Data Assimilation for Chaotic Dynamics

(Article begins on next page)

16 August 2024

Data assimilation for chaotic dynamics

Alberto Carrassi, Marc Bocquet, Jonathan Demaeyer, Colin Grudzien, Patrick Raanes and Stephane Vannitsem

Abstract Chaos is ubiquitous in physical systems. The associated sensitivity to initial conditions is a significant obstacle in forecasting the weather and other geophysical fluid flows. Data assimilation is the process whereby the uncertainty in initial conditions is reduced by the astute combination of model predictions and real-time data. This chapter reviews recent findings from investigations on the impact of chaos on data assimilation methods: for the Kalman filter and smoother in linear systems, analytic results are derived; for their ensemble-based versions and nonlinear dynamics, numerical results provide insights. The focus is on characterizing the asymptotic statistics of the Bayesian posterior in terms of the dynamical instabilities, differentiating between deterministic and stochastic dynamics. We also present two novel results. Firstly, we study the functioning of the ensemble Kalman filter (EnKF) in the context of a chaotic, coupled, atmosphere-ocean model with a quasi-degenerate spectrum of Lyapunov exponents, showing the importance of having sufficient ensemble

Alberto Carrassi

Dept. of Meteorology and National Centre for Earth Observation, University of Reading, Reading, UK,

Mathematical Institute, University of Utrecht, Netherlands e-mail: `n.a.carrassi@reading.ac.uk`

Marc Bocquet

CEREA, joint laboratory École des Ponts ParisTech and EDF R&D, Universitè Paris-Est, France e-mail: `marc.bocquet@enpc.fr`

Jonathan Demaeyer

Royal Meteorological Institute of Belgium, Brussels, Belgium e-mail: `jodemaey@meteo.be`

Colin Grudzien

Department of Mathematics and Statistics, University of Nevada, Reno, Reno, Nevada, USA e-mail: `cgrudzien@unr.edu`

Patrick Raanes

NORCE, Bergen, Norway e-mail: `para@norceresearch.no`

Stephane Vannitsem

Royal Meteorological Institute of Belgium, Brussels, Belgium e-mail: `svn@meteo.be`

1

members to track all of the near-null modes. Secondly, for the fully non-Gaussian method of the particle filter, numerical experiments are conducted to test whether the curse of dimensionality can be mitigated by discarding observations in the directions of little dynamical growth of uncertainty. The results refute this option, most likely because the particles already embody this information on the chaotic system. The results also suggest that it is the rank of the unstable-neutral subspace of the dynamics, and not that of the observation operator, that determines the required number of particles. We finally discuss how knowledge of the random attractor can play a role in the development of future data assimilation schemes for chaotic multiscale systems with large scale separation.

1 Introduction

This chapter attempts a unified and comprehensive discussion of a number of studies that in about a decade have contributed to shape our nowadays's understanding of the implications, impacts and consequences for data assimilation when is applied to chaotic dynamics. The chapter presents a review of the essential results appeared in different studies, but for the first time together in a coherent treatment. In addition we present new original findings addressing two key aspects that were not covered in previous studies. The first treats the impact of data assimilation on a chaotic and multiscale system, the second concerns the consequences for nonlinear, non-Gaussian, data assimilation (a particle filter) in face of the chaotic nature of the underlying dynamics.

The exposition is organised as follows. We first discuss, in Sect. 2, the chaotic character of atmospheric and oceanic flows, and provide a treatment of the key mathematical concepts and tools that allow for characterising chaos and to "measure" the degree of instabilities. In doing so, we review classical results from dynamical system theory, including the *multiplicative ergodic theorem* and associated definition of Lyapunov (forward, backward and covariant) vectors and exponents.

Section 3 analyses how a chaotic dynamics impacts data assimilation. Our focus is on Kalman filter (KF) and smoother (KS) in Sect. 3.1, and on their ensemble-based formulations, the ensemble Kalman filter (EnKF) and smoother (EnKS), in Sect. 3.2. Section 3.1 contains primarily analytic results and treats KF and KS in linear systems, either purely deterministic (Sect. 3.1.1) or with stochastic additive noise (Sect. 3.1.2). Section 3.2 is dedicated to nonlinear systems (deterministic, in Sect. 3.2.1 and stochastic in Sect. 3.2.2) and on how the EnKF (Evensen, 2009a) and the extended Kalman filter (EKF, Ghil and Malanotte-Rizzoli (1991)) works in this scenario. In Sect. 3.2.1 we present original results on the impact of chaos on the performance of the EnKF in a coupled atmosphere-ocean model which possesses a degenerate-like spectrum of Lyapunov exponents, disentangling on the role of the quasi-null exponents.

Section 4 reverses the perspective and instead of studying the effect of chaos on data assimilation, reports on how properties of the dynamics have been used to devise

adaptive observation strategies and ad-hoc data assimilation methods. In particular, in Sect. 4.2, we succinctly review the assimilation in the unstable subspace (AUS, Palatella et al. (2013)), a known approach that exploits the unstable-neutral subspace of the dynamics to perform the analysis. Remarkably, AUS was conceived before the findings reviewed in Sect. 3. The latter work was inspired by these early studies, and the need to provide mathematical rigor to clarify the mechanisms that made these early studies successful. In turn, the later work has furthermore provided the framework to generalize these early ideas to a variety of other types of dynamics.

Section 5 presents what we consider two main areas of future developments. In Sect. 5 the paradigm of AUS is incorporated within a fully nonlinear data assimilation scheme, the particle filter (**?**). It is shown that observing in the directions of instabilities is effective but also that it is not deleterious to observe the stable directions. As opposed to the data or model sizes, our results suggest that the number of particles required to achieve good results scales with the size of the unstable-neutral subspace. Section 5.2 discusses how the concept of random attractor can offer novel ways to handle the data assimilation problem in stochastic chaotic multi-scale systems with large scale separation. It also treats the implications of the numerical scheme on the output of the data assimilation cycle and ensemble-based forecasts, as well as poses some key questions for future studies.

Final conclusions and a summary are drawn in Sect. 6.

2 Chaos in atmospheric and oceanic flows

The atmosphere and the ocean are fluids that are described by the set of classical conservation laws of hydrodynamics, including the conservation of mass, momentum and energy (Vallis, 2017). For the atmosphere, these are often complemented by the conservation of moisture present in the air. These laws lead to a set of local dynamical equations describing the motion of each parcel of fluid. Given that these equations are nonlinear with complex interactions with the boundaries, realistic solutions cannot be obtained analytically and one must rely on numerical simulations starting from appropriate initial conditions.

Numerical simulations are based on discrete approximations in space and time of the dynamical equations, and are often accompanied with simplifications of the equations in order to describe appropriately the scales of interest. One of the most famous approximations is the geostrophic approximation which assumes a balance between the horizontal velocity field and the horizontal pressure gradient. Geostrophic balance is a good approximation for both the ocean and the atmosphere albeit at different spatial scales. The geostrophic approximation is at the basis of the models that will be used later in Sect. 3.2.1

Whatever the scale at which the atmospheric or the ocean fluids are observed, they display an apparently erratic evolution. This erratic behavior is also present in atmospheric, ocean and climate models when appropriate forcing are imposed. This feature should not be confused with randomness as most of the models used

since the start of numerical modelling were purely deterministic. Lorenz (1963) showed in a simple low-order model that this erratic behavior is concomitant with the property of sensitivity to initial conditions, by which whatever small an error in the initial condition is, it will increase rapidly in time. This property implies that given the inevitable error in the initial conditions, any forecast will ultimately become useless as the error finally reaches an amplitude of the same order as the natural variability of the variable considered. The sensitivity to initial conditions and the consequent erratic-like evolution are the key properties of deterministic dynamical systems displaying *chaos*.

This behaviour has been found in many atmospheric, oceanic and climate models (see *e.g.,* Vannitsem, 2017, for a review). In particular, a detailed investigation of the chaotic nature of the coupled ocean-atmosphere system that will be used in Sect. 3.2.1 has been performed by Vannitsem et al. (2015).

2.1 Measuring sensitivity to initial conditions

The notion of sensitivity to initial conditions of dynamical solutions was already discovered and studied in a mathematical context by Poincaré (1899). In the second half of the 20th century, this property was discovered in models of atmospheric and climate relevance (Thompson, 1957; Lorenz, 1963). Its important practical implications drove the regain of interest in developing the appropriate mathematics in support of its description and understanding. These efforts culminated with the development of the ergodic theory of deterministic dynamical systems and chaos theory. In the following we shall briefly summarise what we consider to be the key developments that led to the definition of Lyapunov exponents and vectors.

