Final answer to the referee 3

First of all, we would like to thank the referee for her/his review on our paper and for giving us the opportunity to improve our paper.

Now, we organized the answer to the comments as follows. First, we list some changes afford to the manuscript then detail our answers to the questions raised by the referee.

List of changes for the revision

Major changes

The formalism has been reworded to match the usual notation of data assimilation (Sec. 2). In particular Sec. 2.1 has been split so to better introduce the modelling of the model-error covariance matrix that is now presented in Sec. 2.2. To this end, a new figure 1 has been introduced to sketch the dynamics of the uncertainty in presence of model error and to facilitate the setting of the framework considered here: the case where the numerical model is (more) dissipative compared to the nature. The formalism of the model error has been clarified, with the details of formulations of the forecast error that are given in Appendix A. A specific notation has been introduced for the predictability error of the nature ($\widetilde{\eps}^p$) so to avoid any confusion with the forecast error as usually encountered in data assimilation. Now, every $\widetilde{.}$ symbols refer to the uncertainty of the nature.

Sec 4.2 has been split in two part, one for the nature, the other for the model. The dynamics of the average over the model has been introduced to quantity the evolution of the variance and the length-scale fields. In particular, the Sec. 4.2.2 provides an analytical evolution of the predictability-error variance of the model which better explore the PKF equations. This quantitative evolution of the predictability-error variance is also now explored in section 4.3 for the model-error variance.

A section 5 has been introduced to discuss the results and the connexion with the previous work of Ménard et al 2000s that support the interest of the method introduced here, but in a real data assimilation system.

Minor changes

The detail of the ensemble estimation was missing in the previous version of the manuscript, this has been introduced in the introduction of Sec. 4.2.

There was an error on the definition of the time representation of Fig. 3 and later where the results are shown every $0.1T$, moreover the lead time of the simulation has been indicated $T=2.0$ (this was missing).

P10, line 12: The chordal distance was used in the numerical simulation, but we wrote the distance as |x-y| in the manuscript. While there is no difference between the two distances with the value of D and l_h used here, the chordal distance leads to a valid correlation function on the circle (that is not the case with the previous distance – this theoretical consideration is not discussed in the manuscript but justifications can be found in the article Pannekoucke et al. 2018).

Differences between the two version of the manuscript

To facilitate the comparison between the two version of the manuscript, a companion version of the manuscript lists all the modifications where old (new) statements are in red (blue).

Answer to the question of the referee

We copied your commentary in italics below, we reply in normal blue font.

1. "Equations throughout the discussion version are poorly formatted and awkwardly split, this should be fixed for better readability. This could also use hyperlinking in equation references to make the document more easily searchable."

We send the new version of the manuscript as two-column which is better presented.

2. "I agree with the other review comments that the definitions of quantities such as the model error $\eps^m_ q+1$, and their evolution equations, need to be more clear."

The new version of the manuscript clarify the presentation of the data assimilation so to avoid any confusion in Sec 2.1. Moreover, the introduction of the modelling of the flow-dependent of the model-error covariance matrix is now clarified in the new section 2.2.

Moreover, the new appendix A provides the derivation of the decomposition of the forecast error which contributes to clarify the manuscript.

3. "My understanding is that the CFL condition is stated as, C = uδtδ x ≤ C max where C max is typically taken equal to 1. The CFL condition is stated differently in lines 219, page 8, and 301, page 11, and it does not seem to me that these statements correspond to the above condition. Please check these lines for consistency."

Sorry this was a typos. It is now corrected. Thank you.

4. "In appendix B, lines 495 - 499, page 21: when going through the derivation of the modified equation for the Euler scheme, I find that κ = u /2 (δ x + uδ t) as the term −u^2 δt /2 2 ∂ x c̃ appears on the left-hand-side of the equation on line 495. Please verify the equations for consistency and update the discussion in lines 238 - 246, page 9. Does this affect any of the numerical results such as Fig 1.b? Are there any other places where this would change the interpretation of the results?"

We would like to thank the referee for his/her comment, actually this was a typos in the retranscription of the derivation given in Appendix B, where a minus in the line 491 has been replaced by a plus in the line 492 (of the previous version of the manuscript). And this error has been propagated in the manuscript in line 495. This is now corrected in the appendix B of the new manuscript. With this correction, you will find that the diffusion coefficient $kappa = u/2(dx -u dt)$. Because it was a typos and not a true error in the manuscript, and that all the simulation has been computed with the correct expression of kappa, nothing has been changed elsewhere.

5. "The conclusion of the numerics is too short. The main point of the work, estimating the model error covariance via the PKF formalism is somewhat successful, and the limitation of the decorrelated error assumption is discussed. However, there isn't much quantitative analysis of practical use of this model error estimate. For instance, can the estimated variance in the PKF formalism be used as an upper bound for the true model error variance?"

New quantitative elements have been introduced in Sec 4.2.2 and Sec. 4.3 in order to feature the characteristic of the part of the model error that is introduced here.

In the new version of the conclusion, the quantitative impact of the flow-dependent modelling, explicitly proposed here, relies on the results of Ménard 2000s who measured a positive impact of using this approach in the assimilation of stratospheric tracer. This is discussed in the new section 5. As the flow-dependent modelling Eq.(18) is introduced when the numerical model is dissipative while the nature is not -- or when the numerical model is more dissipative than the nature –, the estimated variance in the PKF should be considered as a lower bound of the true model-error variance, that should be completed by the residual variance Q. This is now stated in the conclusion (see p17, l13-18):

“The flow-dependent formulation being introduced for modelling the situation where the numerical model is more dissipative than the nature, the model-error variance provided by the PKF should be a lower bound of the true model-error variance, that need a residual climatological covariance to account for the bias. ”

6. “Page 5, lines 121 - 123: I find the parentheses in this sentence to be awkward, as it is easy to confuse their use as mathematical arguments. The conditional statements for the forecast, respectively the analysis, should use commas to clarify the meaning of the text.”

This has been rephrased. Than you.

7. “Page 10, line 255: see the typo “details.Note”.”

This has been corrected, thank you.

8. “Page 13, lines 343 - 344: the definition of N sentence doesn’t seem to be complete. I find the notations confusing here, where N is referred to as a model equation. In this context, what equations represent M?”

The N / M notation in Sec. 4.1 have been verified. There were some typos. Thank you.

9. “Page 15, line 385: see typo “legnth-scale”.”

This has been corrected, thank you.

10. “Page 21, Eq. C1: see typo $x\ ki-1$.”

This has been corrected, thank you.

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

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

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

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

Abstract. This contribution ~~adresses~~ addresses 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~~ modelling of the model-error covariance matrix which applies when the numerical model is dissipative compared with the true dynamics. 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 of the nature and of the numerical model, a numerical simulation is proposed which ~~demonstrates the skill of the parametric methodology in reproducing the~~ illustrates the properties of the resulting model-error covariance matrix~~information~~.

Copyright statement. TEXT

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 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 and on model-error covariance in particular. 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 propa-

gation of covariance functions *i.e.* the propagation by the true equations of the dynamics.

Recently, Pannekoucke et al. (2016, 2018b) (P16) have proposed to solve the Kalman filter equations, and its second-order extension for nonlinear dynamics, using approximated covariance ~~matricies~~ matrices 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; ?)~~ (Cohn, 1993; Pannekoucke, 2020). This approach relies on the partial differential equations encountered in geosciences that are often nonlinear.

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~~ can be explored from the results of Pannekoucke et al. (2018a) (P18).

The paper is organized as follows, the ~~uncertainty propagation is first~~ background in data assimilation is reviewed in Section 2 from which the formalism of the model-error covariance matrix ~~can be deduced at least on a theoretical level~~is detailed with the introduction of a modelling that could apply when the numerical model is dissipative. The model-error covariance matrix ~~estimation~~ based on the PKF is ~~detailed~~ illustrated 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~~ model the flow-dependent part of the model-error covariance matrix to numerical schemes in a ~~one-dimensionnal setting.~~ one-dimensional setting. A discussion on the results is proposed in Section 5. Conclusions and perspectives are given in the last section, Section 6.

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 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

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

where the subscript q is used to denote the time t_q.

Due to the imperfect knowledge of the nature and the limitations encountered during the computation, the nature dynamics is only approximated by

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

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

$$\mathcal{X}_{q+1}^t = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}_q^t) - \varepsilon_{q+1}^m(\mathcal{X}_q^t), \tag{4}$$

where $\varepsilon_{q+1}^m(\mathcal{X}_q^t)$ is the *model error* with respect to the true state, and where ε_{q+1}^m is the function defined by

$$\varepsilon_{q+1}^m = \mathcal{M}_{t_{q+1}\leftarrow t_q} - \mathcal{N}_{t_{q+1}\leftarrow t_q}. \tag{5}$$

The model error $\varepsilon_{q+1}^m(\mathcal{X}_q^t)$ represents collectively the numerical discretization error and the effect of unresolved processes. It is often modeled as a random field of zero mean *i.e.* $\mathbb{E}\left[\varepsilon_{q+1}^m(\mathcal{X}_q^t)\right] = 0$, and of covariance matrix

$$\mathbf{P}_{q+1}^m = \mathbb{E}\left[\varepsilon_{q+1}(\mathcal{X}_q^t)^m \left(\left(\varepsilon_{q+1}^m(\mathcal{X}_q^t)\right)^T\right], \tag{6}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator.

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

$$\mathcal{X}_q^a = \mathcal{X}_q^t + \varepsilon_q^a, \tag{7}$$

where ε_q^a stands for the so called *analysis error*~~model~~, that is modeled 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,

$$\mathcal{X}_{q+1}^f = \mathcal{N}\mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}_q^a). \tag{8}$$

Similarily to the analysis state, the forecast state expands as

$$\mathcal{X}_{q+1}^f = \mathcal{X}_{q+1}^t + \varepsilon_{q+1}^f, \tag{9}$$

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

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. (9) its dynamics is given by

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

where $\mathbf{N}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}$ stands (see Eq. (A2) in Appendix A)

$$\varepsilon^f_{q+1} = \mathbf{M}\varepsilon^a_q + \varepsilon^m_{q+1}(\mathcal{X}^t_q), \tag{10}$$

where $\mathbf{M}$ is a simplified notation for $\mathbf{M}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}$ that is for the tangent linear (TL) propagator along the nature TL dynamicsdefined by analysis trajector. The TL dynamics, with respect to the analysis state, is defined by

$$\partial_t \varepsilon = \mathbf{N}\mathbf{M}_{t,\mathcal{X}^a_t}\varepsilon, \tag{11}$$

where $\mathbf{N}_{t,\mathcal{X}^a_t} = d\mathcal{N}_{|t,\mathcal{X}^a_t}$ $\mathbf{M}_{t,\mathcal{X}^a_t} = d\mathcal{M}_{|t,\mathcal{X}^a_t}$ is the differential of $\mathcal{N}$ $\mathcal{M}$ at $(t,\mathcal{X}^a_t)$, and which . This TL model 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 consequenceMoreover, the decomposition of the forecast-error Eq. (10), makes appear the predictability-error ε^p defined by

$$\varepsilon^p_{q+1} = \mathbf{M}\varepsilon^a_q \tag{12}$$

consequently, the forecast error covariance matrix becomes

$$\mathbf{P}^f_{q+1} = \mathbf{N}_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}\mathbf{P}^a_q\mathbf{N}^T_{t_{q+1}\leftarrow t_q, \mathcal{X}^a_q}. \tag{13}$$

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

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

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

$$\mathcal{X}^t_{q+1} = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) + \varepsilon^m_{q+1},$$

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

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

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

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

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

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

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 errorcan be expanded as

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

where

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

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 matrixhas the form

$$\mathbf{P}^f_{q+1} = \mathbf{P}^p_{q+1} + \mathbf{P}^m_{q+1} + \mathbf{V}^{pm}_{q+1} + (\mathbf{V}^{pm}_{q+1})^T, \textit{where} \textit{where} \mathbf{P}^p_{q+1} = \mathbf{MP} \tag{14}$$

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

When the analysis error and the model error are decorrelated, the forecast-error covariance matrix writes

$$\mathbf{P}^f_{q+1} = \mathbf{P}^p_{q+1} + \mathbf{P}^m_{q+1}. \tag{15}$$

Note that, in the case where the true nature is used to forecast the uncertainty, then the forecast-error covariance matrix coincide with the predictability-error covariance matrix. In the latter, the predictability-error with respect to the nature dynamics plays an important role. So in order to avoid any confusion with the predictability error associated with the numerical model, the notation $\widetilde{\cdot}$ is used when the dynamics is the nature *i.e.*

$$\widetilde{\mathbf{P}}^f_{q+1} = \widetilde{\mathbf{P}}^p_{q+1} = \mathbf{N}\mathbf{P}^a_q\mathbf{N}^T, \tag{16}$$

with $\widetilde{\mathbf{P}}^f = \mathbb{E}\left[\widetilde{\varepsilon}^f \left(\widetilde{\varepsilon}^f\right)^T\right]$ where $\widetilde{\varepsilon}^f_{q+1} = \mathbf{N}\varepsilon^a_q$ denotes the forecast error in the particular case where the dynamics is the nature, which coincide with the predictability error in this case *i.e.* $\widetilde{\varepsilon}^f_{q+1} = \widetilde{\varepsilon}^p_{q+1}$; and where $\mathbf{N}$ is a simplified notation for the propagator $\mathbf{N}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}$ solution of the TL dynamics governed by $\mathbf{N}_{t,\mathcal{X}^a_t} = d\mathcal{N}_{|t,\mathcal{X}^a_t}$ (the differential of $\mathcal{N}$ at $(t,\mathcal{X}^a_t)$).

