametric methodology in reproducing the model-error covariance matrix information.

1 Introduction

A significant portion of the work being carried out in state-of-the-art data assimilation concerns the treatment of the forecast-error covariance matrix. The ensemble method has opened up a fantastic playground where variational DA and the Kalman filter are now merged in hybrid or ensemble formulations.

Actually, the forecast-error is composed of two parts. While one part of it is related to the uncertainty in the initial condition, another part is due to the model-error (Daley, 1991; Dee, 1995). The model-error corresponds to the difference between the simulation and the true behavior of a system, and several representations of the model error can be introduced in numerical weather prediction (Houtekamer et al., 2009). For instance, the model error can be related to the misrepresentation of the small-scales and how this influences the large-scales. Stochastic physics such as Stochastic Kinetic Energy Backscatter (Shutts, 2005) or the Stochastically Perturbed Parametrization Tendencies (Palmer et al., 2009) are examples of methods encountered in NWP for this part of the model error.

Although some theoretical studies have been conducted in the past, which elucidate the generic behavior related to the model-error from the dynamical system perspective and in connexion with the data assimilation (*e.g.* Nicolis (2003); Vannitsem

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

and Toth (2002); Carrassi and Vannitsem (2010)), as far as we know there has been little investigation of the effect of the discretization of partial derivative equations on the model error. One reason why the effect of numerical schemes is rarely considered is because it tends to be quite difficult to describe the dynamics of large covariance matrices as encountered in the Kalman filter.

It has been noted in Kalman filtering and EnKF that the propagation of error covariance with a discretized advection model produces a model error (variance) in the form of a variance loss (Ménard et al., 2000, 2020). This error is related to the spatial splitting error in covariance propagation that exists with discretized models and not in continuous propagation of covariance functions.

Recently, Pannekoucke et al. (2016, 2018b) (P16) have proposed to solve the Kalman filter equations using approximated covariance matricies through a covariance model characterized by certain parameters, leading to the so-called *parametric Kalman filter* (PKF). With this approximation, the dynamics of the covariances is replaced by the dynamics of the parameters. For instance, when considering the class of covariance matrices parametrized by the variance field and the local anisotropic tensors (VLATcov), the evolution of the matrices is deduced from the evolution of the variance and the local anisotropic tensors (Cohn, 1993; Pannekoucke, 2020). This approach relies on the partial differential equations encountered in geosciences.

The aims of the present work are to study how the parametric dynamics for covariance matrix evolution can help to characterize the model-error covariance matrix, and more precisely, to determine if is it possible to capture some part of the model-error covariance which is due to the numerical scheme. In this methodological contribution, we will limit ourselves to diffusive numerical-errors whose uncertainty dynamics have been explored by Pannekoucke et al. (2018a) (P18).

The paper is organized as follows, the uncertainty propagation is first reviewed in Section 2 from which the model-error covariance matrix can be deduced at least on a theoretical level. The model-error covariance matrix estimation based on the PKF is detailed for the particular one-dimensional transport equation in Section 3 in the context of the Euler-upwind and semi-Lagrangian schemes. A numerical test bed is proposed in Section 4 to assess the ability of the PKF approach to successfully estimate the model error due to numerical schemes in a one-dimensionnal setting. Conclusions and perspectives are given in the last section, Section 5.

2 Theoretical considerations

2.1 Background in uncertainty propagation and the model error

Here we assume that the *nature* is governed by the deterministic equation

$$\partial_t \mathcal{X} = \mathcal{N}(t, \mathcal{X}), \tag{1}$$

where $\mathcal{X}$ stands for the state. Note that $\mathcal{X}$ can be either discrete or continuous: the discrete case leads to matrix of algebraic relations while the continuous case is suitable for theoretical treatment with partial differential equations. Thereafter, for any state $\mathcal{X}$ of a suitable set, there exists a single trajectory $\mathcal{X}_t = \mathcal{N}_{t\leftarrow 0}(\mathcal{X})$ where $\mathcal{N}_{t\leftarrow 0}$ stands for the propagator of the dynamics

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Eq. (1) from 0 to t. Hence, if $\mathcal{X}_q^t$ denotes the true state of the nature at time t_q, then the true state of the nature at time t_{q+1} is

$$\mathcal{X}_{q+1}^t = \mathcal{N}_{t_{q+1} \leftarrow t_q}(\mathcal{X}_q^t). \tag{2}$$

In practice, the true state $\mathcal{X}_q^t$ is unknown and only an estimation can be deduced from prior informations and the available observations. This estimation is called the *analysis state*, $\mathcal{X}^a$ and it is expanded as

$$\mathcal{X}_q^a = \mathcal{X}_q^t + \varepsilon_q^a, \tag{3}$$

where ε_q^a stands for the so called *analysis error* model as a random field of zero mean and covariance matrix $\mathbf{P}^a = \mathbb{E}\left[\varepsilon_q^a (\varepsilon_q^a)^T\right]$, with $\mathbb{E}[\cdot]$ being the expectation operator. The *forecast state* is the prediction made from the analysis state,

$$\mathcal{X}_{q+1}^f = \mathcal{N}_{t_{q+1} \leftarrow t_q}(\mathcal{X}_q^a). \tag{4}$$

Similarily to the analysis state, the forecast state expands as

$$\mathcal{X}_{q+1}^f = \mathcal{X}_{q+1}^t + \varepsilon_{q+1}^f, \tag{5}$$

where ε_{q+1}^f stands for the so-called *forecast error* modeled as a random field of zero mean and covariance matrix $\mathbf{P}^f = \mathbb{E}\left[\varepsilon_{q+1}^f (\varepsilon_{q+1}^f)^T\right]$.

The forecast error covariance matrix is related to the analysis error covariance matrix through a deterministic relation as follows. From the definition of the forecast error Eq. (5) its dynamics is given by

$$\varepsilon_{q+1}^f = \mathbf{N}_{t_{q+1} \leftarrow t_q, \mathcal{X}_q^a} \varepsilon_q^a,$$

where $\mathbf{N}_{t_{q+1} \leftarrow t_q, \mathcal{X}_q^a}$ stands for the tangent linear (TL) propagator along the nature TL dynamics defined by

$$\partial_t \varepsilon = \mathbf{N}_{t, \mathcal{X}_t^a} \varepsilon, \tag{6}$$

where $\mathbf{N}_{t, \mathcal{X}_t^a} = d\mathcal{N}_{|t, \mathcal{X}_t^a}$ is the differential of $\mathcal{N}$ at $(t, \mathcal{X}_t^a)$, and which governs the evolution of small perturbations along the forecast trajectory starting from the analysis state. Note that the validity of the TL dynamics depends on the error magnitude and on the forecast range. As a consequence, the forecast error covariance matrix becomes

$$\mathbf{P}_{q+1}^f = \mathbf{N}_{t_{q+1} \leftarrow t_q, \mathcal{X}_q^a} \mathbf{P}_q^a \left(\mathbf{N}_{t_{q+1} \leftarrow t_q, \mathcal{X}_q^a}\right)^T. \tag{7}$$

While theoretically correct, the above picture remains a crude idealized shortcut in the realm of numerical predictions. Due to the imperfect knowledge of the nature and the limitations encountered during the computation, the nature dynamics is only approximated by

$$\partial_t \mathcal{X} = \mathcal{M}(t, \mathcal{X}), \tag{8}$$

where $\mathcal{M}$ is the numerical dynamics. Compared with the nature, the time evolution of the true state Eq. (2) is now related to the numerical dynamics as

$$\mathcal{X}_{q+1}^t = \mathcal{M}_{t_{q+1} \leftarrow t_q}(\mathcal{X}_q^t) + \varepsilon_{q+1}^m, \tag{9}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

where ε^m_{q+1} is the *model error*, which can be modeled as a random field $\varepsilon^m_{q+1} = b^m_{q+1} + \eta^m_{q+1}$ field of mean b^m_{q+1} and covariance matrix

$$\mathbf{P}^m_{q+1} = \mathbb{E}\left[\eta^m_{q+1}(\eta^m_{q+1})^T\right] = \mathbb{E}\left[\varepsilon^m_{q+1}(\varepsilon^m_{q+1})^T\right] - b^m_{q+1}(b^m_{q+1})^T. \tag{10}$$

The use of a numerical model in place of the nature requires us to modify the above definitions. Hence, the forecast state Eq. (4) computed by the model is now written as

$$\mathcal{X}^f_{q+1} = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q), \tag{11}$$

while the forecast error evolves as

$$\varepsilon^f_{q+1} = \mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}\varepsilon^a_q - \varepsilon^m_{q+1},$$

where $\mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}$ denotes the propagator of the model TL dynamics

$$\partial_t \varepsilon = \mathbf{M}_{t, \mathcal{X}^a_t}\varepsilon, \tag{12}$$

where $\mathbf{M}_{t,\mathcal{X}^a_t} = d\mathcal{M}_{|t,\mathcal{X}^a_t}$ is the differential of $\mathcal{M}$ at $(t, \mathcal{X}^a_t)$. Hence, the forecast error can be expanded as

$$\varepsilon^f_{q+1} = \varepsilon^p_{q+1} - \varepsilon^m_{q+1}, \tag{13}$$

where

$$\varepsilon^p_{q+1} = \mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}\varepsilon^a_q, \tag{14}$$

is the so-called *predictability error*.