The importance of these mathematical objects stands on their ability to "measure" the degree of instabilities and thus to quantify the aforementioned sensitivity to initial conditions. While more rigorous mathematical treatments can be found in appropriate mathematical literature (see *e.g.*, Pikovsky and Politi, 2016, and references therein), and we will invoke that rigour to a certain extent in the following sections, here we approach the discussion with a physical and intuitive angle. Further details can also be found in Legras and Vautard (1996); Barreira and Pesin (2002); Kuptsov and Parlitz (2012).

Let us write the evolution laws of a deterministic dynamical system in the form of a set of ordinary differential equations (ODEs),

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\sigma}), \tag{1}$$

where $\mathbf{x}$ is a vector containing the entire set of relevant variables $\mathbf{x} = (x_1, ..., x_n)$ and $\boldsymbol{\sigma}$ represents a set of parameters. The discussion that follows holds also when system (1) explicitly depends on time, thus being non-autonomous, provided it remains ergodic.

Since the process of measurement is always subject to finite precision, the initial state is never known exactly. To study the evolution and the implications of such an error, let us consider an initial state displaced slightly from $\mathbf{x}_0$ by an initial error $\delta\mathbf{x}_0$. The perturbed initial state, $\mathbf{x}_0 + \delta\mathbf{x}_0$, generates a new trajectory in phase space and one can define the instantaneous error vector as the vector joining the representative points of the reference trajectory and the perturbed one at a given time, $\delta\mathbf{x}(t)$. Provided that this perturbation is sufficiently small and smooth, its dynamics can be described by the linearized equation,

$$\frac{d\delta\mathbf{x}}{dt} \approx \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\Big|_{\mathbf{x}(t)} \delta\mathbf{x}, \tag{2}$$

with a formal solution,

$$\delta\mathbf{x}(t) \approx \mathbf{M}(t, \mathbf{x}(t_0))\delta\mathbf{x}(t_0). \tag{3}$$

The matrix $\mathbf{M}$ is referred to as the "fundamental matrix" and is the resolvent of Eq. (2), *i.e.* $\mathbf{M}(t, \mathbf{x}(t_0)) = e^{\int_{t_0}^{t} \frac{\partial \mathbf{F}}{\partial \mathbf{x}}|_{\mathbf{x}(t)} dt}$. The fundamental matrix contains the information on the amplification of infinitesimally small perturbations.

In the context of the ergodic theory of deterministic dynamical systems, the Oseledets theorem (Kuptsov and Parlitz, 2012) shows that the limit of the matrix $(\mathbf{M}^{\top}\mathbf{M})^{1/[2(t-t_0)]}$, for time going to infinity, exists; let us refer to this limiting matrix as $\mathbf{S}$. The logarithm of its eigenvalues are called the *Lyapunov exponents* (LEs), whereas the full set of LEs is called the Lyapunov spectrum, and is usually represented in decreasing order. The eigenvectors of $\mathbf{S}$, which are local properties of the flow (they change along the trajectory thus being time-dependent) and depend on the initial time t_0, are called the *forward Lyapunov vectors* (FLVs) (Legras and Vautard, 1996).

The LEs are a powerful tool to "measure" chaos and to characterise the degree of instability of a system. For instance, LEs are averaged (asymptotic) indicators of exponential growth (LE>0) or decay (LE<0) of perturbations under the tangent-linear model. A deterministic chaotic system is uniquely characterised by having at least its leading LE larger than zero, *i.e.* $LE_1 > 0$. The sum of the LEs is equal to the average divergence of flow generated by Eq. (1) (see *e.g.* Pikovsky and Politi (2016)). This means that in dissipative (conservative, *e.g.* Hamiltonian) systems the sum of the LEs is negative (zero): volumes in the phase space of a dissipative (conservative) dynamics reduce (are conserved) on average with time. Furthermore, autonomous continuous-in-time systems generally possess at least one LE=0, unless they converge to a motion-less state. This null exponent is related to perturbations aligned to the system's velocity vector (*i.e.* $\delta\mathbf{x} = \mathbf{f}$), so although they shall fluctuate depending on the local flow, they will not on average decay nor growth. These last two properties can also be used as a check for numerical accuracy when computing the LEs.

The multiplicative ergodic theorem (MET) (Barreira and Pesin, 2002, Theorem 2.1.2) guarantees that, under general hypotheses, the eigenvalues of the matrix $(\mathbf{M}^{\top}\mathbf{M})^{1/[2(t-t_0)]}$ obtained for $t \rightarrow \infty$ are equivalent to the ones of the matrix $\mathbf{S}' = (\mathbf{M}\mathbf{M}^{\top})^{1/[2(t-t_0)]}$ when $t_0 \rightarrow -\infty$. The equivalence of the spectrum of $\mathbf{S}$ and $\mathbf{S}'$

is not generically true for the fundamental matrix of an arbitrary, linear dynamical system; however, the MET guarantees that this is a fairly generic property of the resolvent of the *tangent-linear model of a nonlinear dynamical system*. If the flow of the time derivative $\mathbf{f}$ is a C^1 diffeomorphism of a compact, smooth, Riemannian manifold M, then the MET assures that the LEs are defined equivalently by the log-eigenvalues of $\mathbf{S}$ or $\mathbf{S}'$ for any initial condition $\mathbf{x}(t_0)$ and that the LEs are unique on a subset of M of full measure with respect to any ergodic invariant measure μ of the flow. Note that, without the condition of ergodicity of μ, the LEs may be well-defined point-wise, but the specific values and their multiplicity may depend on the initial condition $\mathbf{x}(t_0)$.

The matrices $\mathbf{S}$ and $\mathbf{S}'$ are symmetric. However, contrary to the eigenvalues, the eigenvectors of these two matrices are not equivalent due to the asymmetric character of the fundamental matrix $\mathbf{M}$ in forward- and reverse-time. The eigenvectors of the latter are called the *backward Lyapunov vectors* (BLVs). Theoretically, each matrix can be evaluated at the same place along the reference trajectory $\mathbf{x}(t)$ and their orthogonal eigenvectors can be computed as $\mathbf{L}_t^{\mathrm{f},i}$ and $\mathbf{L}_t^{\mathrm{b},i}$ for $\mathbf{S}$ and $\mathbf{S}'$, respectively, where it is understood that the time-dependence on t is with respect to the linearization of the dynamics at $\mathbf{x}(t)$. There exist *Oseledet* subspaces W_t^i,

$$W_t^i = \mathbf{L}_t^{\mathrm{b},1} \oplus ... \oplus \mathbf{L}_t^{\mathrm{b},i} \cap \mathbf{L}_t^{\mathrm{f},i} \oplus ... \oplus \mathbf{L}_t^{\mathrm{f},N}, \tag{4}$$

with $\oplus$ being the direct product (Ruelle, 1979), that have the important properties of being invariant under the effect of the fundamental matrix, such that

$$\mathbf{M}(\tau, \mathbf{x}(t))W_t^i = W_\tau^i. \tag{5}$$

Due to their orthogonal nature, the FLVs and the BLVs require by definition the choice of a norm and of an inner product. Nevertheless, the Oseledet subspaces themselves do not have this dependence; in this way, they can be considered to embed more invariant information about the dynamics. The decomposition of the tangent-linear space into these covariant subspaces is commonly known as the "Oseledet splitting" or decomposition. The classical form of the MET thus guarantees that the Oseledet splitting is well-defined and consistent with probability one over all initial conditions of the attractor, with respect to the invariant, ergodic measure. Other covariant splittings of the tangent-linear model, such as by exponential dichotomy, exist under more general forms of the MET (Froyland et al., 2013).

When the Lyapunov spectrum is non-degenerate, one can define a time-varying basis, subordinate to the Oseledet spaces, $W_t^i = \mathrm{span}\left\{\mathbf{L}_t^{\mathrm{c},i}\right\}$, such that

$$\mathbf{M}(\tau, \mathbf{x}(t))\mathbf{L}_t^{\mathrm{c},i} = \alpha_i(\tau, \mathbf{x}(t))\mathbf{L}_\tau^{\mathrm{c},i}, \tag{6}$$

where $\parallel \mathbf{L}_t^{\mathrm{c},i} \parallel = \parallel \mathbf{L}_\tau^{\mathrm{c},i} \parallel$, and $\alpha_i(\tau, \mathbf{x}(t)) \in \mathbb{R}$ describes an amplification factor. The vectors $\mathbf{L}_t^{\mathrm{c},i}$ are known as the *covariant Lyapunov vectors* (CLVs). In the long-time limit, the amplifications $\alpha_i(\tau, \mathbf{x}(t))$ can be associated to the LEs as,

$$\pm \lambda_i = \lim_{\tau \to \pm\infty} \frac{1}{\tau} ln \ |\alpha_i(\tau, \mathbf{x}(t))| = \lim_{\tau \to \pm\infty} \lambda_i^\tau(\mathbf{x}(t)) \tag{7}$$

where we indicate by $\lambda_i^\tau(\mathbf{x}(t))$ the average of the growth rate taken over a time window τ starting at t at position $\mathbf{x}(t)$), and are commonly known as the local Lyapunov exponents (LLEs) (Pikovsky and Politi, 2016, see chapter 5).