2.2 Discussion on the modelling of the model error

The modelling of the model error can be seen as a trade-off between its real properties and the lack of knowledge to address this error. In particular, the various assumptions encountered in data assimilation may be considered as suboptimal ways to model this error. For instance, assuming that the model error is unbiased, leads to model the bias as some variance, and over-estimates the effective model-error variance. Then, assuming a decorrelation between the analysis and the model errors is certainly wrong for deterministic error, as for the model error due to the discretization of the dynamics; but it may not apply for highly non-linear processes as for the turbulent processes and transport by the turbulent. Again, assuming the decorrelation between the analysis and the model errors leads to over-estimate the true effect of the model-error with an over-estimation of the true forecast-error uncertainty. However, with these assumptions, or actually this modelling, some part of the model-error statistics can be estimated from the data. For instance, with the assumption that the analysis and the model errors are decorrelated, leading to Eq. (15), it is possible to estimate the homogeneous correlation and the stationary part of the climatological model-error covariance (Daley, 1992; Boisserie et al., 2013).

By some aspects, the understanding and the specification of the model-error covariance matrix look like to the development of the background-error covariance matrix some decades ago where, in variational data assimilation, the climatological background-error covariance matrix has been progressively replaced by a covariance matrix of the day thanks to the hybridization with ensemble methods, used to characterize the predictability error. Nonetheless, the situation of the model-error covariance matrix is different since, up to now, no equations are known to characterize its properties. It seems that the prospect of estimating the model-error covariance matrix of the day is out of reach.

Because the model error can mean different things, to understand the context we are using model error, let's consider the situation sketched in Fig. 1. This figure mimics the evolution of the analysis uncertainty with respect to the nature and to the numerical model. The initial Gaussian analysis error is characterized by the analysis state (the black point) and the analysis-error covariance matrix (the black ellipse). When this analysis uncertainty evolves by the nature dynamics (blue arrow), it becomes a Gaussian uncertainty (under TL assumption) of mean the analysis forecasted by the nature (blue point) and of covariance matrix the associated predictability-error covariance matrix (that is also the forecast-error covariance) (blue ellipse). When the analysis uncertainty evolves by the numerical model (red arrow), it becomes a Gaussian uncertainty (under TL assumption) of mean the analysis forecasted by the model (red point) and of covariance matrix the associated predictability-error covariance matrix (red ellipse). The evolution of the true state (pink crosses) is also represented (pink arrow). Panels (a) and (b) illustrate what would be the forecast-error covariance matrix (orange ellipse) in two situations:

Panel (a) represents the case where the forecast state $\mathcal{X}^f_{q+1}$ is out of the predictability uncertainty of the nature: in that case, a model-error is needed to enlarge the predictability-error covariance of the numerical model so that the forecast-error covariance is larger enough to account for the uncertainty of the nature. In this situation, it seems difficult to speculate about what would be the characteristic of the model error, beyond any climatological estimate. This situation could be the typical picture for long term forecast.

Panel (b) represents the situation where the time integration is not too long, so the forecast state lies within the predictability uncertainty of the nature. This situation is encountered when the numerical model is more dissipative than the nature *e.g.* the resolution of an advection by a semi-Lagrangian scheme. Then the model-error uncertainty, required to correct the predictability error of the numerical model, should be at least large enough to provide an uncertainty similar to the predictability error covariance of the nature. So if we are able to quantify the predictability-error covariances of the nature and of the numerical model, then it would be able to specify a flow-dependent part of the model-error covariance matrix. To account for the bias, a climatological residual covariance matrix would be necessary.

Thereafter, we consider the situation sketched in panel (b), that suggests to decompose the forecast-error covariance matrix as

$$\mathbf{P}^f_{q+1} \approx \mathbf{P}^p_{q+1} + \mathbf{\Pi}^m_{q+1} + \mathbf{Q}_{q+1}, \tag{17}$$

where

$$\mathbf{\Pi}^m_{q+1} = \mathbf{N}\mathbf{P}^a_q\mathbf{N}^T - \mathbf{M}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}\mathbf{P}^a_q\mathbf{M}_{t_{q+1}\leftarrow t_q,\mathcal{X}^a_q}{}^T, \tag{18}$$

~~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 matrixbetween the predictability error and the model error~~would account for the flow-dependent part of the model-error covariance matrix, while the remaining $\mathbf{Q}_{q+1}$, a residual model-error covariance, would account for the bias and could be estimated from the climatology *e.g.* by considering a chi-squared diagnostic (Ménard et al., 2000).

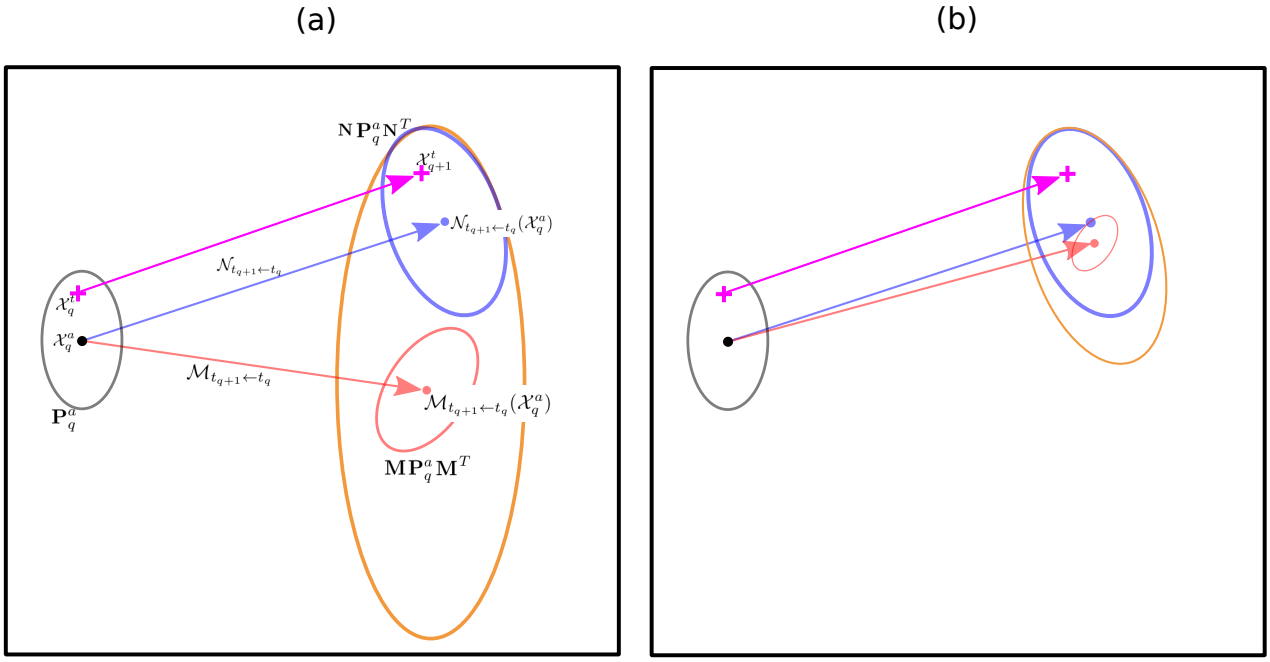

Figure 1. Illustration of the evolution of the uncertainty by the nature and the numerical model: the generic situation (a), and the particular situation particular situation where the forecast lies within the nature uncertainty and where the model is diffusive (b). The predicatbility-error covariance of the nature $\mathbf{N}\mathbf{P}_q^a\mathbf{N}^T$ (of the model $\mathbf{M}\mathbf{P}_q^a\mathbf{M}^T$) is the blue (red) elipse. The forecast-error covariance matrix is the orange elipse.

Thus, $\mathbf{\Pi}_{q+1}^m = \widetilde{\mathbf{P}}_{q+1}^p - \mathbf{P}_{q+1}^p$ measures how the predictability-error covariance of the numerical model should be modifed to find the one of the nature. We think that $\mathbf{\Pi}_{q+1}^m$ could be a usefull proxy to characterize the flow-dependent part of model-error covariance matrix. Note that the matrix $\mathbf{\Pi}_{q+1}^m$ is symmetric but not necessary positive. However, under the assumption depicted in panel (b), we will assume that $\mathbf{\Pi}_{q+1}^m$ is positive. Note also that $\mathbf{\Pi}_{q+1}^m$ is different from the model-error covariance matrix $\mathbf{P}_{q+1}^m$: if there is no analysis uncertainty, then $\mathbf{\Pi}_{q+1}^m$ is zero.

~~The covariance matrices for~~ The decomposition Eq. (17) can be justified from the decomposition of the forecast error ~~and~~ that can be written as (see Eq. (A6) in Appendix A)

$$\varepsilon_{q+1}^f = \widetilde{\varepsilon}_{q+1}^p + \varepsilon_{q+1}^m(\mathcal{X}_q^a), \tag{19}$$

which makes appear the forecast error, ε_{q+1}^f, as the predictability error ~~appear more simple than~~ of the nature, $\widetilde{\varepsilon}_{q+1}^p = \mathbf{N}\varepsilon_q^a$, plus a drift $\varepsilon_{q+1}^m(\mathcal{X}_q^a)$. Note that, the analysis state $\mathcal{X}_q^a$ being known, the model error $\varepsilon_{q+1}^m(\mathcal{X}_q^a)$ is easier to handle than $\varepsilon_{q+1}^m(\mathcal{X}_q^t)$ in Eq. (10), which is defined with respect to the true state $\mathcal{X}_q^t$ that is never known in practice. Now, when assuming that the errors in Eq. (19) are decorrelated and when the model error $\varepsilon_{q+1}^m(\mathcal{X}_q^a)$ is unbiased ($\mathbb{E}\left[\varepsilon_{q+1}^m(\mathcal{X}_q^a)\right] = 0$), and of covariance $\mathbb{E}\left[\varepsilon_{q+1}^m(\mathcal{X}_q^a)\left(\varepsilon_{q+1}^m(\mathcal{X}_q^a)\right)^T\right] = \mathbf{Q}_{q+1}$, it results that the ~~model-error covariance matrix , since they only rely on TL dynamics along the analysis trajectory while~~ forecast-error covariance matrix also writes

$$\mathbf{P}_{q+1}^f = \widetilde{\mathbf{P}}_{q+1}^p + \mathbf{Q}_{q+1}. \tag{20}$$

Hence, the ~~model error depends on the nature trajectory from the unknown true state~~modelling of the model-error covariance as

$$\mathbf{P}_{q+1}^m \approx \mathbf{\Pi}_{q+1}^m + \mathbf{Q}_{q+1}. \tag{21}$$

allows to connect the two formulations Eq. (15) and Eq. (20) of $\mathbf{P}_{q+1}^f$. In fact, while Eq. (15) and Eq. (20) result from a decorrelation assumption of the errors in Eq. (10) and Eq. (19), and because $\mathbf{\Pi}_{q+1}^m$ is not necessary a covariance matrix, then expression of $\mathbf{P}_{q+1}^f$, proposed in Eq. (17), is more like that of Eq. (2.1) where there is no decorrelation assumption.