In the general setting where the model error is serially correlated, the predictability error is correlated with the model error, and the forecast-error covariance matrix has the form

$$\mathbf{P}^f_{q+1} = \mathbf{P}^p_{q+1} + \mathbf{P}^m_{q+1} - \mathbf{V}^{pm}_{q+1} - (\mathbf{V}^{pm}_{q+1})^T, \tag{15}$$

where

$$\mathbf{P}^p_{q+1} = \mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}\mathbf{P}^a_q\left(\mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}\right)^T, \tag{16}$$

is the *predictability-error covariance matrix* (Daley, 1992) and $\mathbf{V}^{pm}_{q+1} = \mathbb{E}\left[\varepsilon^p_{q+1}\left(\eta^m_{q+1}\right)^T\right]$ denotes the cross covariance matrix between the predictability error and the model error.

The covariance matrices for the forecast error Eq. (7) and the predictability error Eq. (16) appear more simple than the model-error covariance matrix, since they only rely on TL dynamics along the analysis trajectory while the model error depends on the nature trajectory from the unknown true state. Hence, considering the approximation that the analysis error and the model

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

error are decorrelated (which is not true but it is a common assumption in estimation theory) a practical estimation of the model-error covariance matrix is given by

$$\mathbf{P}^m_{q+1} \approx \mathbf{N}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}\mathbf{P}^a_q\left(\mathbf{N}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}\right)^T - \mathbf{M}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}\mathbf{P}^a_q\left(\mathbf{M}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}\right)^T,$$

that is

$$\mathbf{P}^m_{q+1} \approx \mathbf{P}^f_{q+1} - \mathbf{P}^p_{q+1}. \tag{17}$$

Of course, Eq. (17) is still quite difficult to solve: for the forecast-error covariance matrix, the nature dynamics is either unknown or, when the nature consists in PDEs, no exact solution exists which is easy to handle or interesting enough for applications. For numerical studies, a high order numerical approximation $\widehat{\mathcal{N}}$ of the nature dynamics can be considered, with the hope that the numerical errors are much smaller than those of $\mathcal{M}$. But the major limitation is due to the large size of the numerical state encountered in geophysics: the direct computation of Eq. (7) or Eq. (16) is impossible in practice, even on supercomputers, which are only able to handle a few numerical states at full resolution.

We now consider an alternative for calculating the temporal evolution of the covariance matrices.

2.2 Parametric dynamics for VLATcov models

The parametric formulation of covariance evolution can be stated as follows. If $\mathbf{P}(\mathcal{P})$ denotes a covariance model characterized by a set of parameters $\mathcal{P} = (p_i)_{i\in I}$, then there exists a set $\mathcal{P}^f_t$ ($\mathcal{P}^a$) featuring the forecast (the analysis) error covariance matrix so that $\mathbf{P}(\mathcal{P}^f_t) \approx \mathbf{P}^f_t$ ($\mathbf{P}(\mathcal{P}^a) \approx \mathbf{P}^a$). In reverse, if the dynamics of the parameters $\mathcal{P}^f_t$ is known, then $\mathbf{P}(\mathcal{P}^f_t)$ approximates the dynamics of $\mathbf{P}^f_t$ without using the full matrix computation. This approach constitutes the so-called parametric Kalman filter (PKF) approximation, introduced by Pannekoucke et al. (2016, 2018a) (P16, P18).

The family of covariance models parametrized by the variance field and the local anisotropic tensors, the VLATcov models, are of particular interest (Pannekoucke, 2020): their parameters are directly related to the grid-point statistics of the error field ε. When the error is modeled as an unbiased random differential field, $\mathbb{E}[\varepsilon] = 0$, the variance at a point $\mathbf{x}$ is written

$$V(\mathbf{x}) = \mathbb{E}\left[\varepsilon(\mathbf{x})^2\right]. \tag{18}$$

The anisotropy of the correlation function $\rho(\mathbf{x},\mathbf{y}) = \frac{1}{\sqrt{V_\mathbf{x}V_\mathbf{y}}}\mathbb{E}[\varepsilon(\mathbf{x})\varepsilon(\mathbf{y})]$ is derived, from the second order expansion

$$\rho(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}_\mathbf{x}}, \tag{19}$$

by the local metric tensor $\mathbf{g}(\mathbf{x})$. An interesting result is that the metric tensor can be obtained from the error as

$$\mathbf{g}_{ij}(\mathbf{x}) = \mathbb{E}\left[\partial_{x^i}\left(\frac{\varepsilon}{\sqrt{V}}\right)\partial_{x^j}\left(\frac{\varepsilon}{\sqrt{V}}\right)\right], \tag{20}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

(see *e.g.* (Pannekoucke, 2020) for details). A VLATcov model is then a covariance model parametrized by V and $\mathbf{g}$, that is $\mathbf{P}(V,\mathbf{g})$.

For instance, the diffusion operator of Weaver and Courtier (2001) is an example of a VLATcov model: the local anisotropic tensors are related to the local diffusion tensors, ν, from

$$\nu_{\mathbf{x}} = \frac{1}{2}\mathbf{g}_{\mathbf{x}}^{-1}, \tag{21}$$

where the superscript $^{-1}$ denotes the matrix inverse operator. Eq. (21) holds under the local homogeneous assumption, that is when the spatial derivatives are negligible.

Following Pannekoucke et al. (2018a), the parametric dynamics of a VLATcov model is deduced from the dynamics of the errors from

$$\partial_t V = 2\mathbb{E}\left[\varepsilon\partial_t\varepsilon\right], \tag{22a}$$

$$\partial_t \mathbf{g}_{ij} = \partial_t\left(\mathbb{E}\left[\partial_{x^i}\left(\frac{\varepsilon}{\sqrt{V}}\right)\partial_{x^j}\left(\frac{\varepsilon}{\sqrt{V}}\right)\right]\right), \tag{22b}$$

where the expectation operator and the temporal derivative commute, $\partial_t\mathbb{E}[\cdot] = \mathbb{E}[\partial_t\cdot]$, as used in Eq. (22a). Therefore, the dynamics of the VLATcov model is written $\mathbf{P}(V_t,\mathbf{g}_t)$ or $\mathbf{P}(V_t,\nu_t)$ which are equivalent.

Now, we apply the parametric covariance dynamics for model-error covariance estimation.

2.3 The model-error VLATcov approximation

With the notations of the previous paragraph, a set $\mathcal{P}_t^p$ also exists for the predictability-error covariance matrix leading to the approximation $\mathbf{P}(\mathcal{P}_t^p) \approx \mathbf{P}_t^p$.

If the dynamics of the parameters $\mathcal{P}_t^p$ is known, then starting from the initial condition $\mathcal{P}_0^p = \mathcal{P}^a$ it is possible to approximately determine $\mathbf{P}_t^p$ without solving Eq. (16) explicitly.

Hence, thanks to the parametric dynamics in the case where the nature is known from its partial derivative equation, a new method to compute the model-error covariance matrix can be proposed as follows. By considering the TL dynamics for the nature Eq. (6) and the model Eq. (12), Equation (22) provides a way to compute both the forecast error covariance matrix $\mathbf{P}^f$, Eq. (7), and the predictability error covariance matrix $\mathbf{P}^p$, Eq. (16) ; from which the model-error covariance matrix $\mathbf{P}^m$ can be diagnosed from Eq. (17). For the covariance model based on the diffusion equation, the model-error variance diagnosed from Eq. (17) is the difference

$$V^m = V^f - V^p, \tag{23a}$$

where V^f (V^p) denotes the forecast-error (predictability-error) variance field. The field of the metric tensor of the model-error is approximately given by

$$\mathbf{g}^m = \frac{1}{V^m}\left(V^f\mathbf{g}^f - V^p\mathbf{g}^p\right), \tag{23b}$$

where $\mathbf{g}^f$ ($\mathbf{g}^p$) denotes the forecast-error (predictability-error) metric tensor field (see Appendix A for details).

In the next section we apply the parametric model-error dynamics to a transport equation.