Throughout the chapter we will denote the matrix with columns corresponding to the full tangent linear space ordered basis of BLVs/FLVs/CLVs at time t as $\mathbf{L}_t^e$ for $e \in \{b, f, c\}$ respectively. A sub-slice of this matrix of Lyapunov vectors corresponding, inclusively, to columns i through j will be denoted $\mathbf{L}_t^{e,i:j}$. In the following, the relationships between these Lyapunov vectors and their asymptotic dynamics will be key to understanding the predictability of chaotic systems. For this purpose, we will largely use the BLVs and the CLVs, and to a lesser extent the FLVs.

Note that the basis $\{\mathbf{L}_t^{c,i}\}_{i=1}^{N_x}$ is not orthogonal in general and, in fact, the angles between any two Oseledec subspaces W_i and W_j are not in general bounded away from zero in their limits in forward- or reverse-time. For this reason, coordinate transformations to covariant Oseledet bases are not generally numerically well-conditioned asymptotically, and therefore hard to compute. The property of *integral separation* (Dieci and Van Vleck, 2002), describing the stability of the Lyapunov spectrum under bounded perturbations of the tangent-linear equations, ensures that the angles between the covariant subspaces will remain bounded away from zero, but this is a strong condition and it is not as generic of a property as the existence of the Oseledet decomposition under the MET. However, if a dynamical system is integrally separated as above, there exists a well-defined, numerically well-conditioned transformation of coordinates of the tangent-linear model for which the action of the resolvent $\mathbf{M}$ can be expressed as a block-diagonal matrix with each block describing the invariant dynamics of a single Oseledet space. For degenerate spectrum, these blocks may be upper-triangular, but for non-degenerate spectrum this representation of the resolvent becomes a strictly diagonal matrix; see Theorem 5.4.9 of Adrianova (1995) for the classical result, or Theorem 5.1 of Dieci and Van Vleck (2007) and Froyland et al. (2013) for more recent extensions.

3 Data assimilation in chaotic systems - How the dynamics impacts the way we assimilate data

The high sensitivity of a chaotic dynamical system to the initial condition makes it hard to forecast it, even when their evolution equations are perfectly known. Indeed the typical error grows exponentially over time with a rate given by the largest positive LE. With a view to forecasting, one has no choice but to regularly correct its trajectory using information on the system state obtained through observations. This is the primary goal of data assimilation (DA) and has been key to the success of numerical weather forecasting. We refer to Kalnay (2003); Asch et al. (2016) and references therein for textbooks and reviews on geophysical DA.

When framed in a Bayesian formalism, the goal of DA is to estimate the conditional probability density function (pdf) of the system state knowing observations of that system. For high-dimensional systems, only approximations of this pdf can be obtained, among which the Gaussian approximation is the most practical and common (Carrassi et al., 2018). In the case of Gaussian statistics of the error and linear dynamics, the conditional pdf can be obtained analytically: it is Gaussian, and can be sequentially computed using the Kalman filter (KF) (Kalman, 1960). Directly inspired from the Kalman filter, the extended Kalman filter (EKF) offers an approximate DA scheme when the operators are nonlinear (Ghil and Malanotte-Rizzoli, 1991). However, it can hardly be used in the context of high-dimensional systems, where it has to be replaced with the ensemble Kalman filter (EnKF) (Evensen, 2009b). In the EnKF, the covariance matrices are represented by state perturbations which are representative of the errors and, together with the state mean, form a limited-size ensemble of state vectors.

Understanding how the ensemble of the EnKF evolves under the action of the DA scheme and the forecast model dynamics is important. Several numerical results suggest that the skills of ensemble-based DA methods in chaotic systems are related to the instabilities of the underlying dynamics (Ng et al., 2011). Numerical evidences exist that some asymptotic properties of the ensemble-based covariance matrices (rank, span, range) relate to the unstable modes of the dynamics (Sakov and Oke, 2008a; Carrassi et al., 2009). Nevertheless, a better, more profound, understanding of these results was needed to aim at designing reduced rank, computationally cheap, formulations of the filters.

Analytic results that have shed lights on the behaviour of filters and smoothers on chaotic dynamics, and explained the numerically observed properties, have been obtained for linear dynamics. Section 3.1.1 reviews those findings in the case of deterministic systems without model error, based on the work by Gurumoorthy et al. (2017) and Bocquet et al. (2017). Section 3.1.2 treats their extension to the case of stochastic systems, following the work by Grudzien et al. (2018a,b). In the case of nonlinear dynamics, robust numerical evidences will be described by the original experiments in Sect. 3.2, that extent the previous findings by Bocquet and Carrassi (2017).

3.1 Linear dynamics: the effect of chaos on the Kalman filter and smoother

3.1.1 Perfect and deterministic dynamics

At time t_k, let $\mathbf{x}_k \in \mathbb{R}^{N_x}$ and $\mathbf{y}_k \in \mathbb{R}^{N_y}$ be the state and observation vector, respectively. Let us assume linear evolution model dynamics $\mathbf{M}_k$ and observation model $\mathbf{H}_k$ such that

$$\mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{w}_k, \tag{8a}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k. \tag{8b}$$

The model and observation noises, $\mathbf{w}_k$ and $\mathbf{v}_k$, are assumed mutually independent, zero-mean Gaussian white sequences with statistics

$$\mathrm{E}[\mathbf{v}_k \mathbf{v}_l^\top] = \delta_{k,l} \mathbf{R}_k, \quad \mathrm{E}[\mathbf{w}_k \mathbf{w}_l^\top] = \delta_{k,l} \mathbf{Q}_k, \quad \mathrm{E}[\mathbf{v}_k \mathbf{w}_l^\top] = \mathbf{0}. \tag{9}$$

In the Kalman filter, which yields the optimal DA solution with such assumptions, the forecast error covariance matrix $\mathbf{P}_k$ satisfies the recurrence

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1} \left(\mathbf{I} + \mathbf{P}_k \mathbf{\Omega}_k \right)^{-1} \mathbf{P}_k \mathbf{M}_{k+1}^\top + \mathbf{Q}_{k+1}, \tag{10}$$

where

$$\mathbf{\Omega}_k \equiv \mathbf{H}_k^\top \mathbf{R}_k^{-1} \mathbf{H}_k, \tag{11}$$

is the *precision matrix* of the observations mapped into state space. In the absence of model error, *i.e.* $\mathbf{Q}_k \equiv \mathbf{0}$, Gurumoorthy et al. (2017) proved rigorously that in the full-rank KF (in particular $\mathbf{P}_0$ is full rank), $\mathbf{P}_k$ collapses onto the unstable-neutral subspace.

We shall summarise in the following some key results that apply to the full-rank but also to the degenerate case, *i.e.* even if $\mathbf{P}_0$ is of arbitrary rank and the initial errors of the DA scheme only lie in a subspace of $\mathbb{R}^{N_x}$. In particular they apply to the EnKF if the dynamical and observations operators are both linear. We recall them here, but readers can find full details in Bocquet et al. (2017); Bocquet and Carrassi (2017).

Result 1: <u>Bound on the free forecast error covariance matrix</u>

Let us define the resolvent of the dynamics from t_l to t_k as $\mathbf{M}_{k:l} = \mathbf{M}_k \mathbf{M}_{k-1} \cdots \mathbf{M}_{l+1}$, with the convention that $\mathbf{M}_{k,k} = \mathbf{I}$. The first key result is the following inequality in the set of the semi-definite symmetric matrices:

$$\mathbf{P}_k \leq \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^\top + \mathbf{\Xi}_k =: \mathbf{P}_k^{\text{free}}, \tag{12}$$

where

$$\mathbf{\Xi}_0 \equiv \mathbf{0} \quad \text{and for } k \geq 1 \quad \mathbf{\Xi}_k \equiv \sum_{l=1}^{k} \mathbf{M}_{k:l} \mathbf{Q}_l \mathbf{M}_{k:l}^\top, \tag{13}$$

is known as the *controllability* matrix (Jazwinski, 1970), and the suffix "free" is used to refer to the forecast error of the system unconstrained by data. In the absence of model noise ($\mathbf{Q}_k \equiv \mathbf{0}$ for the rest of this section), it reads

$$\mathbf{P}_k \leq \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^\top. \tag{14}$$

Assuming the dynamics to be non-singular, the column subspace of the forecast error covariance matrix satisfies

$$\text{Im}(\mathbf{P}_k) = \mathbf{M}_{k:0}\left(\text{Im}(\mathbf{P}_0)\right). \tag{15}$$

If n_0 is the dimension of the unstable-neutral subspace of the dynamics, it can further be shown that

$$\lim_{k\to\infty}\text{rank}(\mathbf{P}_k) \le \min\left\{\text{rank}(\mathbf{P}_0), n_0\right\}, \tag{16}$$

which is a first proof of the collapse of the error covariance matrix (actually its column space) onto the unstable and neutral subspace of the dynamics.