Compared with climatological modelling of the model-error covariance matrix, as usually encountered in data assimilation, the model for $\mathbf{P}^m$ in Eq. (21) is a state-dependent model of the model-error covariance. Note also that, in panel (b), assuming that there is no bias, while there is one, leads to interpret the bias as a residual model-error whose magnitude can be estimated from the

climatology. Hence, ~~considering the approximation that the analysis errorand the model error are decorrelated (which is not true but it is a common assumption in estimation theory)~~ ~~a practical estimation~~ $\mathbf{P}^m$ modeled by Eq. (21) is an hybrid model that balance the model error of the day with the climatological effect of the model error. In particular, if the initial state is perfectly known, then $\mathbf{\Pi}^m_{q+1}$ is zero, and the model-error is characterized by the climatological residual term $\mathbf{Q}_{q+1}$: the source of this uncertainty corresponds to a forcing term that appears in the dynamics of the model error (see *e.g.* Eq. (4) in Nicolis (2003)) ; this source term is not explored here and its contribution is incorporated in $\mathbf{Q}$ whose magnitude depends on the forecast time.

Note that the modeling Eq. (21) for $\mathbf{P}^m$ is actually supported by at least one real experiment. In the assimilation of a chemical tracer using a Kalman filter, Ménard et al. (2000) and Ménard and Chang (2000) (M2000s) have observed a *loss of variance*: the variance they forecasted was lower than the theoretical variance that was transported by the flow for the advection equation (Cohn, 1993). Said differently, in their experiment, the predicability-error variance computed from the numerical model was lower to the predicability-error variance of the nature they considered, and M2000s related the loss of variance to the discretization of the continuous dynamics. This loss of variance is also encountered when considering an ensemble forecasting of the uncertainty, as latter illustrated in the numerical part (see Section 4.2.1) and also observed in 3D domain simulations (Ménard et al., 2020). Accompagning the loss of variance M2000s also observed that the correlation length-scale they predict where larger, due to the same diffusive process that gives rise to the loss of variance. To cope with the loss of variance, M2000s proposed to correct the predictability-error variance (the diagonal of $\mathbf{P}^p$ in Eq. (14)) so that its magnitude is conserved, as it is supposed to be according to the theory. This renormalization introduced an increase of correlation length that was corrected by a Schur product of the new covariances with an homogeneous isotropic correlation model whose length-scale has been determined so that the total covariance is conserved over time. Indeed, M2000s introduced a modeling 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~~

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

similar to the Eq. (21) introduced here, although they did not explicitly formalize it in this way: their objective was not to characterize the model-error covariance matrix, but to correct the predictability error covariance matrix that they considered erroneous from a theoretical point of view. In particular, M2000s have observed that the Kalman filter, with the corrected predictability-error covariance, required less residual model-error $\mathbf{Q}$ (see Ménard et al. (2000), Section 5), and improved the analysis-error statistics (see Ménard et al. (2000), Fig. 11): the flow dependent modeling Eq. (21) of the model error is in a better agreement with the real forecast uncertainty.

~~Of course,~~

At a computational level, $\mathbf{\Pi}^m_{q+1}$ in Eq. (18) ~~is still quite difficult to solve: for the forecast-error covariance matrix~~ appears easier to obtain than the model-error covariance matrix $\mathbf{P}^m_{q+1}$ as defined by Eq. (6): the predictability-error covariance matrices of the model $\mathbf{P}^p_{q+1}$ (Eq. (14)) and of the nature $\widetilde{\mathbf{P}}^p_{q+1}$ (Eq. (16)) are based only on the TL forecasts with respect to the known analysis state $\mathcal{X}^a_q$, while the model error $\varepsilon^m_{q+1}(\mathcal{X}^t_q)$ (Eq. (4)) depends on the true state $\mathcal{X}^t_q$ that is never known. However, computing $\widetilde{\mathbf{P}}^p_{q+1}$ and $\mathbf{P}^p_{q+1}$ remains a challenge. First of all, the nature dynamics ~~is either unknown or~~$\mathcal{N}$ is generally unknown *e.g.* primitive equations are only an approximation of the geophysical fluid dynamics. Then, when the nature ~~consists in PDEs, no exact solution exists which is easy to handle or interesting enough for applications. For numerical studies,~~ dynamics is (assume) known *e.g.* when it is given by PDEs, there is often no analytical solution which means that the problem must be solved numerically: as $\mathcal{M}$ is presicely the numerical approximation of $\mathcal{N}$, the only way to compute $\widetilde{\mathbf{P}}^p_{q+1}$ is to introduce a high order numerical approximation ~~$\widehat{\mathcal{N}}$~~ of the nature dynamics~~can be considered, with the hope that the numerical errors are~~ $\widehat{\mathcal{N}}$, whose numerical error is much smaller than ~~those~~ the one of $\mathcal{M}$. ~~But the major limitation is~~ And finally, it remains to compute $\widetilde{\mathbf{P}}^p_{q+1}$ and $\mathbf{P}^p_{q+1}$. But due to the large size of the numerical state encountered in ~~geophysics:~~ practice, the direct computation of ~~or is impossiblein practice~~$\widetilde{\mathbf{P}}^p_{q+1} \approx \widehat{\mathbf{N}}\mathbf{P}^a_q\widehat{\mathbf{N}}^T$ and $\mathbf{P}^p_{q+1} = \mathbf{M}\mathbf{P}^a_q\mathbf{M}^T$ is impossible, even on supercomputers, which are only able to handle a few numerical states at full resolution: it is the limitation that motivated the ensemble estimation to solve the Kalman filter equations (Evensen, 2009).

~~We now consider an alternative for calculating the temporal evolution~~ To overcome the above limitations, a high order discretization $\widehat{\mathcal{N}}$ of $\mathcal{N}$ will be introduced in the latter numerical simulation in place of $\mathcal{N}$, *e.g.* in the ensemble estimation of the covariance ~~matrices~~matrice $\widetilde{\mathbf{P}}^p_{q+1} \approx \widehat{\mathbf{N}}\mathbf{P}^a_q\widehat{\mathbf{N}}^T$ only used for the validation. But the computation of $\widetilde{\mathbf{P}}^p_{q+1} = \mathbf{N}\mathbf{P}^a_q\mathbf{N}^T$ and $\mathbf{P}^p_{q+1}$ is investigated through an alternative to the ensemble estimation, as now introduced in the next section.

2.3 Parametric dynamics for VLATcov models

The parametric formulation ~~of covariance evolution can be stated~~ provides a framework where a limited number of covariance parameters (based on the continuous PDE) of the nature can be computed. The parametric formulation works 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~~ featuring the forecast-error covariance matrix so that $\mathbf{P}(\mathcal{P}^f_t) \approx \mathbf{P}^f_t$ ~~(~~; and there is a set $\mathcal{P}^a$ featuring the analysis-error covariance matrix so that $\mathbf{P}(\mathcal{P}^a) \approx \mathbf{P}^a$~~)~~. 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~~ Variance field and the ~~local anisotropic tensors~~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

$$V(\mathbf{x}) = \mathbb{E}\left[\varepsilon(\mathbf{x})^2\right]. \tag{22}$$

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~~defined, from the second order expansion

$$\rho(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}_\mathbf{x}\mathbf{g}(\mathbf{x})}, \tag{23}$$

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

$$\mathbf{g}_{ij}(\mathbf{x}) = \mathbb{E}\left[\partial_{x^i}\left(\frac{\varepsilon}{\sqrt{V}}\right)\partial_{x^j}\left(\frac{\varepsilon}{\sqrt{V}}\right)\right], \tag{24}$$

(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

$$\nu_\mathbf{x} = \frac{1}{2}\mathbf{g}^{-1}_\mathbf{x}, \tag{25}$$

where the superscript $^{-1}$ denotes the matrix inverse operator. Eq. (25) 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

$$\partial_t V = 2\mathbb{E}[\varepsilon\partial_t\varepsilon], \tag{26a}$$

$$\partial_t \mathbf{g}_{ij} = \partial_t\left(\mathbb{E}\left[\partial_{x^i}\left(\frac{\varepsilon}{\sqrt{V}}\right)\partial_{x^j}\left(\frac{\varepsilon}{\sqrt{V}}\right)\right]\right), \tag{26b}$$

where the expectation operator and the temporal derivative commute, $\partial_t\mathbb{E}[\cdot] = \mathbb{E}[\partial_t\cdot]$, as used in Eq. (26a). 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.

2.4 The model-error VLATcov approximation

From now, we will assume that $\mathbf{\Pi}^m$ is a covariance matrix, and that there is no residual model-error $\mathbf{Q}$ so to focus on $\mathbf{\Pi}^m$ alone, so that

$$\mathbf{P}^m_{q+1} \approx \mathbf{\Pi}^m_{q+1}, \tag{27}$$

leads to model the forecast-error covariance matrix as

$$\mathbf{P}^f_{q+1} \approx \mathbf{P}^p_{q+1} + \mathbf{\Pi}^m_{q+1}. \tag{28}$$

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

If the dynamics of the parameters $\mathcal{P}^p_t$ is known, then starting from the initial condition $\mathcal{P}^p_0 = \mathcal{P}^a$ it is possible to approximately determine $\mathbf{P}^p_t$ without solving Eq. (14) and Eq. (16) explicitly.

Hence, thanks to the parametric dynamics in the case where the nature is known from its partial derivative ~~equation~~equations, a new method to compute the model-error covariance matrix can be proposed as follows. By considering the TL dynamics for the numerical model and for the nature~~and the model~~, Equation (26) provides a way to compute both the ~~forecast error covariance matrix $\mathbf{P}^f$, , and the predictability error covariance matrix~~ predictability-error covariance matrices $\mathbf{P}^p$ ~~,~~(Eq. (14)), and $\widetilde{\mathbf{P}}^p$ (Eq. (16)) ; from which the ~~model-error covariance matrix~~ model Eq. (27) of $\mathbf{P}^m$ can be ~~diagnosed from~~ evaluated. For the covariance model based on the diffusion equation, the model-error variance diagnosed from Eq. (18) is the difference

$$V^m = V^f\widetilde{V}^p - V^p, \tag{29a}$$

where ~~V^f (~~$\widetilde{V}^p$ and V^p ~~) denotes the forecast-error (~~denote the predictability-error ~~) variance field~~variance fields of the nature and of the numerical model. The field of the metric tensor of the ~~model-error~~ model error is approximately given by

$$\mathbf{g}^m = \frac{1}{V^m}\left(V^f\mathbf{g}^f\widetilde{V}^p\widetilde{\mathbf{g}}^p - V^p\mathbf{g}^p\right), \tag{29b}$$

where ~~$\mathbf{g}^f$ ($\widetilde{\mathbf{g}}^p$ and $\mathbf{g}^p$) denotes the forecast-error (predictability-error) metric tensor field~~ respectively denote the predictability error metric tensor fields of the nature and of the numerical model (see Appendix B for details).

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

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

$$\partial_t c + u\partial_x c = 0, \tag{30}$$

and takes the place of the nature dynamics Eq. (1). Note that dynamics Eq. (30) is linear, meaning that the tangent-linear dynamics is also given by Eq. (30). The advection equation has two aspects. The first side is given by the PDE Eq. (30) 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

$$\frac{dx}{dt} = u, \tag{31a}$$
$$\frac{dc}{dt} = 0. \tag{31b}$$

Each system Eq. (31) 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. (31) remains to compute the trajectory of a mobile point of coordinate $x(t)$, the *characteristic curve*, solution of the dynamics Eq. (31a), 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. (31b). In the present situation, since the right hand side of Eq. (31b) 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. (30).