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

3 Parametric characterization of the model error covariance for the one-dimensional advection equation

The transport equation of a passive scalar c by the wind $u(t,x)$ is written as

$$\partial_t c + u\partial_x c = 0, \tag{24}$$

and takes the place of the nature dynamics Eq. (1). Note that dynamics Eq. (24) is linear, meaning that the tangent-linear dynamics is also given by Eq. (24). The advection equation has two aspects. The first side is given by the PDE Eq. (24) which is referred to as the Euler point of view. The other side is the analytico-geometric perspective known as the method of characteristics (see *e.g.* (Boyd, 2001, chap. 14)) where the dynamics can be solved as a local system of ordinary differential equations, given by

$$\frac{dx}{dt} = u, \tag{25a}$$
$$\frac{dc}{dt} = 0. \tag{25b}$$

Each system Eq. (25) describes the evolution of the couple $(x(t), c(t))$ starting from an initial position $x(0)$ where the scalar value is $c(0, x(0))$. At the geometric level, Eq. (25) remains to compute the trajectory of a mobile point of coordinate $x(t)$, the *characteristic curve*, solution of the dynamics Eq. (25a), and transporting the scalar c whose value $c(t)$ coincide with the field value $c(t, x(t))$. The transported value $c(t)$ evolves following Eq. (25b). In the present situation, since the right hand side of Eq. (25b) is null, c is conserved along the curve. This second point of view is referred to as the Lagrangian description for the transport.

Two discretization methods are interesting to study for the transport equation: the finite difference approach and the semi-Lagrangian method resulting from the Lagrangian interpretation of Eq. (24).

The aim of this section is to detail the model-error covariance matrix for both schemes. This theoretical part is organized as follows. The error covariance parametric dynamics for the nature is first described considering the covariance model based on the diffusion equation, then both finite difference and semi-Lagrangian schemes are introduced with their particular parametric dynamics.

3.1 PKF dynamics for the linear advection equation

To describe the time evolution of the forecast error covariance matrix, Eq. (7), it is necessary to detail what is the TL dynamics, Eq. (6), for the linear transport, Eq. (24). Since this transport dynamics is linear, the error evolves according to the same dynamics, and the TL dynamics can be written as

$$\partial_t \varepsilon^f + u\partial_x \varepsilon^f = 0. \tag{26}$$

The PKF approximation of the forecast-error covariance matrix, relies on the dynamics of the variance and of the diffusion fields deduced from Eq. (22). The equation for the variance is computed from Eq. (22a) by replacing the trend by the TL

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

dynamics Eq. (26), so that

$$\partial_t V^f = 2\mathbb{E}\left[\varepsilon^f\left(-u\partial_x\varepsilon^f\right)\right] = -2u\mathbb{E}\left[\varepsilon^f\partial_x\varepsilon^f\right]. \tag{27}$$

From $\partial_x {\varepsilon^f}^2 = 2\varepsilon^f\partial_x\varepsilon^f$ and by the commutativity between the expectation operator and the spatial derivative, the variance dynamics becomes

$$\partial_t V^f = 2\mathbb{E}\left[\varepsilon^f\left(-u\partial_x\varepsilon^f\right)\right] = -u\partial_x\mathbb{E}\left[\left(\varepsilon^f\right)^2\right]. \tag{28}$$

By using the definition of the variance Eq. (18), it results that the dynamics for the variance can be stated as

$$\partial_t V^f = -u\partial_x V^f. \tag{29}$$

The computation of the metric dynamics Eq. (22b) is similar to the above computation made for the variance dynamics, and is detailed in P16 and P18 where the interested reader is referred to. It results that the PKF evolution for the nature is written

$$\partial_t V^f + u\partial_x V^f = 0, \tag{30a}$$

$$\partial_t \nu^f + u\partial_x \nu^f = (2\partial_x u)\nu^f. \tag{30b}$$

Note that a similar system has been first obtained, in data assimilation, by Cohn (1993) (see their Eq. (4.30a) and Eq.(4.34) when written without model error).

From Eq. (30), it results that the variance and the diffusion are independent quantities. The variance is conserved, while it is transported by the wind. The diffusion is not only transported, it is also modified by the source term $(2\partial_x u)\nu^f$ which results from the deformation of correlations by the gradient of the flow u: the diffusion tensor is not conserved by the flow.

Hence, in this sub-section, the forecast-error covariance dynamics Eq. (7) has been computed for the linear transport Eq. (24) and corresponds to the time integration of the un-coupled system Eq. (30) starting from prescribed analysis error variance and diffusion tensor fields.

The finite difference scheme is now considered as a first numerical integration method for Eq. (24), with the derivation of the predictability-error covariance matrix.

3.2 Finite difference scheme and its equivalent PKF dynamics

When the velocity field u is positive (which is assumed from now without loss of generality), a conditionally stable discretization scheme is given by the Euler-upwind scheme,

$$\frac{c_i^{q+1} - c_i^q}{\delta t} = -u_i\frac{c_i^q - c_{i-1}^q}{\delta x}, \tag{31}$$

Stability is assured as long as the CFL condition $\delta x / \underset{x}{\text{Max}}\,|u| < \delta t$ is satisfied. Moreover the scheme is consistent since in the limit of small δx and δt, the dynamics Eq. (24) is recovered from the discrete equation Eq. (31). Thanks to the consistency and the stability, the equivalence theorem of Lax and Richtmyer (1956) assures to the convergence of Eq. (31) toward the true solution. Equation Eq. (31) stands as an illustration of model dynamics Eq. (8).

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

While the numerical solution computed with the aid of a given numerical scheme can converge toward the true solution as $\delta t \to 0$ and $\delta x \to 0$, when δt and δx are of finie amplitude, the numerical solution often differs from the theoretical one. Actually, there exists another partial differential equation which offers a better fit to the numerical solution and highlithgs the properties of the numerical scheme (Hirt, 1968): the consistency, the stability as well as the dissipative and dispersive nature of the numerical scheme can be deduced trom the so-called *modified equation* (Warming and Hyett, 1974). Hence, while it is supposed to solve Eq. (24) the numerical solution computed from Eq. (31) is actually the solution of the modified equation.

More precisely, if $\tilde{c}$ denotes a smooth function solution of the iterations Eq. (31) with $\tilde{c}(q\delta t, i\delta x) = \tilde{c}_i^q$, then the modified equation is the partial differential equation verified by $\tilde{c}$ and at a given order of precision in δt and δx. Here, it is straightforward to show that at order $\mathcal{O}(\delta t^2, \delta x^2)$, the partial differential equation best fitted by $\tilde{c}$ is given by (see Appendix B)

$$\partial_t \tilde{c} + U \partial_x \tilde{c} = \kappa \partial_x^2 \tilde{c}, \tag{32a}$$

where

$$U = u - \frac{\delta t}{2} \partial_t u + \frac{\delta t}{2} u \partial_x u \tag{32b}$$

and

$$\kappa = \frac{u}{2} (\delta x - u \delta t) \tag{32c}$$

are two functions of t and x.

Compared with the nature Eq. (24), the modified equation that best fits the Euler-upwind numerical scheme Eq. (31) presents a correction of the wind which depends on the trend $\partial_t u$ and the self advection $u \partial_x u$ of the wind u. The magnitude of the correction scales as δt and is null at the limit $\delta t \to 0$. But this is not the only modification of the dynamics, as a more critical difference emerges from the numerical discretization: a diffusion term whose magnitude depends on the CFL number $u\delta t/\delta x$. In particular, the diffusion coefficient is negative when the CFL number is larger than one. The diffusion breaks the conservation property of the initial dynamics Eq. (24). This example shows the importance of the modified equation: this provides a way to understand and characterize the defects due to the numerical resolution. In one dimension, for evolution equation, this can be diffusive processes (associated with derivatives of even order) or dispersive processes (associated with derivatives of odd order).