Result 2: Collapse onto the unstable subspace

Let σ_i^k, for $i = 1, \dots, N_{\text{x}}$ denote the eigenvalues of $\mathbf{P}_k$ ordered as $\sigma_1^k \ge \sigma_2^k \cdots \ge \sigma_{N_{\text{x}}}^k$. It was shown that

$$\sigma_i^k \le \beta_i \exp\left(2\lambda_i^k k\right), \tag{17}$$

for some $\beta_i > 0$, where λ_i^k is a log-singular value of $\mathbf{M}_{k:0}$ that converges to the LE λ_i. This gives an upper bound for all eigenvalues of $\mathbf{P}_k$ and a rate of convergence for the $N_{\text{x}} - n_0$ smallest ones. Moreover, if $\mathbf{P}_k$ is uniformly bounded, it can further be shown that the stable subspace of the dynamics is asymptotically in the null space of $\mathbf{P}_k$, *i.e.*,

$$\lim_{k\to\infty}\left\|\mathbf{P}_k\mathbf{L}_k^{\text{b},i}\right\| = 0 \tag{18}$$

for all $i > n_0$; this extends to any norm and linear combination of these vectors.

Result 3: Explicit dependence of $\mathbf{P}_k$ on $\mathbf{P}_0$

To study the dependence of $\mathbf{P}_k$ on $\mathbf{P}_0$, it has been shown that the forecast error covariance matrix can be written as

$$\mathbf{P}_k = \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^\top\left(\mathbf{I} + \mathbf{\Gamma}_k\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^\top\right)^{-1}, \tag{19}$$

where the matrix

$$\mathbf{\Gamma}_k \equiv \sum_{l=0}^{k-1}\mathbf{M}_{k:l}^{-\top}\mathbf{\Omega}_l\mathbf{M}_{k:l}^{-1}, \tag{20}$$

is known as the *information matrix* and it measures the *observability* of the system by propagating the precision matrices $\mathbf{\Omega}_l$ up to t_k.

An alternative formulation of Eq. (19) is

$$\mathbf{P}_k = \mathbf{M}_{k:0}\mathbf{P}_0\left[\mathbf{I} + \mathbf{\Theta}_k\mathbf{P}_0\right]^{-1}\mathbf{M}_{k:0}^\top, \tag{21}$$

where

$$\boldsymbol{\Theta}_k \equiv \mathbf{M}_{k:0}^\top \boldsymbol{\Gamma}_k \mathbf{M}_{k:0} = \sum_{l=0}^{k-1} \mathbf{M}_{l:0}^\top \boldsymbol{\Omega}_l \mathbf{M}_{l:0}, \tag{22}$$

is also an *information* matrix, directly related to the *observability* of the DA system, but pulled back at the initial time t_0. Equation (21) is of key importance because it allows to study the asymptotic behaviour of $\mathbf{P}_k$, *i.e.* the filter "believed" error, using the asymptotic properties of the dynamics.

Result 4: Asymptotics of $\mathbf{P}_k$

For $\mathbf{P}_k$ to forget about $\mathbf{P}_0$ as k tends to infinity, it was shown that one can impose the following sufficient conditions:

- **Condition 1:** Recall that the FLVs at t_0 associated to the unstable and neutral exponents are the columns of $\mathbf{L}_0^{\mathrm{f},1:n_0} \in \mathbb{R}^{N_x \times n_0}$. Moreover, let define the anomaly matrix $\mathbf{X} \in \mathbb{R}^{N_x \times n_0}$ such that $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$. The condition reads

$$\mathrm{rank}\left(\mathbf{X}_0^\top \mathbf{L}_0^{\mathrm{f},1:n_0}\right) = n_0. \tag{23}$$

  In practice the initial ensemble anomalies $\mathbf{X}_0$ projects onto the first n_0 FLVs at t_0.
- **Condition 2:** The model is sufficiently observed so that the unstable and neutral directions remain under control, *i.e.*, there exits $\varepsilon > 0$ such that

$$\left(\mathbf{L}_k^{\mathrm{b},1:n_0}\right)^\top \boldsymbol{\Gamma}_k \mathbf{L}_k^{\mathrm{b},1:n_0} > \varepsilon \mathbf{I}. \tag{24}$$

- **Condition 3:** Furthermore, for any neutral BLV we have

$$\lim_{k\to\infty} \left(\mathbf{L}_k^{\mathrm{b},n_0}\right)^\top \boldsymbol{\Gamma}_k \mathbf{L}_k^{\mathrm{b},n_0} = \infty, \tag{25}$$

  implying that the neutral direction is well observed and controlled.

  Under these three conditions, we obtain

$$\lim_{k\to\infty}\left\{\mathbf{P}_k - \mathbf{L}_k^{\mathrm{b},1:n_0}\left[\left(\mathbf{L}_k^{\mathrm{b},1:n_0}\right)^\top \boldsymbol{\Gamma}_k \mathbf{L}_k^{\mathrm{b},1:n_0}\right]^{-1}\left(\mathbf{L}_k^{\mathrm{b},1:n_0}\right)^\top\right\} = \mathbf{0}. \tag{26}$$

Hence, the asymptotic sequence does not depend on $\mathbf{P}_0$, but only $\boldsymbol{\Gamma}_k$, *i.e.* the dynamics. It can also be shown that the neutral modes have a peculiar role and a long lasting influence: their influence on the current estimate decreases sub-exponentially.

Result 5: From the degenerate KF to the square-root EnKF and EnKS

The recurrence Eq. (21) can be reformulated in a factorised form which is suited to the square-root EnKF. The standard perturbation decomposition of the forecast error

covariance is

$$\mathbf{P}_k = \mathbf{X}_k \mathbf{X}_k^\top, \tag{27}$$

where $\mathbf{X}_k$ is the matrix of centred perturbations (the anomaly matrix aforementioned). A right-transform update formula can then be obtained from (21):

$$\mathbf{X}_k = \mathbf{M}_{k:0} \mathbf{X}_0 \left[\mathbf{I} + \mathbf{X}_0^\top \mathbf{\Theta}_k \mathbf{X}_0 \right]^{-1/2} \mathbf{\Upsilon}_k, \tag{28}$$

where $\mathbf{\Upsilon}_k$ is an arbitrary orthogonal matrix such that $\mathbf{\Upsilon}_k \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector $[1 \ldots 1]^\top$ defined in the ensemble subspace. It is equivalent to the left-transform update formula

$$\mathbf{X}_k = \left[\mathbf{I} + \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^\top \mathbf{\Gamma}_k \right]^{-1/2} \mathbf{M}_{k:0} \mathbf{X}_0 \mathbf{\Upsilon}_k. \tag{29}$$

The importance of Eqs. (28) and (29) stands on the fact that with linear models, Gaussian observation and initial errors, the (square-root) degenerate KF (with $\mathbf{X} \in \mathbb{R}^{N_x \times n}$ and $n < N_x$) is equivalent to the square-root EnKF and can serve as a proxy to the EnKF applied to nonlinear models.