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~~ predictability-error covariance matrix, Eq. (16), it is necessary to detail what is the TL dynamics ~~, ,~~ for the linear transport, Eq. (30). Since this transport dynamics is linear, the error evolves according to the same dynamics, and the TL dynamics can be written as

$$\partial_t \varepsilon^f \widetilde{\varepsilon}^p + u\partial_x \varepsilon^f \widetilde{\varepsilon}^p = 0. \tag{32}$$

The PKF approximation of the forecast-error covariance matrix, relies on the dynamics of the variance and of the diffusion fields deduced from Eq. (26). The equation for the variance is computed from Eq. (26a) by replacing the trend by the TL dynamics Eq. (32), so that

$$\partial_t V^f \widetilde{V}^p = 2\mathbb{E}\left[\varepsilon^f \left(-u\partial_x \widetilde{\varepsilon}^p\right)\right] = -2u\mathbb{E}\left[\widetilde{\varepsilon}^p \partial_x \widetilde{\varepsilon}^p\right]. \tag{33}$$

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

$$\partial_t V^f \widetilde{V}^p = 2\mathbb{E}\left[\widetilde{\varepsilon}^p \left(-u\partial_x \widetilde{\varepsilon}^p\right)\right] = -u\partial_x \mathbb{E}\left[\left(\widetilde{\varepsilon}^p\right)^2\right]. \tag{34}$$

By using the definition of the variance Eq. (22), it results that the dynamics for the variance can be stated as

$$\partial_t V^f \widetilde{V}^p = -u\partial_x V^f \widetilde{V}^p. \tag{35}$$

The computation of the metric dynamics Eq. (26b) 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

$$\partial_t V^f \widetilde{V}^p + u\partial_x V^f \widetilde{V}^p = 0, \tag{36a}$$
$$\partial_t \nu^f \widetilde{\nu}^p + u\partial_x \nu^f \widetilde{\nu}^p = (2\partial_x u)\nu^f \widetilde{\nu}^p. \tag{36b}$$

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 stochastic model error).

From Eq. (36), 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, but it is also modified by the source term ~~$(2\partial_x u)\nu^f$~~ $(2\partial_x u)\widetilde{\nu}^p$ 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~~ predicability-error covariance for the nature Eq. (16) has been computed for the linear transport Eq. (30) and corresponds to the time integration of the un-coupled system Eq. (36) starting from prescribed ~~analysis error~~ analysis-error variance and diffusion tensor fields.

The finite difference scheme is now considered as a first numerical integration method for Eq. (30), 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,

$$\frac{c_i^{q+1}-c_i^q}{\delta t} = -u_i \frac{c_i^q - c_{i-1}^q}{\delta x}, \tag{37}$$

Stability is assured as long as the CFL condition ~~$\delta x / \mathrm{Max}_x |u| < \delta t$~~ $\delta t \leq \delta x / \mathrm{Max}_x |u|$ is satisfied. Moreover the scheme is consistent since in the limit of small δx and δt, the dynamics Eq. (30) is recovered from the discrete equation Eq. (37). Thanks to the consistency and the stability, the equivalence theorem of Lax and Richtmyer (1956) assures to the convergence of Eq. (37) toward the true solution. Equation Eq. (37) stands as an illustration of model dynamics Eq. (3).

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~~ finite 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~~ highlights 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~~ from the so-called *modified equation* (Warming and Hyett, 1974). Hence, while it is supposed to solve Eq. (30) the numerical solution computed from Eq. (37) is actually the solution of the modified equation.

More precisely, if ~~C~~ C denotes a smooth function solution of the iterations Eq. (37) with ~~$C(q\delta t, i\delta x) = C_i^q$~~ $C(q\delta t, i\delta x) = C_i^q$, then the modified equation is the partial differential equation verified by ~~C~~ 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 C is given by (see Appendix C)

$$\partial_t C + U \partial_x C = \kappa \partial_x^2 C, \tag{38a}$$

where

$$U = u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u \tag{38b}$$

and

$$\kappa = \frac{u}{2}(\delta x - u \delta t) \tag{38c}$$

are two functions of t and x.

Compared with the nature Eq. (30), the modified equation that best fits the Euler-upwind numerical scheme Eq. (37) 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. (30). 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. (38), it is now possible to compute the TL evolution of the predictability error for the Euler-upwind scheme, which can be expressed as

$$\partial_t \varepsilon^p + U\partial_x \varepsilon^p = \kappa \partial_x^2 \varepsilon^p. \tag{39}$$

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 Eq. (39), 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

$$\partial_t V^p + U\partial_x V^p = -\frac{V^p \kappa}{\nu^p} + \kappa \partial_x^2 V^p - \frac{\kappa \left(\partial_x V^p\right)^2}{2V^p} \tag{40a}$$

$$\begin{aligned}
&\partial_t \nu^p + U\partial_x \nu^p = (2\partial_x U)\nu^p + \\
&\kappa \partial_x^2 \nu^p + 2\kappa - \frac{2\left(\partial_x \nu^p\right)^2}{\nu^p}\kappa + \\
&\partial_x \kappa \partial_x \nu^p - \frac{2\partial_x^2 V^p}{V^p}\kappa\nu^p + \\
&\quad \frac{\partial_x V^p}{V}\kappa\partial_x\nu^p - \frac{2\partial_x V^p}{V^p}\nu^p\partial_x\kappa + \frac{2\left(\partial_x V^p\right)^2}{V^{p2}}\kappa\nu^p
\end{aligned} \tag{40b}$$

Compared with the PKF dynamics of the nature Eq. (36), 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, Eq. (27), associated with the Euler-upwind scheme can be deduced from the ~~forecast and the predictability error~~ 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~~ parametric-error covariance equations of the nature, Eq. (36)~~(~~, and of the numerical discretization, Eq. (40)~~)~~ ~~provides the forecast-error (~~, provides the predictability-error ~~) variance V^f and diffusion ν^f (~~ variances $\widetilde{V}^p$ and V^p, and the diffusion $\widetilde{\nu}^p$ and ν^p~~)~~, which are used to compute the model-error covariance parameter Eq. (29).

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~~ advection 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. (30) for the study of its model error.

The Lagrangian perspective Eq. (31) of Eq. (30) 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 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. (30), 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

$$c(t+\delta t, x_i) = c(t, x_i^*), \tag{41}$$

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. (41) into a discrete scheme from which the modified equation is deduced.

In the case where the discretization satisfies the CFL condition ~~$|u(x)|\delta x < \delta t$~~ $\delta t \leq \delta x / \underset{x}{\mathrm{Max}}\,|u(x)|$ and for linear interpolation, it is straightforward to write the semi-Lagrangian procedure Eq. (41) into a discrete scheme (see Appendix D for the details) which is stated as follows:

$$\begin{cases} \frac{c_i^{q+1}-c_i^q}{\delta t} = -u_i \frac{c_i^q - c_{i-1}^q}{\delta x}, & \text{for } u_i > 0 \\ \frac{c_i^{q+1}-c_i^q}{\delta t} = -u_i \frac{c_{i+1}^q - c_i^q}{\delta x}, & \text{for } u_i < 0 \end{cases} \tag{42}$$

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. (42) is given as the PDE verified by a smooth solution C of Eq. (42). From the derivation detailed in Appendix D, the modified equations is

$$\partial_t C + U \partial_x C = \kappa^{\mathrm{SL}} \partial_x^2 C, \tag{43a}$$

where

$$U = u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u \tag{43b}$$

and

$$\kappa^{\mathrm{SL}} = \frac{|u|}{2}(\delta x - |u|\delta t) \tag{43c}$$

are both functions of t and x.

Hence, since this corresponds mainly to the modified equation Eq. (38) encountered for the Euler-upwind scheme Eq. (37), the parametric predictability-error covariance is also given by Eq. (40), 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

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. (30) is set as the stationary field

$$u(x) = 0.4 + \frac{0.6}{2}\left(1 + \cos\left(\frac{2\pi}{D}(x - D/4)\right)\right), \tag{44}$$

showed in Fig. 2-(a), which appears as a jet with the entrance (exit) at $x = 0.75D$ ($x = 0.25D$): the flow accelerates (decelerates) until $x = 0.25D$ ($x = 0.75D$). Latter, the lead time is $T = 2.0$.

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. (38c), associated with this setting is shown in Fig. 2-(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

$$c(0, x) = \exp\left(-\frac{1}{2(0.15D)^2}\sin\left(\frac{\pi}{2}(x - D/2)\right)^2\right) \tag{45}$$

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 $l_h = 0.05 \approx 12\delta x$~~ $\mathbf{P}^f_{t=0}(x,y) = e^{-\frac{d(x,y)^2}{2l_h^2}}$ with $l_h = 0.05D \approx 12\delta x$, where $d(x,y) = \frac{D}{\pi}|\sin\frac{\pi}{D}(x-y)|$ is the chordal distance between the two geographical positions x and y (Pannekoucke et al., 2018a, see Eq.(30)). 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. (30) exists, this dynamics is integrated considering a fourth order Runge-Kutta time scheme applied on the finite difference discretization

$$\partial_t c_i = -u_i\frac{c_{i+1} - c_{i-1}}{2\delta x}, \tag{46}$$

where the spatial derivative is approximated by a centered second order scheme. This constitutes the ~~higher resolution version~~ high order discretization $\widehat{\mathcal{N}}$ of the ~~model~~ nature $\mathcal{N}$ ~~and~~, as introduced in Section ~~2.1~~2.2: $\widehat{\mathcal{N}}$ 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 $1/\sqrt{6400} \approx 1.25\%$~~than the model $\mathcal{M}$.

Figure 3 shows the trajectory computed from the nature approximated by $\widehat{\mathcal{N}}$ and the ~~nature $\mathcal{N}$~~numerical model $\mathcal{M}$. 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 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 ~~$\mathcal{N}$~~ $\mathcal{M}$ fails to maintain the magnitudes of the extrema (panel (b)), in accordance with the modified equation Eq. (38a) of the Euler-upwind Eq. (37) 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. 2-(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. (38a).

Having validated the two numerical models $\widehat{\mathcal{N}}$ and $\mathcal{M}$, 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~~ predictability-error covariance dynamics of the nature and of the numerical model

The PKF ~~forecast-error~~ predictability-error covariance matrix dynamics for the transport equation Eq. (30) is given by the system Eq. (36). The PKF predictability-error covariance matrix dynamics resulting from the Euler-upwind integration Eq. (37) is given by Eq. (40). Both systems are

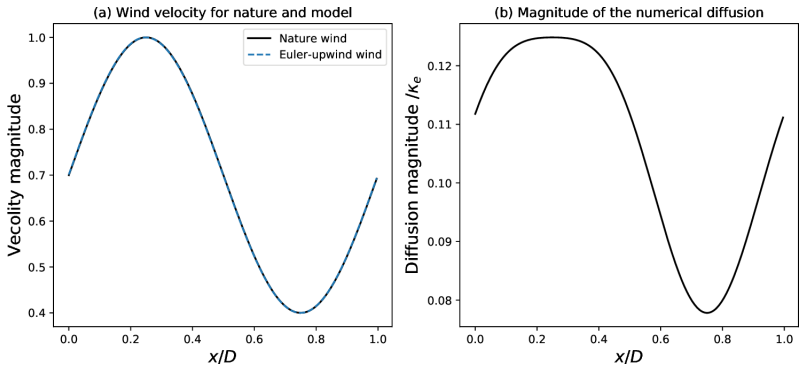

Figure 2. (a) Wind field specified for the nature dynamics and the one seen in the discretized model from Eq. (38b). Panel (b) represents the numerical diffusion coefficient due to the discretization Eq. (38c), normalized by $\kappa_e = \delta x^2/\delta t$.

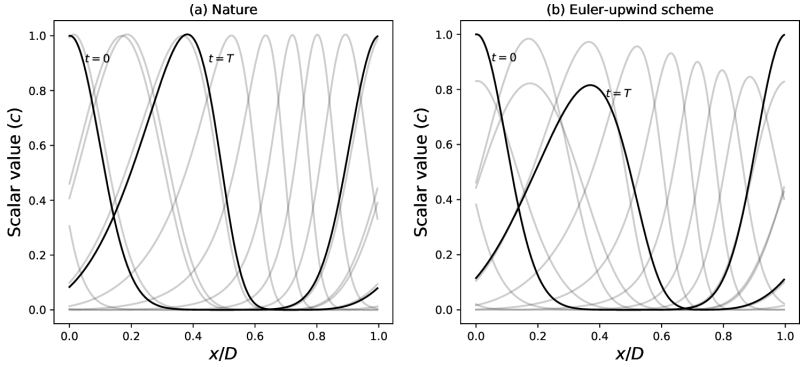

Figure 3. Nature (a) and numerical model (b) runs for times from $t = 0$ to $t = T$ and represented each ~~0.2T~~0.1T.