From the PKF point of view, the modified equation is crucial since it converts a discrete dynamics into a partial differential equation, which appeared from P16 and P18, much simpler to handle when considering error covariance dynamics. Thanks to the modified equation Eq. (32), it is now possible to compute the TL evolution of the predictability error for the Euler-upwind scheme, which can be expressed as

$$\partial_t \varepsilon^p + U \partial_x \varepsilon^p = \kappa \partial_x^2 \varepsilon^p. \tag{33}$$

Equations of the PKF forecast can be computed under a similar derivation as in the above Section 3.1. To simplify the computation workflow, a splitting method has been introduced in P16 and P18. Due to the diffusion process appearing in

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Eq. (33), the PKF formulation faces a closure issue for which a closure scheme has been successfully proposed in P18, the *Gaussian closure*. The interested reader is referred to P18 for the details.Note that an alternative to the Gaussian closure can be deduced from the data through machine-learning (Pannekoucke and Fablet, 2020). Hence, the resulting dynamics for the parameter of the predictability-error covariance model is given by

$$\partial_t V^p + U\partial_x V^p = -\frac{V^p\kappa}{\nu^p} + \kappa\partial_x^2 V^p - \frac{\kappa\left(\partial_x V^p\right)^2}{2V^p} \quad (34a)$$

$$\begin{aligned}\partial_t \nu^p + U\partial_x \nu^p = (2\partial_x U)\nu^p + \\ \kappa\partial_x^2\nu^p + 2\kappa - \frac{2\left(\partial_x\nu^p\right)^2}{\nu^p}\kappa + \\ \partial_x\kappa\partial_x\nu^p - \frac{2\partial_x^2 V^p}{V^p}\kappa\nu^p + \\ \frac{\partial_x V^p}{V}\kappa\partial_x\nu^p - \frac{2\partial_x V^p}{V^p}\nu^p\partial_x\kappa + \frac{2\left(\partial_x V^p\right)^2}{{V^p}^2}\kappa\nu^p\end{aligned} \quad (34b)$$

Compared with the PKF dynamics of the nature Eq. (30), the PKF for the Euler-upwind scheme gives rise to additional terms which result from the numerical diffusion of magnitude κ. Moreover, this time, the PKF for the Euler-upwind scheme presents a coupling between the variance and the diffusion, the coupling being a consequence of the numerical diffusion only. Note that a coupling between the variance and the correlation scale also appeared in Eq. (4.30a) and Eq. (4.34) of Cohn (1993), but without a link to the discretization scheme.

The model-error covariance matrix associated with the Euler-upwind scheme can be deduced from the forecast and the predictability error covariance matrix approximations: starting from the initial analysis-error variance and diffusion field, integration of the parametric forecast-error (predictability-error) covariance equation Eq. (30) (Eq. (34)) provides the forecast-error (predictability-error) variance V^f and diffusion ν^f (V^p and ν^p) which are used to compute the model-error covariance parameter Eq. (23).

As another example, the model-error parameters for the semi-Lagrangian scheme are now discussed.

3.3 Semi-Lagrangian scheme and its equivalent PKF dynamics

The modified equation technique has been previously considered for SL schemes. For instance, McCalpin (1988) has shown for the case of constant advecting velocity that a linear interpolation leads to an effective Laplacian dissipation while the quadratic and cubic interpolations lead to a biharmonic dissipation.

Because we want to focus on the method to address the issue of the model error, and since uncertainty prediction of diffusive dynamics has been detailed by P18, we limit the presentation to the linear interpolation in the semi-Lagrangian. and we present the modified equation of Eq. (24) for the study of its model error.

The Lagrangian perspective Eq. (25) of Eq. (24) suggests to build curves along which c is constant. While simple, the drawback of this analytico-geometric method is the possible occurrence of curve trajectory collapses which prevent us from

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

describing the time evolution of c throughout the geographical domain. It is possible to take advantage of the geometrical resolution while avoiding the collapse by considering the so-called semi-Lagrangian procedure.

In the Lagrangian way of thinking, starting from a given position x_o, the question is where the mobile point lies along the time axis, which makes evolving the computation grid forward in time. The semi-Lagrangian perspective reverses this question by asking from which position x_o^* originates the mobile point arriving at x_o at a given time. Hence, the semi-Lagrangian leaves the computation grid unchanged over the time steps of the integration, while letting the scalar field c evolve. More precisely for the particular dynamics of Eq. (24), by assuming the scalar field at time t known for each points of the computational grid, for grid point x_i, the scalar field evolves as

$$c(t+\delta t, x_i) = c(t, x_i^*), \tag{35}$$

where x_i^* is the origin of the trajectory at time t which arrives at x_i at time $t+\delta t$. Since the point of origin x_i^* is unlikely to be a point of the computational grid (except for very particular situations), the value $c(t, x_i^*)$ is computed as an interpolation of the known values of c at time t.

In its present form, the semi-Lagrangian procedure is not suited to the PKF method since it does not give rise any partial differential equation which lies at the core of the parametric approximation for covariance dynamics. To proceed further and to obtain PDEs, additional assumptions are introduced to translate the semi-Lagrangian procedure Eq. (35) into a discrete scheme from which the modified equation is deduced.

In the case where the discretization satisfies the CFL condition $|u(x)|\delta x < \delta t$ and for linear interpolation, it is straightforward to write the semi-Lagrangian procedure Eq. (35) into a discrete scheme (see Appendix C for the details) which is stated as follows:

$$\begin{cases} \frac{c_i^{q+1}-c_i^q}{\delta t} = -u_i \frac{c_i^q - c_{i-1}^q}{\delta x}, & \text{for } u_i > 0 \\ \frac{c_i^{q+1}-c_i^q}{\delta t} = -u_i \frac{c_{i+1}^q - c_i^q}{\delta x}, & \text{for } u_i < 0 \end{cases} \tag{36}$$

which give rise to the Euler-upwind/downwind schemes. Then following the same derivation as previously presented in Section 3.2, the modified equation resulting from the scheme Eq. (36) is given as the PDE verified by a smooth solution $\tilde{c}$ of Eq. (36). From the derivation detailed in Appendix C, the modified equations is

$$\partial_t \tilde{c} + U \partial_x \tilde{c} = \kappa^{\mathrm{SL}} \partial_x^2 \tilde{c}, \tag{37a}$$

where

$$U = u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u \tag{37b}$$

and

$$\kappa^{\mathrm{SL}} = \frac{|u|}{2}(\delta x - |u|\delta t) \tag{37c}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

are both functions of t and x.

Hence, since this corresponds mainly to the modified equation Eq. (32) encountered for the Euler-upwind scheme Eq. (31), the parametric predictability-error covariance is also given by Eq. (34), replacing κ by its SL counterpart value κ^{SL}

Note that the derivation leading to the Euler-upwind and Euler-downwind schemes is due to the choice of the linear interpolation. The bridge between the SL and the Euler-up/down-wind procedures is not a novelty. The derivation has been carried out since it offers an insight into how to build a modified equation for the SL scheme, and also for the self consistency of the presentation. In the general situation, the modified equation for the SL scheme is hard to obtain, if at all possible, and it is not the idea to claim the procedure as universal. But it provides a new insight into the model-error covariance matrix for the SL scheme, which is one of the main goals of the present contribution.

The next section presents the numerical experiments carried out to assess the ability of the PKF to characterize the model-error covariance matrix.

4 Numerical validation

4.1 Setting and illustration

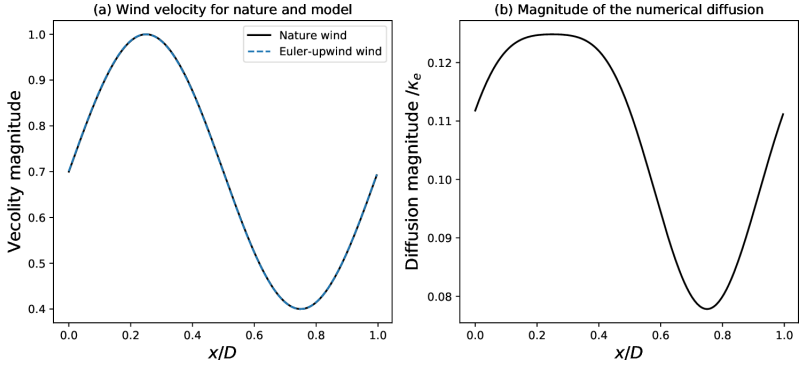

Figure 1. (a) Wind field specified for the nature dynamics and the one seen in the discretized model from Eq. (32b). Panel (b) represents the numerical diffusion coefficient due to the discretization Eq. (32c).

In this experimental test bed, the domain is assumed to be the one dimensional segment $[0, D)$ with periodic boundary conditions, where $D = 1$. The domain is discretized into a regular grid of $n = 241$ points $x_i = i\delta x$ for $i \in [0, 240]$ and $\delta x = D/n \approx 4.1\,10^{-3}$.