All of these results can be generalised to linear smoothers. In particular, the smoother forecast error covariance matrix is similar to that of the filter, *i.e.* given by (19) (or (21)) but with the following modified information matrix:

$$\widehat{\mathbf{\Gamma}}_k = \mathbf{\Gamma}_k + \sum_{l=k}^{k+L-S} \mathbf{M}_{k:l}^{-\top} \mathbf{\Omega}_k \mathbf{M}_{k:l}^{-1}, \tag{30}$$

where L is the lag of the smoother (how far in the past observations are accounted for) and S tells by how many time steps the smoother's window is shifted between two consecutive updates. Note that $\widehat{\mathbf{\Gamma}}_k \geq \mathbf{\Gamma}_k$ (using the Loewner order on the set of semi-definite positive matrices) reflecting the general higher amount of information incorporated within a smoother update relative to a filter. Therefore the asymptotic sequences for the filter (right-hand side) and smoother (left-hand-side) follow the inequality:

$$
\begin{aligned}
\mathbf{L}_k^{\mathrm{b},1:n_0} \left[\left(\mathbf{L}_k^{\mathrm{b},1:n_0} \right)^\top \widehat{\mathbf{\Gamma}}_k \mathbf{L}_k^{\mathrm{b},1:n_0} \right]^{-1} \left(\mathbf{L}_k^{\mathrm{b},1:n_0} \right)^\top \leq \\
\mathbf{L}_k^{\mathrm{b},1:n_0} \left[\left(\mathbf{L}_k^{\mathrm{b},1:n_0} \right)^\top \mathbf{\Gamma}_k \mathbf{L}_k^{\mathrm{b},1:n_0} \right]^{-1} \left(\mathbf{L}_k^{\mathrm{b},1:n_0} \right)^\top.
\end{aligned}
\tag{31}
$$

The linear smoothers can serve as a proxy to the ensemble Kalman smoother (Evensen, 2009b) or the iterative ensemble Kalman smoother (IEnKS) (Bocquet and Sakov, 2014) applied to nonlinear models (Bocquet and Carrassi, 2017). With the IEnKS, where the true state trajectory is even better estimated and the errors are reduced, the collapse of the perturbations onto the unstable and neutral subspace is expected to be even faster.

3.1.2 Stochastic dynamics

The above perfect-deterministic model configuration demonstrates how chaos shapes the inferences of the posterior. Asymptotic statistics are determined by the ability of the filter to control the growth of initial error in the unstable-neutral subspace with respect to its sensitivity to observations therein. However, in realistic DA, additional forecast errors are introduced throughout the forecast cycle due to the inadequacy of numerical models in representing reality. One classical approach to treat these model errors is to represent them as additive or multiplicative noise (Jazwinski, 1970). Oseledet's theorem and the Lyapunov spectrum are also formulated for such systems of stochastic differential equations (SDEs) and discrete maps.

Suppose $\{\mathbf{v}^i\}_{i=0}^d$ is a collection of C^3 vector fields on $M \subset \mathbb{R}^{N_x}$, a smooth manifold without boundary. Define $\{W_t^i\}_{i=1}^d$ where each W_t^i is an independent Wiener process defined on the probability space $(\Omega, \mathcal{F}, P)$. This describes a generic Stratonovich SDE,

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\sigma}, \omega) \triangleq \mathbf{v}^0(\mathbf{x}, \boldsymbol{\sigma}) + \sum_{i=1}^d \mathbf{v}^i(\mathbf{x}, \boldsymbol{\sigma}) \circ W_t^i(\omega). \tag{32}$$

For fixed $\boldsymbol{\sigma}$, if $\{\mathbf{v}^i\}_{i=1}^d$ span the tangent space $T_{\mathbf{x}}M$ for each $\mathbf{x} \in M$, then the system of SDEs gives rise to a unique probability measure μ on M that is invariant with respect to the random flow induced by the system of SDEs. For $\mu \times P$ almost every $(\mathbf{x}, \omega) \in M \times \Omega$, the Lyapunov exponents and their multiplicity are defined and only depend on $\mathbf{x}$ (Liu and Qian, 2006)[theorem 2.1]. For the SDE, $\mathbf{f}(\mathbf{x}, \boldsymbol{\sigma}, \omega)$, the tangent linear model is once again defined as in Eq. (2) from which the exponents and vectors can be computed as usual (Pikovsky and Politi, 2016)[chapters 2 and 8].

The analysis from perfect-deterministic models thus extends to stochastically forced models as in Eq. 8a, but with key differences. Firstly, the forecast error covariance can generally be considered to be of full rank due to the injection of the stochastic forcing $\mathbf{w}_k$ into arbitrary subspaces; the standard controllability assumption actually guarantees that it is of full rank after a sufficient lead time (Jazwinski, 1970)[lemma 7.3]. Stochastic perturbations $\mathbf{w}_k$ are, moreover, subject to growth and decay rates of the LLEs, λ_i^τ of Eq. (7). These LLEs are distributed about the LE such that even asymptotically stable modes may have transient periods of rapid growth.

To make the analysis tractable, assume that the dynamics are stationary in the sense that the recursive QR algorithm (Shimada and Nagashima, 1979; Benettin et al., 1980) converges uniformly to the theoretical LEs uniformly in number of iterations from any initial time. While the forecast error covariance is not generally reduced-rank, the stable dynamics is actually sufficient to uniformly bound the forecast error variances in the stable subspace *without any assimilation*.

Result 6: Uniform bounds on the variance of errors in the stable subspace

Recall that $\mathbf{L}_k^{b,i}$ is the i-th BLV and that $\mathbf{P}_k^{\text{free}}$ is the free forecast error covariance (*e.g.* without DA). For $\lambda_i < 0$, define $\epsilon > 0$ such that $\exp\{2\lambda_i + \epsilon\} < 1$. By the stationarity assumed above, there exists N_ϵ such that if $k - l > N_\epsilon$, then

$$-\epsilon < \frac{1}{k-l} \log\left(\| \mathbf{M}_{k:l}^\top \mathbf{L}_k^{b,i} \|\right) - \lambda_i < \epsilon. \tag{33}$$

Assuming that the system is uniformly completely controllable and $\mathbf{Q}_k \le q_{\sup}\mathbf{I}_n < \infty$ for all k, the controllability matrix, Eq. (13), is uniformly bounded above by $C_{N_\epsilon}\mathbf{I}_n$. A variation on Eq. (12) can be used to obtain a uniform bound on the i-th variance of $\mathbf{P}_k^{\text{free}}$ in the basis of the BLVs as

$$\limsup_{k\to\infty} \left(\mathbf{L}_k^{b,i}\right)^\top \mathbf{P}_k^{\text{free}}\mathbf{L}_k^{b,i} \le C_{N_\epsilon} + \frac{q_{\sup}\exp\{2(\lambda_i + \epsilon)N_\epsilon + 1\}}{1 - \exp\{2(\lambda_i + \epsilon)\}}. \tag{34}$$

This bound represents the competing forces of the transient growth rates of recently introduced perturbations in the controllability matrix, and perturbations that adhere to their aysmptotic log-average rate of decay in the stable subspace within a margin of ϵ (Grudzien et al., 2018a)[proposition 2 and corollary 3]. This demonstrates that, if assimilation prevents error growth in the span of the unstable-neutral BLVs, errors in the span of the stable BLVs can be neglected without relinquishing filter boundedness. However, this does not state whether the error in the span of the stable subspace will remain within tolerable bounds.

Redefine $q_{\sup}$ such that $\mathbf{P}_0^{\text{free}}, \mathbf{Q}_k \le q_{\sup}\mathbf{I}_n$ for all k. The variance of the free forecast in the i-th BLV is bounded directly as

$$\left(\mathbf{L}_k^{b,i}\right)^\top \mathbf{P}_k^{\text{free}}\mathbf{L}_k^{b,i} = \left(\mathbf{L}_k^{b,i}\right)^\top \mathbf{M}_{k:0}\mathbf{P}_0^{\text{free}}\mathbf{M}_{k:0}^\top\mathbf{L}_k^{b,i} + \sum_{l=1}^{k} \left(\mathbf{L}_k^{b,i}\right)^\top \mathbf{M}_{k:l}\mathbf{Q}_l\mathbf{M}_{k:l}^\top\mathbf{L}_k^{b,i}$$

$$\le q_{\sup} \sum_{l=0}^{k} \left(\mathbf{L}_k^{b,i}\right)^\top \mathbf{M}_{k:l}\mathbf{M}_{k:l}^\top\mathbf{L}_k^{b,i} = q_{\sup} \sum_{l=0}^{k} \| \left(\mathbf{T}_{k:l}^\top\right)^i \|^2, \tag{35}$$

where $\| \left(\mathbf{T}_{k:l}^\top\right)^i \|$ is the norm of the i-th row of $\mathbf{T}_{k:l}$ in the recursive QR decomposition of the propagator, $\mathbf{M}_{k:l} = \mathbf{L}_k^b\mathbf{T}_{k:l}\left(\mathbf{L}_l^b\right)^\top$. The sum

$$\Psi_k^i \triangleq \sum_{l=0}^{k} \| \left(\mathbf{T}_{k:l}^\top\right)^i \|^2 \tag{36}$$

describes the invariant evolution for the i-th variance of the free forecast error covariance matrix in the basis of BLVs. In the case that $\mathbf{Q}_k = q\mathbf{I}_n$ for all k, Eq. (35) becomes an equality and Ψ_k^i can be interpreted as the evolution of the i-th variance

when $q = 1$. As $k - l$ grows $\| \left(\mathbf{T}_{k:l}^{\top} \right)^{i} \|$ converges exponentially to zero while for $k - l$ close to one this describes transient dynamics in the basis of the BLVs. Although Ψ_k^i is guaranteed to be uniformly bounded in k (Grudzien et al., 2018a)[corollary 3], numerical simulations demonstrate how this uniform bound can be extremely large.