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~~ fields are shown in Fig. 4. The ~~forecast-error (~~predictability-error ~~)~~ correlation length-scale ~~field~~fields, defined from the one-dimensional diffusion ~~field by $L^f = \sqrt{2\nu^f}$, (~~fields by $\widetilde{L}^p = \sqrt{2\widetilde{\nu}^p}$ (nature) and , $L^p = \sqrt{2\nu^p}$ ~~), is ~~(numerical model), are shown in Fig. 5. The variance and the length-scale, are shown for the PKF and ~~the~~ an 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).

To do so, an ensemble of $N_e = 6400$ analysis errors has been generated, $(\varepsilon^a_{0,k})_{k\in[1,N_e]}$, where each member is computed as $\varepsilon^a_{0,k} = (\mathbf{P}^f_{t=0})^{1/2}\zeta_k$ with ζ_k a sample of the Gaussian random vector of zero mean and of covariance matrix the identity matrix $\mathbf{I}$. This large size limits the sampling noise to a relative error of $1/\sqrt{N_e} \approx 1.25\%$.

Because the dynamics are linear, the TL nature and model are independent of any analysis state, and the ensemble is computed from the forecasts, by the high order discretization of the nature $\widehat{\mathcal{N}}$ and the model $\mathcal{M}$, of the ensemble of analysis errors $(\varepsilon^a_{0,k})$.

~~The forecast-error covariance dynamics~~

4.2.1 Validation of the PKF for the nature

The predictability-error covariance dynamics for the nature is first considered. Since the variance of the nature Eq. (36a) 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. 4-(a) from the PKF integration while the ensemble estimation, Fig. 4-(c) also shows this stationary but to within the sampling noise. The length-scale (Fig. 5-(a)) shows a periodic evolution where, starting from the homogeneous field of ~~L_h~~value l_b, 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. 5-(c) presents the same variations (again to within the sampling noise), which validates the PKF dynamics for the nature. As a consequence, the PKF dynamics Eq. (36) can be used to understand the dynamics of the uncertainty. In particular,

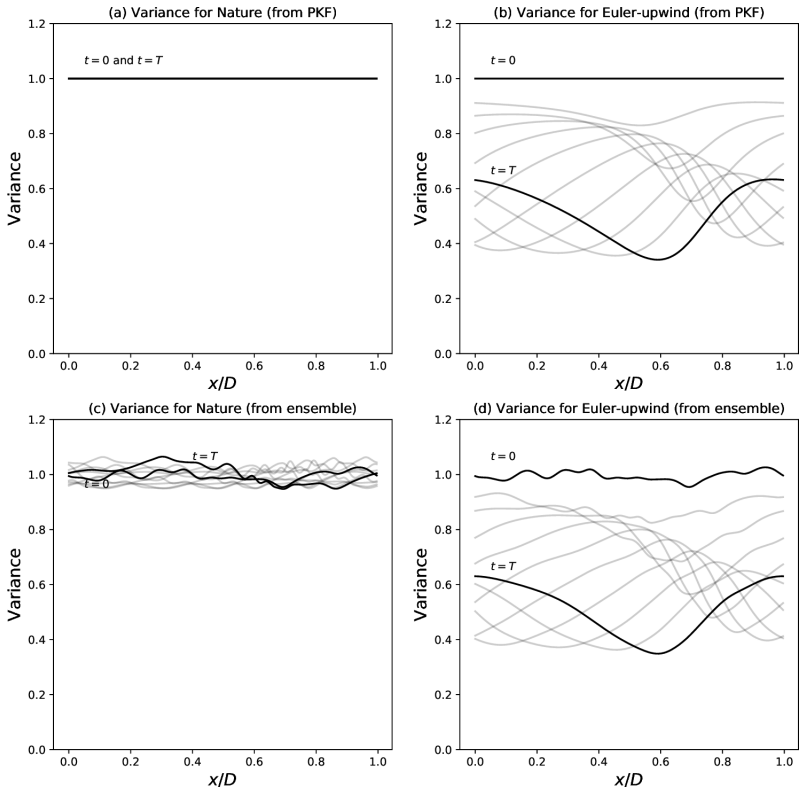

Figure 4. ~~Forecast-error~~ Predictability-error variance field, ~~$V^f(t,x)$~~ $\widetilde{V}^p(t,x)$, for the nature Eq. (30), computed from the PKF Eq. (36) (panel a), and ~~predicatbility-error~~ predictability-error variance field, $V^p(t,x)$, for the numerical model resulting from the finite difference & Euler discretization Eq. (37), computed from the PKF Eq. (40) (panel b). Panels (c) and (d) are the ensemble estimation for panels (a) and (b), where the nature dynamics is approximated by Eq. (46) dynamics in panel (c) (6400 members are used here). Fields are represented for times from $t=0$ to $t=T$ and represented each ~~$0.2T$~~ $0.1T$.

the length-scale field at $t=0.1T$ is well explained by the source/sink term $2(\partial_x u)\widetilde{\nu}^p$ in Eq. (36b) whose magnitude, that lies from -0.004 to 0.004, implies a rapid emergence of a heterogeneity leading to large (small) length-scales for $x\in[0,0.25D]\cup[0.75D,D]$ (for $x\in[0.25D,0.75D]$) where $\partial_x u>0$ ($\partial_x u<0$); and by the transport term $u\partial_x\widetilde{\nu}^p$ that shifts the fields to the right. Note that, by introducing the spatial average operator defined for any function f by $\langle f\rangle(t)=\frac{1}{D}\int f(t,x)dx$ as represented in Fig. 6, the averaged length-scale $\langle\widetilde{L}^p\rangle(t)$ ranges in $[12\delta x, 17.5\delta x]$ (see Fig. 6-(b)) while the $\langle\widetilde{V}^p\rangle(t)$ is the constant 1. (see Fig. 6-(b)).

4.2.2 Validation of the PKF for the numerical model

The predictability-error covariance dynamics for the numerical model is now discussed. For the Euler-upwind scheme, the numerical diffusion resulting from the spatio-temporal discretization in Eq. (38a) implies a damping of the variance along the time axis (see Fig. 4-(b)). The attenuation of the uncertainty governed by Eq. (40), 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. 5-(b) increases by the diffusion while the shear produces similar patterns as for the forecast-error statistics. The ensemble estimation in Fig. 4-(d) and Fig. 5-(d) shows the same signal as the PKF prediction (within the sampling noise) which validates the system Eq. (40). As for the nature, it appears that the PKF dynamics for the numerical model, Eq. (40), explains the dynamics of the uncertainty. In particular, again, the length-scale field at $t=0.1T$ is well explained by the source/sink strain term $2(\partial_x u)\widetilde{\nu}^p$ in Eq. (40b) and by the transport term $u\partial_x\widetilde{\nu}^p$, but this time, compared with Eq. (36b), the source term 2κ in Eq. (40b) implies an increase of the length-scale L^p. Note that the influence of the remaining terms in Eq. (40b) can be neglected at the prime instants of the dynamics: this is because at $t=0$, V^p and ν^p are constant fields ($V^p(t=0)=1$ and $\nu^p(t=0)=l_h^2/2$). Compared with the nature, the behavior of the predictability-error variance of the numerical model presents some source/sink terms (rhs of Eq. (40a)) that explain the emergence of a heterogeneity of the variance field. In particular, the term $-\frac{\kappa}{\nu^p}V^p$ being strictly negative, it is responsible of the damping of the

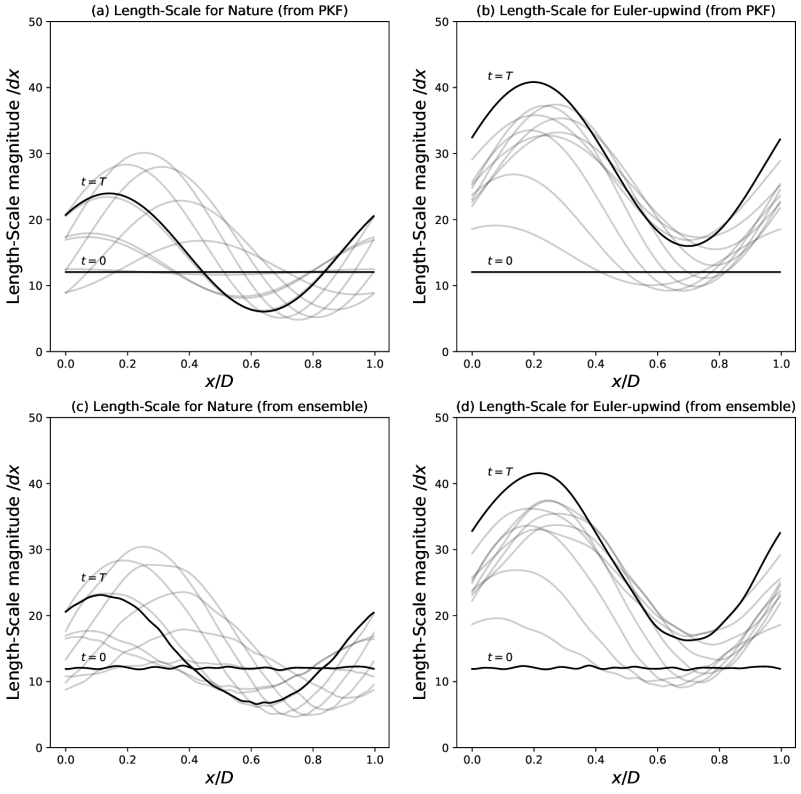

Figure 5. The length-scale counterpart of Fig. 4 representing the ~~forecast-error (the predicatbility-error)~~ predictability-error length-scale ~~field L^f~~ fields $\widetilde{L}^p$ (nature) and L^p (numerical model) in panels (a,c) (in panels (b-d)). The ~~length-scale is~~ length-scales are diagnosed from the diffusion ~~coefficient ν^f (ν^p) as $L^f = \sqrt{2\nu^f}$ ($L^p = \sqrt{2\nu^p}$)~~ coefficients from the formula $L = \sqrt{2\nu}$ 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. 4. Fields are represented for times from $t = 0$ to $t = T$ and represented each ~~$0.2T$~~ $0.1T$

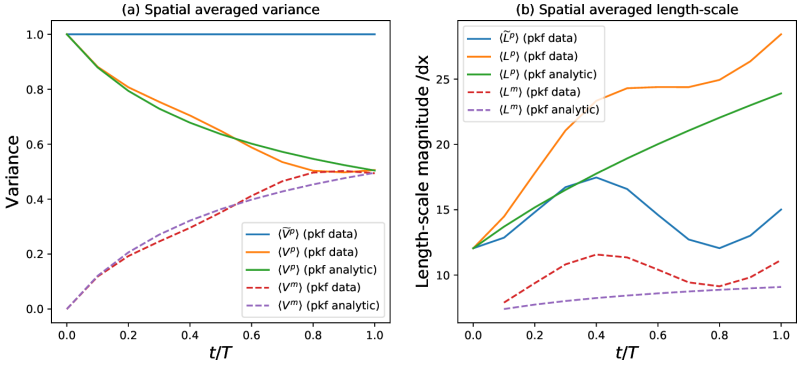

Figure 6. Time evolution of the spatial average over the domain of the predictability-error variance (a) and length-scale (b), computed from the PKF for the nature (blue) and the numerical model (orange). The analytical PKF approximation Eq. (48) for the numerical model is in green. The model error variance (Eq. (29a)) and length-scale (Eq. (49)) are also represented (in dashed lines) for the spatial averaged of the PKF results shown in Fig. 7 (a) and (b) (red), and the analytical approximation (purple).

variance ; it is also responsible of the heterogeneity at the prime instants: the length-scale L^p being heterogeneous, the damping will be more (less) intense in the areas of small (large) length-scales (see Fig. 4-(b) versus Fig. 5-(b) for $t = 0.1T$). In terms of spatial average, with the assumption that the variations around each averaged field are smalls so

that for any fields f and g the approximation $\langle fg\rangle \approx \langle f\rangle\langle g\rangle$ applies, then the spatial averaged of the dynamics Eq. (40) writes

$$\partial_t\langle V^p\rangle \approx -\frac{\langle\kappa\rangle}{\langle\nu^p\rangle}\langle V\rangle, \quad (47a)$$

$$\partial_t\langle\nu^p\rangle \approx 2\langle\kappa\rangle, \quad (47b)$$

where the property that for any function f and integer $k>0$, $\langle\partial_x^k f\rangle = 0$ has been used to eliminate all the other terms. Eq. (47) can be solved analytically, ad its solutions writes

$$\langle V^p\rangle(t) \approx \langle V^p\rangle(0)\left(\frac{\langle\nu^p\rangle(0)}{\langle\nu^p\rangle(0)+2\langle\kappa\rangle t}\right)^{1/2}, \quad (48a)$$

$$\langle\nu^p\rangle(t) \approx \langle\nu^p\rangle(0)+2\langle\kappa\rangle t. \quad (48b)$$

The analytical solution Eq. (48) successfully reproduces the time evolution of the statistics in the present experiment. For the length-scale, Eq. (48b) reproduces the increase (see Fig. 6-(b)), with an underestimation because this solution doesn't account for the oscillation due to the strain term that has been neglected in the dynamics Eq. (47b). For the variance, Eq. (48a) explains a linear decrease at the prime instant, followed by an attenuation in $t^{-1/2}$ (see Fig. 6-(a)).