The wind field u for the one-dimensional transport Eq. (24) is set as the stationary field

$$u(x) = 0.4 + \frac{0.6}{2}\left(1 + \cos\left(\frac{2\pi}{D}(x - D/4)\right)\right), \tag{38}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

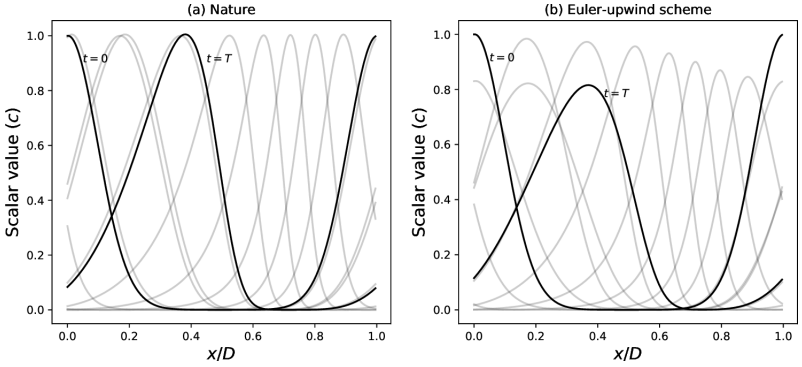

Figure 2. Nature (a) and model (b) runs for times from $t = 0$ to $t = T$ and represented each $0.2T$.

showed in Fig. 1-(a), which appears as a jet with the entrance (exit) at $x = 0.75D$ ($x = 0.25D$): the flow accelerates (decelerates) until $x = 0.25D$ ($x = 0.75D$).

In order to verify the CFL condition, the time step for the numerical simulation is set to $\delta t = 0.002$ leading to a CFL value of $0.48 < 1$. The magnitude of the numerical diffusion κ, Eq. (32c), associated with this setting is shown in Fig. 1-(b), normalized by the diffusion coefficient $\kappa_e = \delta x^2/\delta t$.

For the numerical experiment, the initial state for c is set to

$$c(0,x) = \exp\left(-\frac{1}{2(0.15D)^2}\sin\left(\frac{\pi}{2}(x - D/2)\right)^2\right) \tag{39}$$

while the initial analysis-error covariance matrix is set as the homogeneous Gaussian covariance matrix $\mathbf{P}^f_{t=0}(x,y) = e^{-\frac{(x-y)^2}{2l_h^2}}$ where $l_h = 0.05 \approx 12\delta x$. The analysis-error standard-deviation is set to the homogeneous value 1.0.

For numerical validation, since no simple analytical solution of the partial differential equation Eq. (24) exists, this dynamics is integrated considering a fourth order Runge-Kutta time scheme applied on the finite difference discretization

$$\partial_t c_i = -u_i \frac{c_{i+1} - c_{i-1}}{2\delta x}, \tag{40}$$

where the spatial derivative is approximated by a centered second order scheme. This constitutes the higher resolution version $\widehat{\mathcal{N}}$ of the model $\mathcal{N}$ and introduced in Section 2.1: $\widehat{\mathcal{N}}$ is assumed to better reproduce the nature $\mathcal{N}$. In order to compute the true covariance dynamics, a very large ensemble of forecast, integrated with $\widehat{\mathcal{N}}$, has been considered with $N_e = 6400$ forecasts. This large size limits the sampling noise to a relative error of $1/\sqrt{6400} \approx 1.25\%$.

Figure 2 shows the trajectory computed from the nature approximated by $\widehat{\mathcal{N}}$ and the nature $\mathcal{N}$. Since the transport equation conserves the value of the field c, the extremal values of c do not change along the integration and the wind $u > 0$ causes the initial structure to move to the right. While the field is conserved, it is also deformed by the wind. For the particular choice of the initial condition made here, the signal is of larger (smaller) scale in the region $x \in [0, 0.5]$ ($x \in [0.5, 1]$) than its initial shape. Panel (a) shows that the nature approximation $\widehat{\mathcal{N}}$ is able to reproduce the conservation of c as well as the stretching of

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

the signal along the time axis. Hence, the nature approximation $\widehat{\mathcal{N}}$ is good enough to capture the main features of the nature dynamics, which justifies the use of this approximation in place of the true dynamics in the following. At the opposite, the model $\mathcal{N}$ fails to maintain the magnitudes of the extrema (panel (b)), in accordance with the modified equation Eq. (32a) of the Euler-upwind Eq. (31) which presents a non-physical diffusion process resulting from the numerical discretization. Note that the coefficient of the numerical diffusion is heterogeneous over the domain with a typical value of thereabout $0.1\kappa_e$ (see Fig. 1-(b)). This heterogeneity is due to the scale variation of the signal, stretched by the wind shear: when the signal is of smaller (larger) scale than its initial shape, the second order derivative is larger (smaller), which leads to an intensification (reduction) in the numerical diffusion term in Eq. (32a).

Having validated the two numerical models, it is now possible to look at the covariance dynamics and how the model-error covariance error can be estimated from the PKF prediction.

4.2 Assessment of the PKF in predicting the forecast-error and the predicability-error covariance dynamics

The PKF forecast-error covariance matrix dynamics for the transport equation Eq. (24) is given by the system Eq. (30). The PKF predictability-error covariance matrix dynamics resulting from the Euler-upwind integration Eq. (31) is given by Eq. (34). Both systems are numerically integrated by considering respectively an explicit RK4 time scheme for the nature and an Euler time scheme for the Euler-upwind scheme. The time step used for the integration is $\delta t = 0.002$. The forecast-error and the predictability-error variance field are shown in Fig. 3. The forecast-error (predictability-error) correlation length-scale field, defined from the one-dimensional diffusion field by $L^f = \sqrt{2\nu^f}$, ($L^p = \sqrt{2\nu^p}$), is shown in Fig. 4. The variance and the length-scale, are shown for the PKF and the ensemble estimation, the latter being only computed for the validation of the PKF (the ensembles are not needed neither used for the computation of the PKF systems).

The forecast-error covariance dynamics is first considered. Since the variance of the nature Eq. (30a) is conserved, it results that with the choice of an initial homogeneous variance, the trend is null and the variance field is the stationary homogeneous field 1.0. This theoretical result is well reproduced in Fig. 3-(a) from the PKF integration while the ensemble estimation, Fig. 3-(c) also shows this stationary but to within the sampling noise. The length-scale (Fig. 4-(a)) shows a periodic evolution where, starting from the homogeneous field of L_h, the length-scale first increases (decreases) in the entrance (exit) of the jet, then these evolutions are attenuated then compensated with the transport. Then ensemble estimation Fig. 4-(c) presents the same variations (again to within the sampling noise), which validates the PKF dynamics for the nature.

The predictability-error covariance dynamics is now discussed. For the Euler-upwind scheme, the numerical diffusion resulting from the spatio-temporal discretization in Eq. (32a) implies a damping of the variance along the time axis (see Fig. 3-(b)). The attenuation of the uncertainty governed by Eq. (34), leads to a heterogeneous damping over the domain and appears much stronger in the middle of the domain ($x = 0.5$) than near the boundaries ($x = 0$ and $x = 1$), while transported by the flow. The length-scale, Fig. 4-(b) increases by the diffusion while the shear produces similar patterns as for the forecast-error statistics. The ensemble estimation in Fig. 3-(d) and Fig. 4-(d) shows the same signal as the PKF prediction (within the sampling noise) which validates the system Eq. (34).

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

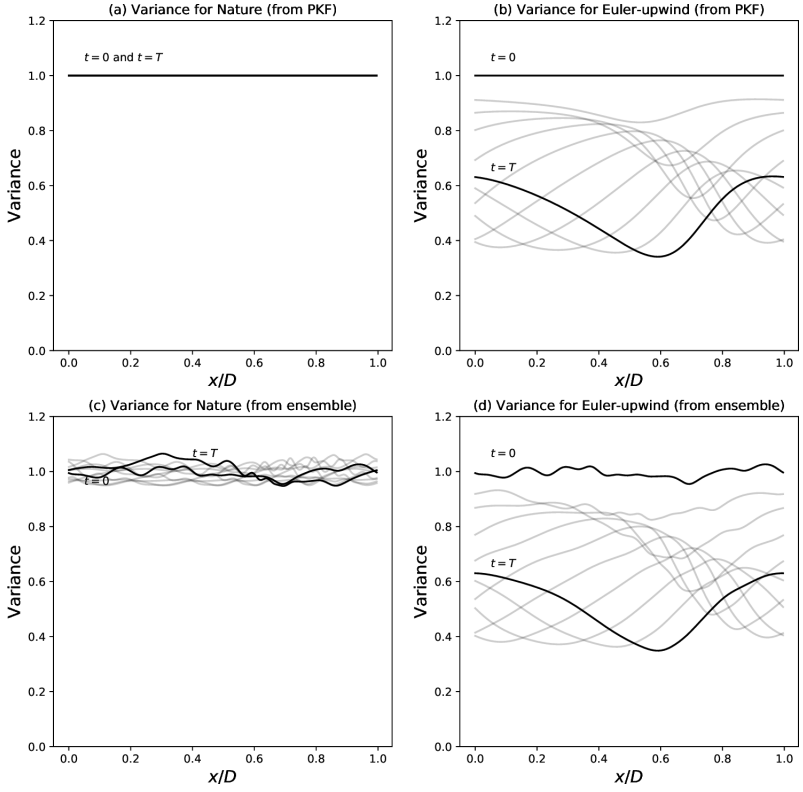

Figure 3. Forecast-error variance field, $V^f(t,x)$, for the nature Eq. (24), computed from the PKF Eq. (30) (panel a), and predicatbility-error variance field, $V^p(t,x)$, for the numerical model resulting from the finite difference & Euler discretization Eq. (31), computed from the PKF Eq. (34) (panel b). Panels (c) and (d) are the ensemble estimation for panels (a) and (b), where the nature dynamics is approximated by Eq. (40) dynamics in panel (c) (6400 members are used here). Fields are represented for times from $t = 0$ to $t = T$ and represented each $0.2T$.