Figure 1, presents an example from Grudzien et al. (2018a) of the free forecast error variance, Ψ_k^i, over 10^4 forecast cycles where the model propagator $\mathbf{M}_k$ is defined by the evolution of the tangent linear model of the Lorenz-96 system (Lorenz, 1996) in $N_x = 10$ dimensions, with an interval between observations of 0.1, and $\mathbf{Q}_k \triangleq \mathbf{I}_{N_x}$. Although this model is generated from the underlying nonlinear Lorenz system, the state model for the experiment is treated as a discrete linear model using only the tangen-linear resolvent as above. This model has three unstable, one neutral and six stable LEs. While λ_5 and λ_6 are negative, $\mathbf{L}_k^{b,5}$ and $\mathbf{L}_k^{b,6}$ experience frequent transient instabilities in the timeseries of their LLEs (see bottom row in Fig. 1). The LLEs of $\mathbf{L}_k^{b,5}$ have more intense growth, reflected in the differences between Ψ_k^5 and Ψ_k^6 (top row): the maximum of Ψ_k^6 is on the order of $O(10^2)$ and the mean is approximately 28; for Ψ_k^5 the max is of $O\left(10^3\right)$ and the mean is approximately 808. This suggests that, as opposed to the case of deterministic dynamics, for successful DA in stochastic systems with additive model error, it is necessary to control the growth of forecast errors also in the span of weakly stable BLVs. While errors in this span will not grow indefinitely and remain bounded, if those directions are left uncontrolled their error bounds can be practically too large for any meaningful state estimation purposes.
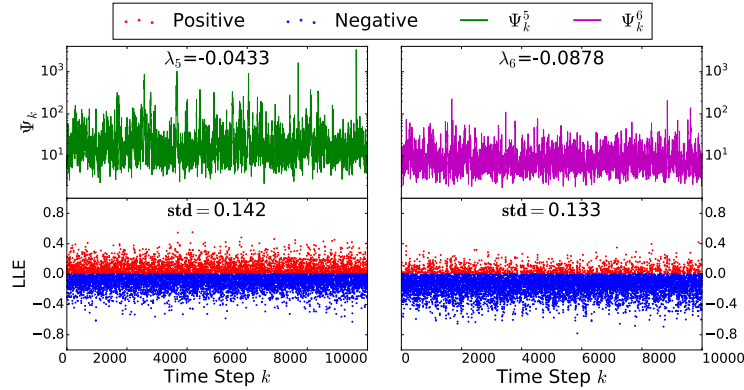
Fig. 1 Upper: time series of Ψ_k^5 and Ψ_k^6 as defined in Eq. (36). **Lower:** LLEs of $\mathbf{L}^{b,5}$ and $\mathbf{L}^{b,6}$. Adapted from Grudzien et al. (2018a)

Result 7: The unfiltered-to-filtered error upwell and the need for inflation

Motivated by the results above, suppose that an approximate, reduced-rank Kalman estimator is defined such that the resulting forecast error covariance and the Kalman gain have image (column) spaces constrained to the span of the leading $r \geq n_0$ BLVs, $\mathbf{L}_k^{b,1:r}$. For perfect and deterministic dynamics, this estimator is asymptotically equivalent to the optimal Kalman filter by the results of Sect. 3.1.1. On the other hand, Result 6 establishes that, in the presence of additive model errors, the variance of the unfiltered error in the span of the trailing $N_x - n_0$ BLVs will remain finite, albeit bounded as per Eq. (35). When $r > n_0$ this reduced-rank Kalman filter will correct the $r - n_0$ stable modes in addition to unstable-neutral modes, thereby reducing the variance of the free forecast errors below the bounds pictured in Fig. (1).

Grudzien et al. (2018b) derive the full forecast error covariance dynamics for the reduced-rank Kalman estimator described above. Note that, as opposed to the free forecast error covariance described in relation to Result 6, the discussion pertains now to the forecast error covariance cycled within the KF. While the standard reduced-rank KF formalism would write the recursion for the forecast error covariance entirely within the span of $\mathbf{L}^{b,1:r}$, it is proven in Grudzien et al. (2018b) that forecast errors in the column span of trailing $\mathbf{L}_k^{b,r+1:N_x}$ (those left "uncorrected" by DA) are transmitted into the column span of $\mathbf{L}_k^{b,1:r}$. This "error upwell" is a consequence of the KF rank-reduction within the first r BLVs and is driven by the upper triangular dynamics of the BLVs in the recursive QR algorithm. Therefore, neglecting the contribution of the "upwelling" of error from the trailing to the leading BLVs, as in the standard recursion, leads to a systematic underestimation of the true forecast error in the presence of additive noise. Furthermore, because the leading r BLVs share the same span as the leading r Oseledet spaces, Eq. (4), the upwelling of errors from the span of the trailing BLVs to the leading BLVs holds for any estimator that is restricted to the span of the leading r covariant subspaces.

Figure 2 presents an example from Grudzien et al. (2018b), using the same tangent linear model from the 10-dimensional Lorenz-96 system as in Fig. 1, and fixing each of $\mathbf{H}_k = \mathbf{R}_k = \mathbf{Q}_k = \mathbf{I}$. In each window, the eigenvalues of the forecast error covariance matrix of the optimal full-rank KF (yellow) and of the "exact", reduced-rank estimator (red) are averaged over 10^5 analysis cycles and plotted with triangles. By exact it is meant here that the full covariance equation is evolved analytically, including the covariance within the unfiltered trailing BLVs and their cross covariances with the leading filtered modes. The rank, r, of the reduced-rank estimator is varied to examine the differences between the forecast error covariances arising from the optimal KF and the reduced-rank one when only the unstable-neutral subspace is corrected by the gain ($r = n_0 = 4$), the first stable mode is corrected by the gain ($r = 5$) and so on. As suggested by Fig. 1, correcting the first stable mode reduces the leading eigenvalue of reduced-rank estimator's forecast error covariance by an order of magnitude versus the case when only the unstable-neutral subspace has been corrected (*cf* the red curves between the two top windows).

In each window the projection coefficients of the reduced-rank estimator's forecast error covariance into the basis of BLVs are also plotted (green line). For the full-
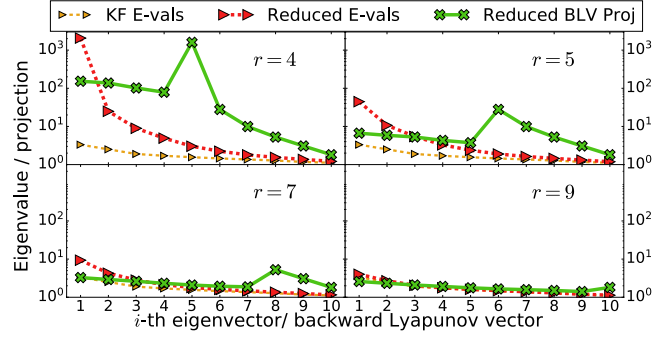
Fig. 2 Eigenvalues of the KF and the reduced-rank estimator covariances plotted with triangles. Projection coefficients of the reduced-rank estimator covariance plotted with X's.

rank optimal KF, the projection coefficients closely follow the eigenvalues and this is not shown due to redundancy. By contrast, it is clear that for the reduced-rank estimator the leading eigenvector is typically close to the first BLV that is contained in the null space of the reduced-rank gain, *i.e.* the first among the unconstrained directions. Figure 2 demonstrates thus a fundamental difference between the perfect model and stochastically forced models configurations: in stochastic systems, using a reduced-rank $n_0 \leq r < N_x$ Kalman estimator, the leading order forecast errors can actually lie in the stable, unfiltered modes. Those modes must be taken thus into account in the DA procedure by, for instance, appropriately enlarging the ensemble size or otherwise augmenting the span of the ensemble-based gain.

The exact recursion for the reduced-rank gain (red lines in Fig. 2) requires resolving the full error covariance dynamics and therefore cannot be used practically for DA. Nevertheless, it was used here to render apparent, the upwelling phenomena: a fundamental source of uncertainty that is not captured by the standard reduced-rank KF (and thus almost all EnKF) recursion. This mechanism is ubiquitous whenever one solves for a reduced rank estimator and is present whenever the forecast error evolution can be well approximated by the tangent-linear dynamics (see Sect. 3.2.2). Notably, it also provides one basic, mathematically grounded, justification for using covariance inflation (Grudzien et al., 2018b), a powerful common fix used in ensemble-based DA (that are commonly rank-deficient by construction) to mitigate for sampling, and sometimes model, error (see *e.g.* Carrassi et al., 2018, their Section 4.4.2 and references therein). Finally, the exact recursion also demonstrates the asymptotic characteristics of the forecast error covariance when using a reduced-rank gain, in the absence of sampling error. It is extremely important to note that in the exact error dynamics for the reduced rank estimator, the leading order forecast errors lie in the directions that are asymptotically stable but unfiltered. This also highlights

the importance of localization (Sakov and Bertino, 2011) and gain-hybridization (Penny, 2017) as effective means for preventing the growth of forecast errors that lie outside of the ensemble span in the EnKF.