4.2.3 Intermediate result

As a conclusion of this section, the PKF ~~is~~ appears able to predict the variance and the ~~legnth-scale feature of the forecast-error~~ length-scale features of the predicability-error covariance dynamics of the nature ~~dynamics~~ Eq. (30) and of the ~~predictability-error covariance dynamics resulting from~~ numerical model, that corresponds to the discretization of the true dynamics given by Eq. (37). These results are now considered to provide an estimation of the model-error covariances.

4.3 Model-error covariance ~~diagnosis~~ from the PKF prediction

~~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 $\epsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$. Fields are represented for times from $t=0.2T$ to $t=T$, each $0.2T$ (for $t=0$ the model error is null)~~

~~As discussed in Section 2.1,~~ From the previous section, the Euler-upwind discretization of the advection Eq. (30) leads to an heterogeneous dissipative term, which affects the dynamics of the numerical model uncertainty by damping the variance while increasing the correlation length-scale. When the bias due to the model-error is lower than the predictability-error variance of the nature and that the numerical model is dissipative, then the modelling Eq. (21) of the model-error covariance matrix can be ~~estimated from the difference between the forecast-error and~~ introduced, that is a flow dependent modelling of the model-error covariance plus a climatological residual. This is the situation encountered in the present numerical setting: the predictability-error ~~covariance matrix, , 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 relationship . The question is to know if it is possible to capture some information about the model-error covariance from the approximation~~ variance of the nature is 1 which is larger than the bias (that is at most 0.2 when comparing the nature and the numerical model evolution in Fig. 3), while the predictability-error variance of the numerical model rapidly fails with at its worst a reduction of 60% of the predictability-error variance of the nature (see the reduction at $x=0.6D$ when comparing panels (a) and (b) in Fig. 4). It results that the flow-dependent modelling Eq. (21) may apply here.

In order to ~~tackle this issue, the computation of $\mathbf{P}^m{}_t$ is first made from , considering the PKF estimation of the forecast-error covariance $\mathbf{P}_t^f$ and of the predictability-error covariance $\mathbf{P}_t^p$, following . Then this estimation of $\mathbf{P}^m{}_t$, is compared to the direct ensemble estimation of the~~ focus on the flow dependent part of Eq. (21), the approximation Eq. (27) is considered. Here, $\mathbf{P}^m$ is computed from the parametric approach discussed in Section 2.4, with the parameters Eq. (29), where the predictability-error covariance statistics are computed from Eq. (36) for the nature and Eq. (40) for the numerical model. Note that in this 1D domain situation, Eq. (29b) is equivalent to the computation of the local correlation length-scales by

$$L^m(t,x) = \sqrt{\frac{V^m}{\widetilde{V}^p/(\widetilde{L}^p)^2 - V^p/(L^p)^2}}. \quad (49)$$

The flow-dependent model-error covariance ~~computed from the model error computation $\epsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$, 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~~ parameters are shown in Fig. 7, with the variance in panel (a) and the length-scale in panel (b).

At the initial time, as there is no model-error, the model-error variance is zero. But then, the model-error variance should increase linearly because, the sink term $\frac{\kappa}{\nu^p}V^p$ that is the only non-zero right hand side term in Eq. (36a) and Eq. (40a) (see also the spatially averaged dynamics Eq. (47a)) is a source of model-error variance at the initial time, so that for small t, the order of magnitude of

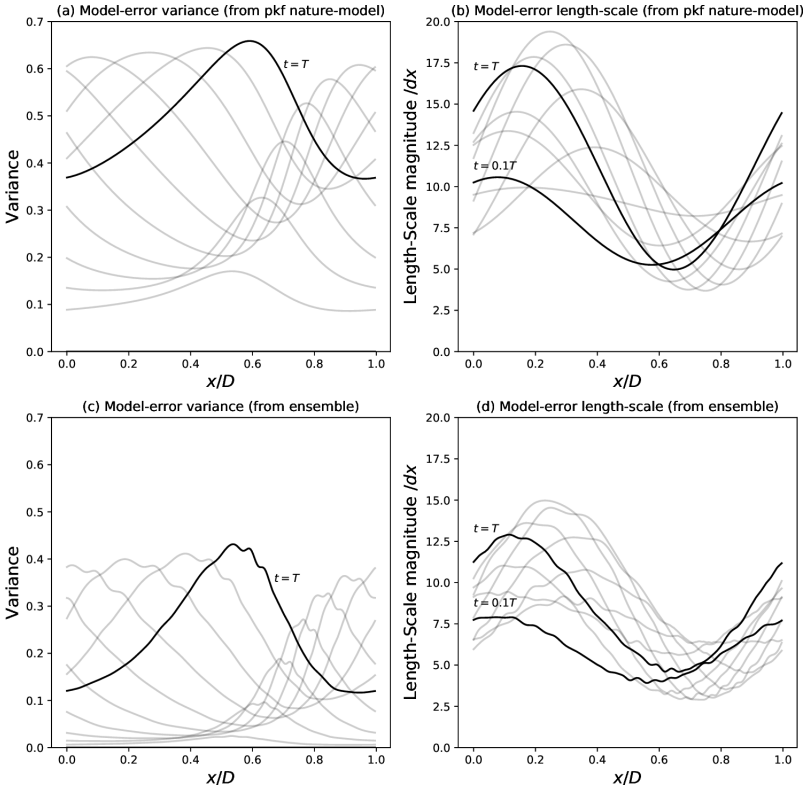

Figure 7. Flow-dependent Model-error ~~variance: (a) computed under decorrelation assumption~~ covariance, modelled from Eq. (21) as ~~the difference between the variance~~ $\mathbf{P}^m = \mathbf{\Pi}^m + \mathbf{Q}$, and computed from the PKF for the nature and ~~shown in Fig. 4~~the Euler-upwind scheme. The variance (a) and the length-scale (bnormalized by dx) ~~;~~ (b) ~~diagnosed from ensemble of differences $\epsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$. Fields~~ are represented for times from ~~$t = 0.2T$~~ $t = 0.1T$ to $t = T$ ~~and represented~~ at each ~~$0.2T$~~ $0.1T$ (for $t = 0$ the model error is null). Comparison with the ensemble estimation of the variance (c) and the length-scale (d) of $\mathbf{P}^{ma}$, Eq. (53).

V^m is given by

$$\langle V^m \rangle(t) \sim t \frac{\langle \kappa \rangle}{\langle \nu^p \rangle(0)} \langle V^p \rangle(0), \tag{50}$$

which relates the increase of the model error variance to the numerical diffusion. Note that the numerical diffusion is not the only process that induces a model error *e.g.* the phase shift due to the correction of the numerical velocity $\frac{\delta t}{2} u \partial_x u$ in Eq. (38b) is also a source term while it has been removed from by the averaging here. Hence, Eq. (50) provides the order of magnitude of the model-error ~~covariance is made by looking at~~ variance at time $t = 0.1T$: when considering the initial conditions $\nu^p(t=0) = l_h^2/2$ and $V^p(t=0) = 1$, and the order of magnitude of the diffusion coefficient $\langle \kappa \rangle \sim 0.1 \delta x^2 / \delta t$ (see Fig. 2-(b)), then the typical values of the model-error variance ~~and length-scale fields.~~

~~Figure ?? shows~~ computed from Eq. (50) is $\langle V^m \rangle(0.1T) \sim 0.12$. This is in accordance with the typical values observed in Fig. 7-(a), for that time. Note that the heterogeneity of the model-error variance field ~~V_t^m approximated by the computation along the simulation as the difference $V_t^f - V_t^p$~~ (panel is due to the heterogeneity of the diffusion field ν^p as discussed in the previous section 4.2.2.

Then, the model-error variance continues to grow, with a peak of uncertainty that evolves with the flow. In this numerical experiment, the magnitude of the $\tilde{V}^p$ being constant and equals to 1., the magnitude of the model-error variance $V^m = \tilde{V}^p - V^p$, shown in Fig. 6-(a), ~~and estimated from the ensemble of model error $\epsilon_k^m = \mathcal{X}_k^f - \mathcal{X}_k^p$ (panel b)~~ ~~this estimation being the approximation of the true model-error statistics. Both estimates of the~~ evolves from Eq. (48a) as

$$\langle V^m \rangle(t) \sim 1 - \left(\frac{l_h^2}{l_h^2 + 4\langle \kappa \rangle t} \right)^{1/2}, \tag{51}$$

when using the initial values $\langle \nu^p \rangle(0) = \frac{1}{2} l_h^2$ and $\langle V^p \rangle(0) = 1.0$. Note that Eq. (51) asymptotically behaves as $1 - \frac{1}{2} \left(\frac{t}{\tau} \right)^{-1/2}$ where $\tau = \frac{l_h^2}{\langle \kappa \rangle} \approx 1.3T$ is the half-magnitude time, that is in accordance with the simulation since $\langle V^m \rangle(T) \sim 0.5$ at the end of the simulation.

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 with the true statistics (panel~~ length-scale, given by Eq. (49), is more difficult to interpret (Fig. 7-(b)) because of the oscillation due to the periodic domain. However, the evolution of the spatial average of the length-scale fields (red dashed line in Fig. 6-(b)) shows an increase of the averaged length-scale with the time, that is in accordance the order of magnitude for the model-error length-scale Eq. (49) computed from the analytical approximations Eq. (48) and Eq. (51), with $\langle \widetilde{V}^p \rangle(t) = 1$ and $\langle \widetilde{L}^p \rangle(t) \sim l_h$ (purple dashed line in Fig. 6-(b))

Note that the model-error length-scale is much smaller, but not null, that will balance the large length-scale of the predictability-error covariance matrix $\mathbf{P}^p$. Hence, as expected, the model-error modeled by Eq. (21) is a heterogeneous covariance that depends on the state and the time: it is flow dependent.