As a conclusion of this section, the PKF is able to predict the variance and the legnth-scale feature of the forecast-error covariance dynamics of the nature dynamics Eq. (24) and of the predictability-error covariance dynamics resulting from the discretization of the true dynamics given by Eq. (31). These results are now considered to provide an estimation of the model-error covariances.

4.3 Model-error covariance diagnosis from the PKF prediction

As discussed in Section 2.1, the model-error covariance matrix can be estimated from the difference between the forecast-error and the predictability-error covariance matrix, Eq. (17), when the analysis error and the model error are decorrelated. This latter assumption is quite restrictive since in real applications such a correlation certainly exists, leading to the much more complex

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

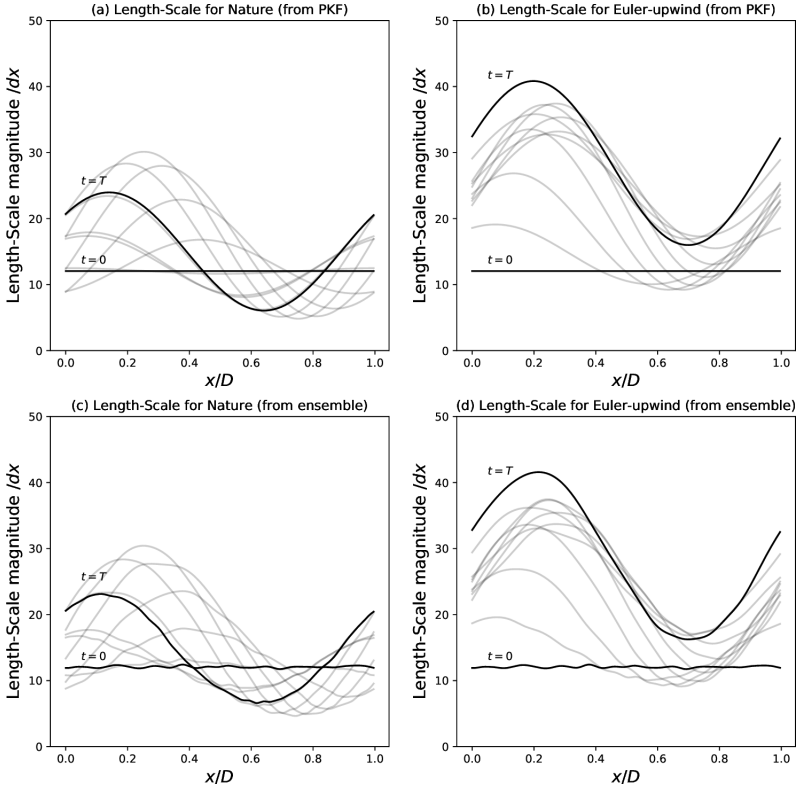

Figure 4. The length-scale counterpart of Fig. 3 representing the forecast-error (the predicatbility-error) length-scale field L^f (L^p) in panels (a,c) (in panels (b-d)). The length-scale is diagnosed from the diffusion coefficient ν^f (ν^p) as $L^f = \sqrt{2\nu^f}$ ($L^p = \sqrt{2\nu^p}$) and normalized by the grid spacing δx. Top panels are computed from the PKF while the bottom panels are estimated from the same large ensemble of forecasts as considered in Fig. 3. Fields are represented for times from $t = 0$ to $t = T$ and represented each $0.2T$

relationship Eq. (15). The question is to know if it is possible to capture some information about the model-error covariance from the approximation Eq. (17).

In order to tackle this issue, the computation of $\mathbf{P}^m{}_t$ is first made from Eq. (17), considering the PKF estimation of the forecast-error covariance $\mathbf{P}^f_t$ and of the predictability-error covariance $\mathbf{P}^p_t$, following Eq. (23). Then this estimation of $\mathbf{P}^m{}_t$, is compared to the direct ensemble estimation of the model-error covariance computed from the model error computation $\varepsilon^m_k = \mathcal{X}^f_k - \mathcal{X}^p_k$, which is made possible here since the nature is known (to within the approximation of $\mathcal{N}$ by $\widehat{\mathcal{N}}$). The comparison of both estimates of the model-error covariance is made by looking at the model-error variance and length-scale fields.

Figure 5 shows the model-error variance field V^m_t approximated by the computation along the simulation as the difference $V^f_t - V^p_t$ (panel a), and estimated from the ensemble of model error $\varepsilon^m_k = \mathcal{X}^f_k - \mathcal{X}^p_k$ (panel b) – this estimation being the approximation of the true model-error statistics. Both estimates of the model-error variance predict a variance bump that increases and translates over time. The location of the bump predicted from the PKF estimation (panel a) is in accordance

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

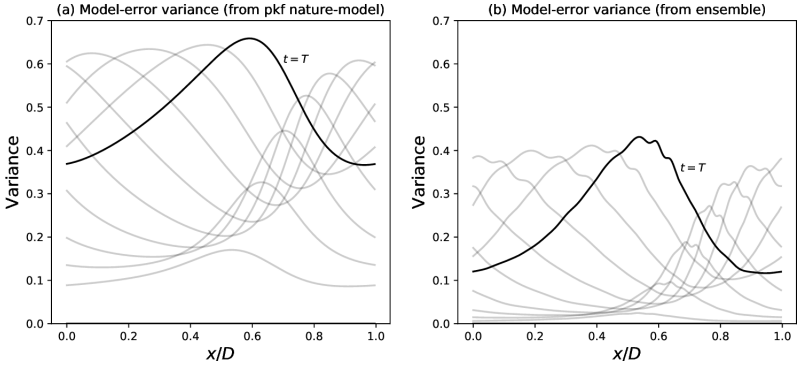

Figure 5. Model-error variance: (a) computed under decorrelation assumption as the difference between the variance computed from the PKF and shown in Fig. 3(a) and (b); (b) diagnosed from ensemble of differences $\varepsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$. Fields are represented for times from $t = 0.2T$ to $t = T$ and represented each $0.2T$ (for $t = 0$ the model error is null).

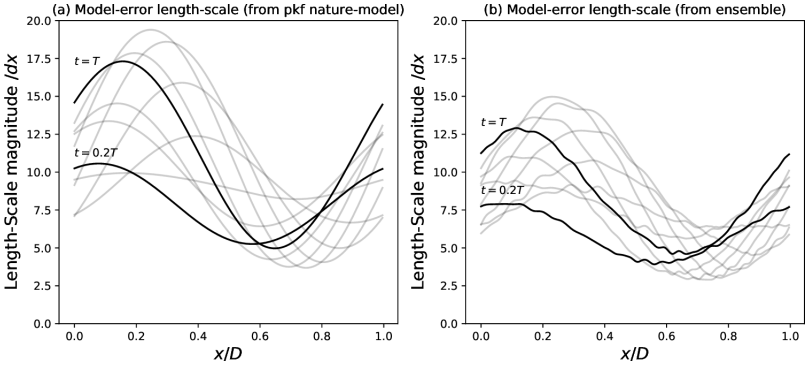

Figure 6. Model-error length-scale: (a) under decorrelation of errors as computed from the PKF for the nature and the Euler-upwind scheme, (b) diagnosed from the ensemble of differences $\varepsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$. Fields are represented for times from $t = 0.2T$ to $t = T$, each $0.2T$ (for $t = 0$ the model error is null)

with the true statistics (panel b). But the magnitude of the variance differs: under the analysis-error/model-error decorrelation assumption, the variance averaged over the domain (panel a) increases much faster than for the true model-error variance (panel b).

The length-scale diagnosis, shown in Figure 6, illustrates a similar behaviour. At $t = 0$, the length-scale is not determined since the model-error covariance matrix is null for both simulations, so $t = 0.2T$ is represented in replacement. The length-scale translates with the flow and increases or decreases due to the velocity shear along the domain. Despite their similarities, the decorrelation assumption (panel a) produces larger length-scales than for the real model-error length-scale statistics, estimated from the ensemble of simulations (panel b).

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

In this simulation, the parametric Kalman filter appears as a theoretical tool able to investigate the model-error statistics by providing some estimation of the variance field and the length-scale field.