3.2 Nonlinear dynamics: the effect of chaos on the ensemble Kalman filter

3.2.1 Perfect and deterministic dynamics

The performance and the functioning mechanisms of the EnKF in nonlinear systems are studied with the aid of numerical simulations performed using the `qgs` code platform (Demaeyer et al., Submitted; Demaeyer and De Cruz, 2020).

We consider first a spectral 2-layer channel quasi-geostrophic atmospheric model. The Fourier modes decomposition is truncated at wavenumber 2 in both meridional and zonal directions on a beta-plane, leading to a set of 20 ODEs for the time evolution of the first 10 components of the atmospheric streamfunction ψ, and temperature θ (Reinhold and Pierrehumbert, 1982). The model dimension is $N_x = 20$ and its spectrum of LEs includes 3 positive and one neutral exponents, so that $n_0 = 4$.

The model is integrated with a time step of approximately 15 minutes and is spun-up for 3 years to ensure the solution has reached the model attractor. Afterward we initialise the DA experiments with the following protocol: a "true" trajectory is computed for $4, 6$ years, and synthetic observations are generated by first sampling this trajectory at regular analysis time t_k, each separated by a time interval Δt, $t_{k+1} = t_k + \Delta t$. The observations are then obtained by adding a zero mean Gaussian random error sampled from $\mathcal{N}(\mathbf{0}, \mathbf{R})$, with $\mathbf{R}$ being the (assumed to be known) observation error covariance matrix. It is assumed that we observe the spectral components directly and that the full system is observed, implying that the observation operator is the identity matrix, $\mathbf{H} = \mathbf{I}_{N_x} \in \mathbb{R}^{N_x \times N_x}$. Although the former hypothesis cannot be met in practice (instrumental devices do not observe the spectral modes) and the latter rarely holds in high-dimensional applications, they are done here for the sake of clarity and will facilitate the study of the dynamical behaviour we intend to discuss. Furthermore, observational error is supposed to be spatially (in spectral space) uncorrelated with an amplitude proportional to the corresponding model variable's standard deviation, $\sigma_{\text{md}}^i, i = 1, \ldots, N_x$. These imply $\mathbf{R} = \sigma^{\%} \text{diag}(\sigma_{\text{md}}^1, \ldots, \sigma_{\text{md}}^{N_x})$

Data assimilation is performed using the EnKF-N, hereafter simply referred to as EnKF (Bocquet, 2011). This EnKF belongs to the family of deterministic filters but it furthermore possesses the appealing property that the ensemble covariance multiplicative inflation is computed automatically as part of the DA process. Inflation is one of the unavoidable feature making the EnKF methods suitable for high dimensional problems (Carrassi et al., 2018). It comes under two ways known as multiplicative and additive inflation, that are often used together. We shall use multiplicative inflation alone because, as opposed to the additive version, it does not change the rank and span of the ensemble error covariance but it only inflates the

matrix entries amplitude. This allows us to study the effect (if any) on the ensemble subspace (reflected into the ensemble covariance rank and span) that comes from the dynamics, without artifacts from the DA procedure. At each analysis step, the forecast anomaly matrix is inflated as $\alpha \mathbf{X}^{\text{f}} \leftarrow \mathbf{X}^{\text{f}}$, $\alpha \geq 1$.

We will study the properties of the EnKF ensemble subspace, its dimension and alignment to the unstable subspace of the underlying dynamics, and will investigate how those will relate to the skill of the EnKF. Following Bocquet and Carrassi (2017), at each analysis time t_k, the alignment between the ensemble and the unstable-neutral subspaces, $\mathcal{U}_k$, is computed as

$$\cos^2(\theta_k^i) = \sum_{p=1}^{n_0} \cos^2(\theta_k^{i,p}) = \sum_{p=1}^{n_0} \frac{\left\{ (\mathbf{u}_k^p)^{\text{T}} \mathbf{v}_k^i \right\}^2}{\left\| \mathbf{v}_k^i \right\|^2} \quad 1 \leq i \leq N, \qquad (37)$$

where $\theta_k^i \in [0, \pi]$ is the angle between the anomaly $\mathbf{v}_k^i$ and $\mathcal{U}_k$, $\mathbf{u}_k^p$ is the p-th BLV, and N is the number of ensemble members.

The RMSE of the EnKF analyses and the angle θ_k^i are shown on the plane $(x, y) = (\Delta t, \sigma^{\%})$ in Fig. 3. Values are averaged over 4 years of simulated time after discarding the first 200 analyses. The RMSE of the analyses is also averaged over all model variables, once the error on each variable is normalised with respect to the corresponding standard deviation σ_{md}. The number of ensemble members is set to $N = 10$.
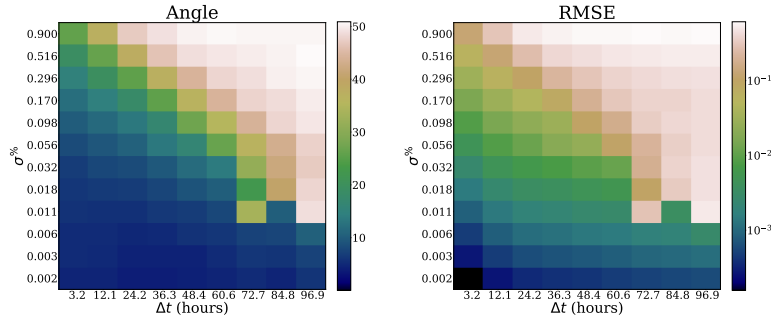
Fig. 3 Atmospheric quasi-geostrophic model. Time- and ensemble-averaged angle, from Eq. (37) (in degree), between the anomalies of the EnKF and the unstable–neutral subspace (left panel, shadow colours, in degree), and time averaged normalised RMSE of the EnKF analysis (right panel, shadow colours), both on the plane $(x, y) = (\Delta t, \sigma^{\%})$. The set-up is $\mathbf{H} = \mathbf{I}_d$, $\mathbf{R} = \sigma^{\%} \text{diag}(\sigma_{\text{md}}^1, \ldots, \sigma_{\text{md}}^{N_x})$ and $N = 10$. Note that the logarithmic scale is used on the right panel.

Figure 3 immediately shows the strong resemblance of patterns between the angle and the RMSE. In practice, whenever the observation time interval Δt and error $\sigma^{\%}$ are small enough then the angle between the two subspaces get smaller and the

RMSE decreases. Similarly to what is discussed in Bocquet and Carrassi (2017), we also see here how increasing observation frequency is more effective in reducing the RMSE than reducing the observation error (see discussion in the Appendix of Bocquet and Carrassi (2017)). This result also suggests that the use of a large number and frequent observations, like the ones produced by non-conventional systems (*e.g.* crowdsourcing) has the potential for improving analyses, despite that observational errors may be larger than in conventional measurement systems (Nipen et al., 2019).