It is interesting to compare $\mathbf{\Pi}^m_{q+1}$ with the covariance of the unbiased error $\varepsilon^{ma}_{q+1} = (\mathbf{N} - \mathbf{M})\varepsilon^a_q$ that appears in the decomposition of the forecast error (see Eq. (A3) in Appendix A)

$$\varepsilon^f_{q+1} = \varepsilon^p_{q+1} + \varepsilon^{ma}_{q+1} + \varepsilon^m_{q+1}(\mathcal{X}^a_q). \tag{52}$$

Indeed, if the errors in the right hand side of Eq. (52) were decorrelated (that is not) then, $\mathbf{\Pi}^m_{q+1}$ in Eq. (17) would have been replaced by the covariance matrix $\mathbf{P}^{ma} = \mathbb{E}\left[\varepsilon^{ma}_{q+1}(\varepsilon^{ma}_{q+1})^T\right]$ given by (see Eq. (A5) in Appendix A):

$$\mathbf{P}^{ma}_{q+1} = \mathbf{\Pi}^m_{q+1} + \left[\left(\mathbf{M}\mathbf{P}^a_q\mathbf{D}^T\right) + \left(\mathbf{M}\mathbf{P}^a_q\mathbf{D}^T\right)^T\right]. \tag{53}$$

with $\mathbf{D} = \mathbf{M} - \mathbf{N}$. ~~But the magnitude of the variance differs: under the analysis-error~~/In practice, $\mathbf{P}^{ma}$ can be estimated from the ensemble of 6400 errors $\varepsilon^{ma}_{q,k} = \widehat{\mathbf{N}}\varepsilon^a_{0,k} - \mathbf{M}\varepsilon^a_{0,k}$ where $\varepsilon^a_{0,k}$ is one of the analysis errors detailed in Section 4.2, and where $\widehat{\mathbf{N}}$ is the TL dynamics associated with the high order numerical approximation $\widehat{\mathcal{N}}$ of $\mathcal{N}$. Because in the present experiment the dynamics of the nature and of the model are linear, $\varepsilon^{ma}_{q,k}$ is computed here as $\varepsilon^{ma}_{q,k} = \widehat{\mathcal{N}}(\varepsilon^a_{0,k}) - \mathcal{M}(\varepsilon^a_{0,k})$. The estimated variance and length-scale fields of $\mathbf{P}^{ma}$ are shown in Fig. 7 (c) and (d). Compared with the PKF modelling (panel a and b), the time evolution shows a similar behavior, but the variance of $\mathbf{P}^{ma}$ is smaller, as well as its length-scale. In this simulation, the contribution of the terms in $\mathbf{D}$, Eq. (53), is to reduce the variance with a maximum of 0.4 at the end of the simulation. However, the minimum of variance of the predictability error is also nearly 0.4. Thus if $\mathbf{P}^{ma}$ were considered in place of $\mathbf{\Pi}^m$, then a residual variance of order 0.2 would be needed (*e.g.* in $\mathbf{Q}$) so to obtain a magnitude of forecast error similar to the predictability of the nature.

Hence, the present numerical experiment illustrated and characterized the flow-dependent part of the model-error ~~decorrelation assumption, the variance averaged over the domain (panel a) increases much faster than for the true~~ covariance $\mathbf{P}^m$, modeled by Eq. (21), in the situation where the model error is related to the discretization of the advection by an heterogeneous wind, leading to a numerical model that is more diffusive than the nature. In this experiment, a linear increase in time, followed by a saturation in $t^{-1/2}$ has been found for the order of magnitude of the model-error variance~~(panel b~~. The residual climatological covariance, $\mathbf{Q}$ in Eq. (21), has yet to be estimated (not considered here).

5 Discussion

Before to conclude, we end this work by addressing some general points about the flow-dependent model which has been introduced here.

The originality of the present contribution is twofold. First, we have formulated a theoretical background corresponding to the model-error covariance matrix and introduced a modelling for its flow-dependent part, Eq. (21). This provides a theoretical framework to the correction of the predictability-error introduced in M2000s. Then, we have provided theoretical and quantitative results about the diffusive effect due to the discretization that can leads to a loss of variance as observed in M2000s: this has been done by combining the formalism of the PKF and the modified equation. The interest for this modelling of the model-error covariance is supported by the results of M2000s who have observed an improvement of the quality of the analysis in their data assimilation system of stratospheric observations.

The ~~length-scale diagnosis, shown in Figure ??, illustrates a similar behaviour. At $t = 0$, the length-scale is not determined since the~~ flow-dependent component of the model-error covariance introduced here can be computed in practice, because it relies on: (1) the analysis uncertainty as characterized by the analysis state and its error covariance that can be estimated in data assimilation; and (2) the time evolution of the analysis-error covariance by the nature and by the numerical model that can be computed from an ensemble method or from the PKF approach.

Note that, if the difference between a low and a high resolution forecast is often used to compute the model-error at a given time, this does not tell anything about the model-error covariances at that time. At most, the model errors collected for a large number of dates, and for the same forecast time, can be used to compute the climatological bias and the climatological model-error covariance~~matrix is null for both simulations, so $t = 0.2T$ is represented in replacement. The~~. To capture the error

of the day following Eq. (21), the computation of the predictability-error covariance matrices is needed.

Hence, the employ of the PKF is important because Eq. (21) needs to estimate not only the predictability-error covariance matrix of the numerical model, but also the one of the nature. If an ensemble estimation of the latter matrix is possible in the research *e.g.* by computing an ensemble of high-resolution forecast with $\widetilde{\mathcal{N}}$ in place of the nature $\mathcal{N}$, it is too costly for real time applications. It results that it is difficult to use Eq. (21) in an ensemble method. Compared with an ensemble method, the PKF remains to compute the evolution of a reduced set of covariance parameters, by computing equations similar to the one encountered in geosciences. For the passive tracer in 1D, the PKF dynamics consists in three equations: for the transport of the concentration, the dynamics of the variance and the dynamics of the local anisotropy (here a diffusion coefficient, related to the correlation length-scale~~translates with the flow and increases or decreases due to~~). So, the numerical cost of the PKF (three equations) for 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~~tracer (one equation) is amount three times the computation of a single forecast, compared to the dozen of members often used in ensemble methods (from which the statistics are corrupted by the sampling noise).

~~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~~For the dynamics of a tracer, the PKF applies in 1D as well as in 2D and 3D domains, where the number of equations are this time of five in 2D, and eight in 3D (the additional equation are for the components of the local anisotropic tensor). However, in general, the use of the PKF is limited by the knowledge of the parameters dynamics. The formalism of the PKF is adapted for dynamics given by partial differential equations, as for the advection of a tracer, but the design of a multivariate PKF formulation is needed so to address multivariate dynamics. Note that for the model error as presented here, the knowledge of the modified equation is a prerequisite that can be difficult to determine in general.

While the PKF is designed from the TL approximation, it is a second order Gaussian filter that is a particular implementation of *nonlinear Kalman-like filters* (Cohn, 1993): for non-linear dynamics, the PKF equation of the mean state depends on the second order moments. However, for long-term predictions, or when the magnitude of the error is too large, the PKF would fails to provide an accurate estimation of the covariance matrices.

6 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 (PKF). The PKF approach applies for a system whose dynamics is given by a set of partial differential equations (PDEs). In the PKF formulation, covariances are approximated by covariance models ~~which are~~ characterized by a set of ~~parameters~~covariance parameters, whose dynamics is deduced from the PDEs of the system, supplemented by an appropriate closure if necessary. We focused on the class of covariance model distinguished by the ~~variance~~ Variance field and the ~~local anisotropic tensors~~ 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,~~ In the case where the numerical model presents a dissipation due to the discretization, or where the numerical model is more dissipative than the nature, we introduced a modelling of the model-error covariance, where its flow-dependent part is approximated as the difference between the parametric approximation of the ~~forecast-error and the predictability-error covariance matrices~~predictability-error covariance matrix of the nature and of the numerical model, plus a residual climatological covariance matrix. This modelling of the flow-dependent part can be computed in real application because it relies on quantities that can be estimated: the analysis state and its analysis-error covariance matrix (or some of its characteristics). For a dynamics given by a partial differential equation, the parametric ~~forecast-error covariance matrix~~ predictability-error covariance matrix of the nature is deduced from the evolution equation while the ~~predictability error covariance matrix~~ predictability-error covariance matrix of the numerical model is computed from the modified evolution, *i.e.* the partial differential ~~equation~~ equations 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~~ forecast 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~~ discretization (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~~ Then, the modelling of the flow dependent part of the model-error covariance ~~, the difference between the forecast-error and the predictability-error covariances was not able to perfectly recover~~ matrix has been computed and discussed. In particular, we discussed the growth of the model-error variance from the understanding of the PKF dynamics, showing a linear increase in time followed by a saturation in $t^{-1/2}$.

The flow-dependent formulation being introduced for modelling the situation where the numerical model is more dissipative than the nature, the model-error variance provided by the PKF should be a lower bound of 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~~ variance, that need a residual climatological covariance to account for the bias.

While there is no data assimilation experiment here, this contribution provides a theoretical background on 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~~ covariance that sheds light on a study previously done by Ménard et al. (2000) and Ménard and Chang (2000) (M2000s), who have observed a loss of variance in the assimilation of a stratospheric tracer by using a Kalman filter: the variance forecasted was lower than the theoretical variance that is supposed to be conserved for the advection (Cohn, 1993). Actually, interpreted as an account of the model error due to the discretization scheme, the correction made by M2000s is similar to the modelling of the flow-dependent part of the model-error ~~covariance characteristics~~covariance matrix we proposed here. In particular, M2000s have observed that the Kalman filter, with the corrected predictability-error covariance, required less residual climatological model-error (see Ménard et al. (2000), Section 5), and an improvement of the analysis-error statistics (see Ménard et al. (2000), Fig. 11), and thus indicating that the modeling of the model error, as proposed here, is in a better agreement with optimality of the nature. Hence, the benefit of the flow-dependent modelling introduced here appears as supported by the improvement of the analysis observed by M2000s in their experiment.

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

Code availability. TEXT

Data availability. TEXT

Code and data availability. TEXT

Sample availability. TEXT

Video supplement. TEXT

Appendix A: Expressions for the forecast error

The aim of this section is to provide the demonstrations of some decompositions of the forecast error: the usual expression as encountered in data assimilation, an expression where the model error is considered with respect to the analysis state, and an expression that makes appear the predictability error with respect to the nature.

A1 Expression of the forecast error as usually encountered in data assimilation

The forecast error is defined in Eq. (9), as the difference $\varepsilon^f_{q+1} = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) - \mathcal{X}^t_{q+1}$. Thanks to Eq. (4), the true state at time t_{q+1} can be replaced so that

$$\varepsilon^f_{q+1} = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) - \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) + \varepsilon^m_{q+1}(\mathcal{X}^t_q), \tag{A1}$$

that makes appear the model error defined by Eq. (5) as $\varepsilon^m_{q+1} = \mathcal{M}_{t_{q+1}\leftarrow t_q} - \mathcal{N}_{t_{q+1}\leftarrow t_q}$. However, with $\mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) = \mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q - \varepsilon^a_q)$ which expands for small analysis error as

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

($\mathbf{M}$ denotes the propagator of the TL model along the analysis state trajectory, see Section 2.1 for details) the forecast error Eq. (A1) becomes

$$\varepsilon^f_{q+1} = \mathbf{M}\varepsilon^a_q + \varepsilon^m_{q+1}(\mathcal{X}^t_q),$$

that is written as

$$\varepsilon^f_{q+1} = \varepsilon^p_{q+1} + \varepsilon^m_{q+1}(\mathcal{X}^t_q), \tag{A2}$$

where $\varepsilon^p_{q+1} = \mathbf{M}\varepsilon^a_q$ is the predictability error, Eq. (12), with respect to the model. The expression Eq. (A2) is the expression of the forecast error usually introduced in data assimilation (Daley, 1992, see Eq. (2.8)). Note that in this expression, the model error is evaluated at the true state $\mathcal{X}^t_q$ while it is never known in practice. It would be interesting to consider an expression with known quantities *e.g.* with the analysis state ; this is now detailed in the next subsection.