5 Conclusions

In this contribution, the part of the model-error covariance due to the spatio-temporal discretization scheme is explored by considering the parametric approximation for the Kalman filter. In the PKF formulation, covariances are approximated by covariance models which are characterized by a set of parameters. We focused on the class of covariance model distinguished by the variance field and the local anisotropic tensors (VLATcov). Therefore, for VLATcov matrices, the covariance dynamics is given by the dynamics of the variance and the local anisotropic tensors, whose dynamics are deduced from the partial differential equations of the system.

Under the decorrelation assumption of the analysis error and the model error, the model-error covariance is approximated as the difference between the parametric approximation of the forecast-error and the predictability-error covariance matrices. For a dynamics given by a partial differential equation, the parametric forecast-error covariance matrix is deduced from the evolution equation while the predictability error covariance matrix is computed from the modified evolution, *i.e.* the partial differential equation that best fits the numerical solution.

The ability of the parametric approach to characterize part of the model-error covariance dynamics has been illustrated in a numerical test bed in 1D. We have considered the transport of a scalar by a heterogeneous velocity field. In this case, the parametric dynamics of the foreacast error shows that the variance is conserved along the flow, while the local anisotropic tensor is transported by the flow and deformed by the gradient of the velocity.

For this transport dynamics, two numerical schemes have been considered: an Euler-upwind scheme and a semi-Lagrangian scheme in the case of a linear interpolation. The modified equations of both schemes make appear an additional heterogeneous dissipation and a perturbation of the velocity, whose characteristics depend on the spatio-temporal discretisation (dt, dx), the trend and the shear of the flow. Because of the numerical diffusion, the variance of the predictability error is not conserved and a coupling with the anisotropy appears. This effect has been noted as well in 3D global transport models (Ménard et al., 2020) where the loss of error variance is stronger for short correlation length-scales.

An ensemble of forecasts has been introduced, taken as the reference, to compare the true covariance evolution with the parametric approximation. The numerical experiment shows the ability of the parametric dynamics to reproduce the forecast-error and the predictability-error covariance dynamics. For the model-error covariance, the difference between the forecast-error and the predictability-error covariances was not able to perfectly recover the true model-error covariance dynamics: this is due to the cross-correlation between the predictability error and model error that is not taken into account here. Nonetheless, the model-error variance and local anisotropy obtained from the PKF shared some similarities with the ensemble estimation, which indicates that the PKF approximation can provide an estimation of some model-error covariance characteristics.

The methodology introduced here has shown the potential of exploring the model-error covariance from the parametric dynamics. While the characterization of the model-error covariance is a challenge, as in air quality forecasts (Emili et al.,

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

445 2016), the parametric approach appears as a new theoretical tool to tackle this issue. In order to represent the uncertainty of the small scales, it would be interesting to combine the parametric approach with other new methods *e.g.* the modelling under location uncertainty (Resseguier et al., 2017).

However, the parametric dynamics faces closure issues that have to be adressed depending on applications. Here, the investigation of diffusive model errors has been made possible thanks to the Gaussian closure of P18. For other kind of numerical
450 errors, an appropriate closure will have to be specified, either from theoretical closures or from the data as suggested by the data-driven and physics-informed identification of uncertainty dynamics of Pannekoucke and Fablet (2020).

Appendix A: Approximation of the model-error metric tensor field

Here, we consider the particular case where the model-error covariance model is approximated under the white noise assumption as Eq. (17) *i.e.*

$$\mathbf{P}^m = \mathbf{P}^f - \mathbf{P}^p.$$

The local metric tensor can be diagnosed from the Taylor expansion of the model-error correlation function

460 $$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) = \frac{1}{\sqrt{\mathbf{P}^m(\mathbf{x},\mathbf{x})\mathbf{P}^m(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})}} \left(\mathbf{P}^f(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) - \mathbf{P}^p(\mathbf{x},\mathbf{x}+\delta\mathbf{x})\right). \quad \text{(A1)}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Under an assumption of local homogeneity of the variance, $\mathbf{P}^m(\mathbf{x},\mathbf{x}) \approx \mathbf{P}^m(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$, $\mathbf{P}^f(\mathbf{x},\mathbf{x}) \approx \mathbf{P}^f(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$, and $\mathbf{P}^p(\mathbf{x},\mathbf{x}) \approx \mathbf{P}^p(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$, which leads to the expansion

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx \frac{\mathbf{P}^f(\mathbf{x},\mathbf{x})}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(1-\frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^f_\mathbf{x}}\right) - \frac{\mathbf{P}^p(\mathbf{x},\mathbf{x})}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(1-\frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^p_\mathbf{x}}\right). \quad \text{(A2)}$$

Since, $||\delta\mathbf{x}||^2_{\mathbf{g}_\mathbf{x}} = \delta\mathbf{x}^T\mathbf{g}_\mathbf{x}\delta\mathbf{x}$, the correlation is expanded as

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}\delta\mathbf{x}^T\left[\frac{1}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(\mathbf{P}^f(\mathbf{x},\mathbf{x})\mathbf{g}^f_\mathbf{x} - \mathbf{P}^p(\mathbf{x},\mathbf{x})\mathbf{g}^p_\mathbf{x}\right)\right]\delta\mathbf{x}. \quad \text{(A3)}$$

After identification with the expected form of the expansion

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^m_\mathbf{x}}, \quad \text{(A4)}$$

it follows that

$$\mathbf{g}^m_\mathbf{x} = \frac{1}{V^f(\mathbf{x}) - V^p(\mathbf{x})}\left(V^f(\mathbf{x})\mathbf{g}^f_\mathbf{x} - V^p(\mathbf{x})\mathbf{g}^p_\mathbf{x}\right), \quad \text{(A5)}$$

where the variance are denoted by $\mathbf{P}^f(\mathbf{x},\mathbf{x}) = V^f(\mathbf{x})$ and $\mathbf{P}^p(\mathbf{x},\mathbf{x}) = V^p(\mathbf{x})$.

Appendix B: Computation of the modified equation for Euler scheme

The modified partial differential equation associated with the numerical scheme Eq. (31) is the partial differential equation of a smooth function $\tilde{c}$, solution of the scheme, so that $\tilde{c}(q\delta t, i\delta x) = \tilde{c}^q_i$ *i.e.*

$$\frac{\tilde{c}^{q+1}_i - \tilde{c}^q_i}{\delta t} = -u_i\frac{\tilde{c}^q_i - \tilde{c}^q_{i-1}}{\delta x}, \quad \text{(B1)}$$

for which the Taylor formula in time and space at order $\mathcal{O}(\delta t^2, \delta x^2)$ is

$$\partial_t\tilde{c} + \frac{\delta t}{2}\partial^2_t\tilde{c} + \mathcal{O}(\delta t^2) = -u\left(\partial_x\tilde{c} - \frac{\delta x}{2}\partial^2_x\tilde{c} + \mathcal{O}(\delta x^2)\right) \quad \text{(B2)}$$

The second order time derivation can be replaced from the equation Eq. (B2) itself, at an appropriate order. Due to the δt, an expansion at order $\mathcal{O}(\delta t)$ only requires to express the second order derivative at the lead order, that is from

$$\partial_t\tilde{c} = -u\partial_x\tilde{c} + \mathcal{O}(\delta t, \delta x). \quad \text{(B3)}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Then, from the time derivation, the second order derivative can be replaced by

$$\begin{aligned}\partial_t^2\tilde{c} &= \partial_t\left(-u\partial_x\tilde{c}\right)+\mathcal{O}(\delta t,\delta x),\\ &= -\partial_t u\partial_x\tilde{c}-u\partial_{xt}^2\tilde{c}+\mathcal{O}(\delta t,\delta x),\end{aligned}$$

then, the second order derivative $\partial_{xt}^2\tilde{c}$ can be deduced from spatial derivation of Eq. (B3), and writes

$$\begin{aligned}\partial_{xt}^2\tilde{c} &= -\partial_x(u\partial_x\tilde{c})+\mathcal{O}(\delta t,\delta x),\\ &= -\partial_x u\partial_x\tilde{c}+u\partial_x^2\tilde{c}+\mathcal{O}(\delta t,\delta x).\end{aligned}$$

It results that Eq. (B2) writes

$$\partial_t\tilde{c}+\frac{\delta t}{2}\left[-\partial_t u\partial_x\tilde{c}-u\left(-\partial_x u\partial_x\tilde{c}+u\partial_x^2\tilde{c}\right)\right]= -u\left(\partial_x\tilde{c}-\frac{\delta x}{2}\partial_x^2\tilde{c}\right)+\mathcal{O}(\delta t^2,\delta x^2)$$

then

$$\partial_t\tilde{c}+U\partial_x\tilde{c}=\kappa\partial_x^2\tilde{c}+\mathcal{O}(\delta t^2,\delta x^2), \tag{B4}$$

where $U=u-\frac{\delta t}{2}\partial_t u+\frac{\delta t}{2}u\partial_x u$ and $\kappa=\frac{u}{2}\left(\delta x-u\delta t\right)$ are two functions of t and x.