A2 Expression of the forecast error considering the model error with respect to the analysis state

The forecast error Eq. (A2) can be obtained by rewriting the model error term as $\varepsilon^m_{q+1}(\mathcal{X}^t_q) = \varepsilon^m_{q+1}(\mathcal{X}^a_q - \varepsilon^a_q)$. Hence, the taylor expansion of ε^m_{q+1} with respect to $\mathcal{X}^a_q$ for small error and lead time, leads to

$$\varepsilon^m_{q+1}(\mathcal{X}^t_q) = \varepsilon^m_{q+1}(\mathcal{X}^a_q) - d\varepsilon^m_{q+1,\mathcal{X}^a_q}\varepsilon^a_q,$$

where $d\varepsilon^m$ denotes the differential of the model error $\varepsilon^m = \mathcal{M} - \mathcal{N}$ (Eq. (5)) which exists when $\mathcal{M}$ and $\mathcal{N}$ are both differentiable, so that $d\varepsilon^m = d\mathcal{M} - d\mathcal{N}$. It results that

$$\varepsilon^m_{q+1}(\mathcal{X}^t_q) = \varepsilon^m_{q+1}(\mathcal{X}^a_q) - (\mathbf{M} - \mathbf{N})\varepsilon^a_q,$$

where $\mathbf{N}$ is the propagator of the TL nature along the analysis state trajectory (see Section 2.1 for details). Then, the forecast error Eq. (A2) expands as

$$\varepsilon^f_{q+1} = \varepsilon^p_{q+1} + \varepsilon^{ma}_{q+1} + \varepsilon^m_{q+1}(\mathcal{X}^a_q), \tag{A3}$$

where ε^{ma}_{q+1} is defined by

$$\varepsilon^{ma}_{q+1} = (\mathbf{N} - \mathbf{M})\varepsilon^a_q. \tag{A4}$$

Note that ε^{ma}_{q+1} is unbiased (at least when the analysis error is unbiased) *i.e.* $\mathbb{E}\left[\varepsilon^{ma}_{q+1}\right] = 0$, so that is covariance matrix is $\mathbf{P}^{ma}_{q+1} = \mathbb{E}\left[\varepsilon^{ma}_{q+1}(\varepsilon^{ma}_{q+1})^T\right]$ which expands as

$$\mathbf{P}^{ma}_{q+1} = \mathbf{N}\mathbf{P}^a_q\mathbf{N}^T + \mathbf{M}\mathbf{P}^a_q\mathbf{M}^T - \left[\left(\mathbf{M}\mathbf{P}^a_q\mathbf{N}^T\right) + \left(\mathbf{N}\mathbf{P}^a_q\mathbf{M}^T\right)^T\right].$$

Replacing the TL model $\mathbf{M}$ by $\mathbf{N} = \mathbf{M} - \mathbf{D}$, leads to

$$\mathbf{P}^{ma}_{q+1} = \mathbf{\Pi}^m_{q+1} + \left[\left(\mathbf{M}\mathbf{P}^a_q\mathbf{D}^T\right) + \left(\mathbf{M}\mathbf{P}^a_q\mathbf{D}^T\right)^T\right]. \tag{A5}$$

where $\mathbf{\Pi}^m_{q+1} = \mathbf{N}\mathbf{P}^a_q\mathbf{N}^T - \mathbf{M}\mathbf{P}^a_q\mathbf{M}^T$ (see Eq. (18)).

As ε^{ma}_{q+1} contains the predictability error, a final expression of the forecast error can be obtained as shown now.

A3 Expression of the forecast error formulated in term of nature predictability

Considering the definition of the predictability error Eq. (12), the forecast error Eq. (A3) rewrites as

$$\varepsilon^f_{q+1} = \mathbf{N}\varepsilon^a_q + \varepsilon^m_{q+1}(\mathcal{X}^a_q), \tag{A6}$$

which makes appear the predictability error with respect to the nature, $\varepsilon^p_{q+1} = \mathbf{N}\varepsilon^a_q$.

Note that Eq. (A6) can be obtained directly from the definition of the forecast error Eq. (9) as follows. By replacing the forecast by $\mathcal{M}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) = \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) + \varepsilon^m_{q+1}(\mathcal{X}^a_q)$, the forecast error first writes $\varepsilon^f_{q+1} = \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) - \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) + \varepsilon^m_{q+1}(\mathcal{X}^a_q)$, where the definition of the nature $\mathcal{X}^t_{q+1} = \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q)$ has been used. Then, rewriting $\mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) = \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q - \varepsilon^a_q)$, whose Taylor expansion is $\mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^t_q) = \mathcal{N}_{t_{q+1}\leftarrow t_q}(\mathcal{X}^a_q) - \mathbf{N}\varepsilon^a_q$ leads to the forecast error Eq. (A6).

Appendix B: 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~~ as Eq. (27) *i.e.*

$$\mathbf{P}^m \approx \mathbf{\Pi}^m = \mathbf{P}^f\widetilde{\mathbf{P}}^p - \mathbf{P}^p,$$

assuming this matrix is a covariance matrix. The local metric tensor can be diagnosed from the Taylor expansion of the model-error correlation function

$$\rho^m(\mathbf{x}, \mathbf{x}+\delta\mathbf{x}) = \frac{1}{\sqrt{\mathbf{P}^m(\mathbf{x},\mathbf{x})\mathbf{P}^m(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})}} \left(\mathbf{P}^f\widetilde{\mathbf{P}}^p(\mathbf{x}, \mathbf{x}+\delta\mathbf{x}) - \mathbf{P}^p(\mathbf{x}, \mathbf{x}+\delta\mathbf{x})\right). \tag{B1}$$

Under an assumption of local homogeneity of the variance, $\mathbf{P}^m(\mathbf{x},\mathbf{x}) \approx \mathbf{P}^m(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$, ~~$\mathbf{P}^f(\mathbf{x},\mathbf{x}) \approx \mathbf{P}^f(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$~~$\widetilde{\mathbf{P}}^p(\mathbf{x},\mathbf{x}) \approx \widetilde{\mathbf{P}}^p(\mathbf{x}+\delta\mathbf{x},\mathbf{x}+\delta\mathbf{x})$, 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

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx \frac{\mathbf{P}^f(\mathbf{x},\mathbf{x})}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\frac{\widetilde{\mathbf{P}}^p(\mathbf{x},\mathbf{x})}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(1-\frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^f_{\mathbf{x}}\widetilde{\mathbf{g}}^p_{\mathbf{x}}}\right) - \frac{\mathbf{P}^p(\mathbf{x},\mathbf{x})}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(1-\frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^p_{\mathbf{x}}}\right). \quad \text{(B2)}$$

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

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}\delta\mathbf{x}^T\left[\frac{1}{\mathbf{P}^m(\mathbf{x},\mathbf{x})}\left(\mathbf{P}^f\widetilde{\mathbf{P}}^p(\mathbf{x},\mathbf{x})\mathbf{g}_\mathbf{x}\widetilde{\mathbf{g}}_\mathbf{x}^{fp} - \mathbf{P}^p(\mathbf{x},\mathbf{x})\mathbf{g}^p_\mathbf{x}\right)\right]\delta\mathbf{x}. \quad \text{(B3)}$$

After identification with the expected form of the expansion

$$\rho^m(\mathbf{x},\mathbf{x}+\delta\mathbf{x}) \approx 1 - \frac{1}{2}||\delta\mathbf{x}||^2_{\mathbf{g}^m_\mathbf{x}}, \quad \text{(B4)}$$

it follows that

$$\mathbf{g}^m_\mathbf{x} = \frac{1}{V^f(\mathbf{x})-V^p(\mathbf{x})}\frac{1}{\widetilde{V}^p(\mathbf{x})-V^p(\mathbf{x})}\left(V^f\widetilde{V}^p(\mathbf{x})\mathbf{g}_\mathbf{x}\widetilde{\mathbf{g}}_\mathbf{x}^{fp} - V^p(\mathbf{x})\mathbf{g}^p_\mathbf{x}\right) \quad \text{(B5)}$$

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

Appendix C: Computation of the modified equation for Euler scheme

The modified partial differential equation associated with the numerical scheme Eq. (37) is the partial differential equation of a smooth function C, solution of the scheme, so that $C(q\delta t, i\delta x) = C^q_i$ *i.e.*

$$\frac{C^{q+1}_i - C^q_i}{\delta t} = -u_i\frac{C^q_i - C^q_{i-1}}{\delta x}, \quad \text{(C1)}$$

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

$$\partial_t C + \frac{\delta t}{2}\partial^2_t C + \mathcal{O}(\delta t^2) = -u\left(\partial_x C - \frac{\delta x}{2}\partial^2_x C + \mathcal{O}(\delta x^2)\right) \quad \text{(C2)}$$

The second order time ~~derivation~~ derivative can be replaced from the equation Eq. (C2) 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

$$\partial_t C = -u\partial_x C + \mathcal{O}(\delta t, \delta x). \quad \text{(C3)}$$

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

$$\begin{aligned}\partial^2_t C &= \partial_t\left(-u\partial_x C\right) + \mathcal{O}(\delta t, \delta x),\\ &= -\partial_t u\partial_x C - u\partial^2_{xt}C + \mathcal{O}(\delta t, \delta x),\end{aligned}$$

~~then~~consequently, the second order derivative $\partial^2_{xt}C$ can be deduced from spatial ~~derivation~~ derivative of Eq. (C3), and writes

$$\begin{aligned}\partial^2_{xt}C &= -\partial_x(u\partial_x C) + \mathcal{O}(\delta t, \delta x),\\ &= -\partial_x u\partial_x C \pm u\partial^2_x C + \mathcal{O}(\delta t, \delta x).\end{aligned}$$

It results that Eq. (C2) writes

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

~~then~~ so that

$$\partial_t C + U\partial_x C = \kappa\partial^2_x C + \mathcal{O}(\delta t^2, \delta x^2), \quad \text{(C4)}$$

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

Appendix D: 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

$$c(t_q, x^*_i) = \left(1 - \frac{x^*_i - x_{i-1}}{x_i - x_{ki-1}}\frac{x^*_i - x_{i-1}}{x_i - x_{i-1}}\right)c^q_{i-1} + \left(\frac{x^*_i - x_{i-1}}{x_i - x_{i-1}}\right)c^q_i = \frac{u_i\delta t}{\delta x}c^q_{i-1} + \left(1 - \frac{u_i\delta t}{\delta x}\right)c^q_i, \quad \text{(D1)}$$

Hence, the numerical scheme writes

$$c_{q+1,i} = \frac{u_i \delta t}{\delta x} c_{q,i-1} + \left(1 - \frac{u_i \delta t}{\delta x}\right) c_{q,i}. \quad \text{(D2)}$$

The modified differential equation is obtained by replacing c by a smooth function $\tilde{c}$, solution of the numerical scheme Eq. (D2). The computation of the modified equation is similar to the Euler case detailed in Appendix C, leading to

$$\partial_t C + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right) \partial_x C = \left(\frac{1}{2} u \delta x - \frac{1}{2} u^2 \delta t\right) \partial_x^2 C. \quad \text{(D3)}$$

When $u < 0$, the differential equation writes

$$\partial_t C + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right) \partial_x C = \left(\frac{1}{2}(-u) \delta x - \frac{1}{2} u^2 \delta t\right) \partial_x^2 C \quad \text{(D4)}$$

Hence, in the general situation,

$$\partial_t C + \left(u - \frac{\delta t}{2}\partial_t u + \frac{\delta t}{2} u \partial_x u\right) \partial_x C = \left(\frac{1}{2} |u| \delta x - \frac{1}{2} u^2 \delta t\right) \partial_x^2 C, \quad \text{(D5)}$$

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. We would like to thank the three anonymous referees for their fruitful comments which have contribute to improve the manuscript. This work was supported by the KAPA project (Kalman ~~Parametrique~~Paramétrique), funded by the French LEFE INSU program.

References

Boisserie, M., Arbogast, P., Descamps, L., Pannekoucke, O., and Raynaud, L.: Estimating and diagnosing model error variances in the Meteo-France global NWP model, Quarterly Journal of the Royal Meteorological Society, 140, 846–854, https://doi.org/10.1002/qj.2173, http://dx.doi.org/10.1002/qj.2173, 2013.

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.

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer, 2009.

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. and Chang, L.-P.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: chi-2-Validated Results and Analysis of Variance and Correlation Dynamics, Monthly Weather Review, 128, 2672–2686, https://doi.org/10.1175/1520-0493(2000)128<2672:AOSCTO>2.0.CO;2, 2000.

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, submitted, 2020.

Pannekoucke, O. and Fablet, R.: PDE-NetGen 1.0: from symbolic partial differential equation (PDE) representations of physical processes to trainable neural network representations, Geoscientific Model Development, 13, 3373–3382, https://doi.org/10.5194/gmd-13-3373-2020, 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/https://doi.org/10.5194/npg-2018-10, https://www.nonlin-processes-geophys-discuss.net/npg-2018-10/, 2018a.

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.