Appendix C: Computation of the modified equation for Semi-Lagrangian scheme

The aims of this section is twofold, the first goal is to obtain a discrete scheme from the semi-Lagrangian procedure, then to deduce the modified equation of the discrete scheme.

For the sake of simplicity, the linear advection dynamics $\partial_t c+u\partial_x c=0$ is first considered with a velocity $u>0$.

From the characteristic curve resolution it follows that $c(t_{q+1},x_i)=c(t_q,x_i^*)$, where the originate point x_i^* is assumed in between points x_{i-1} and x_i, which means that the CFL constraint $u\delta t<\delta x$ is verified. This originate point can be approximated as $x_i^*=x_i-u_i\delta t$, and if a linear interpolation is considered for the computation of $c(t,x_i^*)$, it follows that

$$\begin{aligned}&c(t_q,x_i^*)=\\ &\left(1-\frac{x_i^*-x_{i-1}}{x_i-x_{ki-1}}\right)c_{i-1}^q+\left(\frac{x_i^*-x_{i-1}}{x_i-x_{i-1}}\right)c_i^q\\ &= \frac{u_i\delta t}{\delta x}c_{i-1}^q+\left(1-\frac{u_i\delta t}{\delta x}\right)c_i^q,\end{aligned} \tag{C1}$$

Hence, the numerical scheme writes

$$c_{q+1,i}=\frac{u_i\delta t}{\delta x}c_{q,i-1}+\left(1-\frac{u_i\delta t}{\delta x}\right)c_{q,i}. \tag{C2}$$

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

The modified differential equation is obtained by replacing c by a smooth function $\tilde{c}$, solution of the numerical scheme Eq. (C2). The computation of the modified equation is similar to the Euler case detailed in Appendix B, leading to

$$\partial_t \tilde{c} + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right)\partial_x \tilde{c} = \left(\frac{1}{2}u\delta x - \frac{1}{2}u^2\delta t\right)\partial_x^2 \tilde{c}. \quad \text{(C3)}$$

When $u < 0$, the differential equation writes

$$\partial_t \tilde{c} + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right)\partial_x \tilde{c} = \left(\frac{1}{2}(-u)\delta x - \frac{1}{2}u^2\delta t\right)\partial_x^2 \tilde{c} \quad \text{(C4)}$$

Hence, in the general situation,

$$\partial_t \tilde{c} + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right)\partial_x \tilde{c} = \left(\frac{1}{2}|u|\delta x - \frac{1}{2}u^2\delta t\right)\partial_x^2 \tilde{c}, \quad \text{(C5)}$$

whatever the sign of the velocity u.

Author contributions. OP and RM imagined to explore the influence of the numerical scheme on the error model. OP linked the modified equation to the parametric formulation of the uncertainty prediction. MEA contributed to the simulation during its training period, supervised by OP and MP.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. TEXT

Acknowledgements. We would like to thank Matheusz Reszka for proofreading the English. This work was supported by the KAPA project (Kalman Parametrique), funded by the French LEFE INSU program.


https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

References

Boyd, J.: Chebyshev and Fourier Spectral Methods, Dover Publications, second edn., 2001.

Carrassi, A. and Vannitsem, S.: Accounting for Model Error in Variational Data Assimilation: A Deterministic Formulation, Monthly Weather Review, 138, 3369–3386, https://doi.org/10.1175/2010MWR3192.1, http://dx.doi.org/10.1175/2010MWR3192.1, 2010.

Cohn, S.: Dynamics of Short-Term Univariate Forecast Error Covariances, Monthly Weather Review, 121, 3123–3149, 1993.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, 1991.

Daley, R.: Estimating Model-Error Covariances for Application to Atmospheric Data Assimilation, Monthly Weather Review, 120, 1735–1746, 1992.

Dee, D.: On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation, Monthly Weather Review, 123, 1128–1145, 1995.

Emili, E., Gürol, S., and Cariolle, D.: Accounting for model error in air quality forecasts: an application of 4DEnVar to the assimilation of atmospheric composition using QG-Chem 1.0, Geoscientific Model Development, 9, 3933–3959, https://doi.org/10.5194/gmd-9-3933-2016, http://www.geosci-model-dev.net/9/3933/2016/, 2016.

Hirt, C.: Heuristic stability theory for finite-difference equations, Journal of Computational Physics, 2, 339–355, https://doi.org/10.1016/0021-9991(68)90041-7, 1968.

Houtekamer, P. L., Mitchell, H. L., and Deng, X.: Model Error Representation in an Operational Ensemble Kalman Filter, Monthly Weather Review, 137, 2126–2143, 2009.

Lax, P. D. and Richtmyer, R. D.: Survey of the stability of linear finite difference equations, Communications on Pure and Applied Mathematics, 9, 267–293, https://doi.org/10.1002/cpa.3160090206, https://doi.org/10.1002/cpa.3160090206, 1956.

McCalpin, J. D.: A Quantitative Analysis of the Dissipation Inherent in Semi-Lagrangian Advection, Mon. Wea. Rev., 116, 2330–2336, 1988.

Ménard, R., Cohn, S., Chang, L.-P., and Lyster, P. M.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation, Monthly Weather Review, 128, 2654–2671, 2000.

Ménard, R., Skachko, S., and Pannekoucke, O.: Spatial splitting error in covariance propagation: The need for inflation (to be submitted), Monthly Weather Review, 2020.

Nicolis, C.: Dynamics of Model Error: Some Generic Features, Journal Atmospheric Sciences, 60, 2208–2218, 2003.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., and Weisheimer, A.: Stochastic Parametrization and Model Uncertainty, Tech Memo 598, ECMWF, 2009.

Pannekoucke, O.: An anisotropic formulation of the parametric Kalman filter analysis step, Quaterly Journal Royal Meteorological Society, 2020.

Pannekoucke, O. and Fablet, R.: PDE-NetGen 1.0: from symbolic PDE representations of physical processes to trainable neural network representations, Geoscientific Model Development Discussions, https://doi.org/https://doi.org/10.5194/gmd-2020-35, 2020.

Pannekoucke, O., Ricci, S., Barthelemy, S., Ménard, R., and Thual, O.: Parametric Kalman Filter for chemical transport model, Tellus, 68, 31 547, https://doi.org/10.3402/tellusa.v68.31547, 2016.

Pannekoucke, O., Bocquet, M., and Ménard, R.: Parametric covariance dynamics for the nonlinear diffusive Burgers' equation, Nonlinear Processes in Geophysics, 2018, 1–21, https://doi.org/10.5194/npg-2018-10, https://www.nonlin-processes-geophys-discuss.net/npg-2018-10/, 2018a.

https://doi.org/10.5194/npg-2020-14
Preprint. Discussion started: 8 May 2020

Pannekoucke, O., Ricci, S., Barthelemy, S., Ménard, R., and Thual., O.: Parametric Kalman filter for chemical transport models - Corrigendum, Tellus A: Dynamic Meteorology and Oceanography, 70, 1–2, https://doi.org/10.1080/16000870.2018.1472954, 2018b.

Resseguier, V., Mémin, E., and Chapron, B.: Geophysical flows under location uncertainty, Part I Random transport and general models, Geophysical & Astrophysical Fluid Dynamics, 111, 149–176, https://doi.org/10.1080/03091929.2017.1310210, 2017.

Shutts, G. J.: A kinetic energy backscatter algorithm for use in ensemble prediction systems, Quarterly Journal of the Royal Meteorological Society, 131, 3079–3102, http://onlinelibrary.wiley.com/doi/10.1256/qj.04.106/pdf, 2005.

Vannitsem, S. and Toth, Z.: Short-Term Dynamics of Model Errors, Journal Atmospheric Sciences, 59, 2594–2604, https://doi.org/10.1175/1520-0469(2002)059<2594:STDOME>2.0.CO;2, http://dx.doi.org/10.1175/1520-0469(2002)059<2594:STDOME>2.0.CO;2, 2002.

Warming, R. and Hyett, B.: The modified equation approach to the stability and accuracy analysis of finite-difference methods, Journal of Computational Physics, 14, 159–179, https://doi.org/10.1016/0021-9991(74)90011-4, 1974.

Weaver, A. and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation (Tech. Memo. ECMWF, num. 306), Quarterly Journal of the Royal Meteorological Society, 127, 1815–1846, 2001.