A complementary picture of the relation between the filter skill and the subspaces alignment is provided in Fig. 4, that displays the angle (left y-axis) and RMSE of the EnKF analysis (right y-axis) both against the ensemble size N. The remaining experimental set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathrm{T}}$ with $\sigma^{\%} = 0.08$ and $\Delta t = 12.11$ hours. Recalling that the unstable-neutral subspace has dimension $n_0 = 4$, Fig. 4 shows that as soon as $N \geq n_0 + 1$, *i.e.* the ensemble fully spans the unstable-neutral subspace, the RMSE suddenly reduces to very small values and it does not further decrease when N is increased beyond $n_0 + 1$. The fact that the convergence occurs for $n_0 + 1$ instead of n_0 is due to the ensemble anomalies subspace to be at most of rank $N - 1$ because one degree of freedom is removed when computing the ensemble mean. The behavior depicted in Fig. 4 is peculiar of the EnKF applied to chaotic dynamics and suggests a natural way to reduce computational cost by applying the EnKF with "only" $N = n_0 + 1$ members.
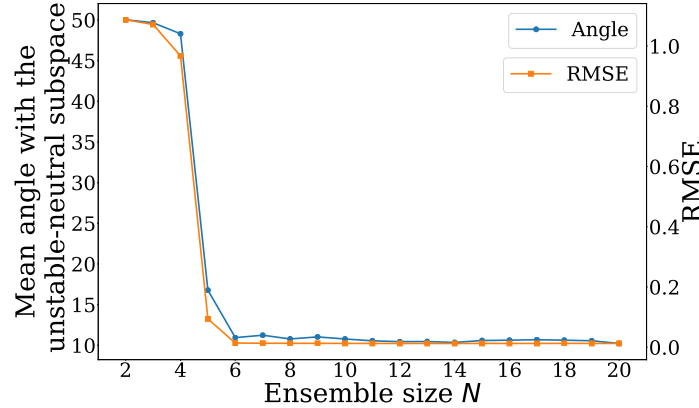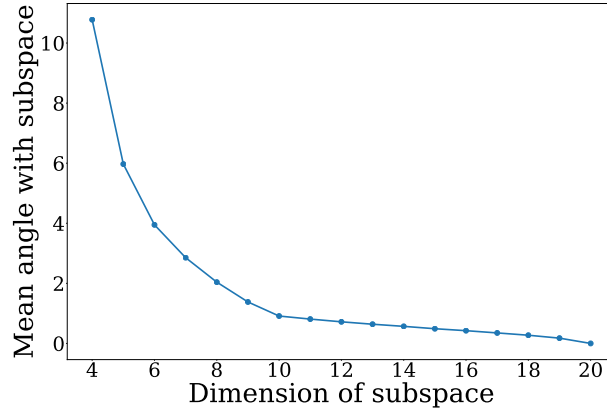

Fig. 4 Atmospheric quasi-geostrophic model. Time- and ensemble-averaged angle, Eq. (37) (in degree), between the ensemble anomalies of the EnKF and the unstable-neutral subspace as a function of the ensemble size N (left y axis), and the corresponding time-averaged RMSE of the EnKF (right y axis). The set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathrm{T}}$ with $\sigma^{\%} = 0.08$ and $\Delta t = 12.11$ hours.

At the convergence, the angle between the ensemble and unstable-neutral subspace (of dimension $n_0 = 4$) is about 10 degrees (see Fig. 4): a remaining small portion of the ensemble subspace is projecting outside the unstable-neutral space. To investigate

how large is such a portion, we compute the angle between an ensemble subspace with fixed $N = 10$ and a Lyapunov/Oseledets subspace of increasing dimension beyond $n_0 + 1$; results are shown in Fig. 5. The angle between the two subspaces decreases monotonically with the size of the subspace, eventually reaching zero once the full phase space ($n = 20$) is considered. Interestingly, the rate of convergence is initially faster until approximately dimension 10, and slower afterwards. This indicates that the additional projection beyond the asymptotic unstable-neutral subspace is largely confined to the less stable (*i.e.* close to neutral) directions, those that can often be locally unstable. This mechanism is a reminiscent of what was shown by (Grudzien et al., 2018b) for stochastic systems and reviewed in Section 3.1.2. It suggests that the addition of few ensemble members beyond $n_0 + 1$ may lead to modest improvements in the filter skill, even though the long term performance is not much impacted.


Fig. 5 Atmospheric quasi-geostrophic model. Time- and ensemble-averaged angle (in degree) between the ensemble anomalies of the EnKF with $N = 10$ members, and subspaces spanned by the BLVs of increasing dimensions. These subspaces are constructed by starting from the unstable-neutral subspace ($n_0 = 4$) and adding one by one stable directions ordered decreasingly by their Lyapunov exponent. The set-up is $\mathbf{H} = \mathbf{I}_d$, $\mathbf{R} = \sigma\sigma^{\mathrm{T}}$ with $\sigma^{\%} = 0.08$ and $\Delta t = 12.11$ hours.

To discuss the impact of multiple timescales on the ensemble Kalman filtering, we consider now the addition of a shallow-water ocean component to the previous atmospheric model. The atmospheric and oceanic models are coupled together through wind and radiative forcing as well as through heat exchanges, yielding the Modular Arbitrary-Order Ocean-Atmosphere Model MAOOAM (De Cruz et al., 2016). MAOOAM has the same 20 ODEs of the atmospheric quasi-geostrophic model, and additional 16 ODEs for the ocean. Amongst these 16 equations, the first 8 govern the time evolution of the first eights components of the oceanic streamfunction ψ_o, while last 8 are relative to the first eights components of the oceanic temperature anomaly δT_o. The model dimension is thus $N_x = 36$ and its parameters are cho-

sen such that a decadal low-frequency variability appears as a consequence of the coupling (Vannitsem et al., 2015; De Cruz et al., 2016).

The model is integrated with a time step of approximately 15 minutes and is spun-up for as many as $18,500$ years to ensure the solution has reached the model attractor, given the low-frequency (*i.e.* slow time-scale) of the model. From this spun-up trajectory we start DA experiments using the EnKF-N with the same protocol used for the atmospheric model detailed above. In this case the true trajectory lasts 185 years, and it is again assumed that we observe the spectral components directly and completely. The EnKF is used in a strongly-coupled DA mode (see *e.g.* Penny and Hamill (2017)) so that, at analysis steps, atmospheric data can impact the ocean and vice-versa. MAOOAM has been already used as a prototypical coupled model to study coupled DA by Penny et al. (2019); Tondeur et al. (2020). Again, observational error is assumed to be spatially (in spectral space) uncorrelated with an amplitude proportional to the corresponding model variable's standard deviation, σ_{md}^i, $i = 1, \ldots, N_x$. These imply $\mathbf{R} = \sigma^{\%} \text{diag}(\sigma_{md}^1, \ldots, \sigma_{md}^{N_x})$.

The spectrum of LEs of MAOOAM is displayed in Fig. 6 and presents some key features.
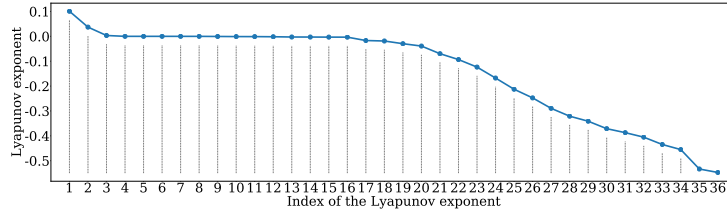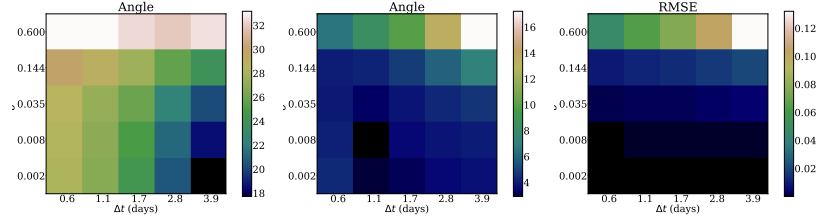

Fig. 6 MAOOAM coupled Ocean-Atmosphere model. Time-averaged Lyapunov exponents in days^{-1} computed along the true trajectory of the DA experiments.

With some arbitrariness, it can be decomposed into three subsets. A first subset of LEs corresponds to the unstable ($\lambda_i > 0$) and neutral ($\lambda_{n_0} = 0$) directions. The dimension of the subspace spanned by these directions is $n_0 = 6$. A second subset of small negative LEs ($\lambda_i \in [-5 \times 10^{-3}, 0[\text{ day}^{-1})$ corresponds to *nearly neutral directions*. The subspace they span has dimension $n_1 = 10$ and we hereafter define the *unstable-near–neutral subspace* as the one spanned by the first $n_0 + n_1$ directions. The presence of these nearly-neutral directions amounts as a form of degeneracy of the neutral direction. Although the model is theoretically possessing only one single neutral mode (*cf* Sect. 2), it is computationally extremely difficult to distinguish it from other nearly neutral ones, and the model can be said to degenerate in the neutral direction in any practical sense. We shall see how this feature has important consequences on the functioning and performance of DA. Finally, after a clear gap, the nearly neutral Lyapunov exponents are followed by the remaining stable directions which form the last subset of the spectrum.

Similarly to Fig. 3, we study the filter performance along with its ensemble subspace alignment to the unstable subspace in Fig. 7. The figure shows the RMSE of the EnKF analyses and the angle θ_k^i given by Eq. (37), between the ensemble subspace and the unstable-neutral (left) and unstable-near–neutral (middle) subspace. Values are averaged over 185 years of simulated time after discarding the first 100 analyses. RMSE of the analysis is also averaged over all model variables, once the error on each variable is normalised with respect to the corresponding standard deviation $\sigma_{\rm md}$. The number of ensemble members is set to $N = 20$.
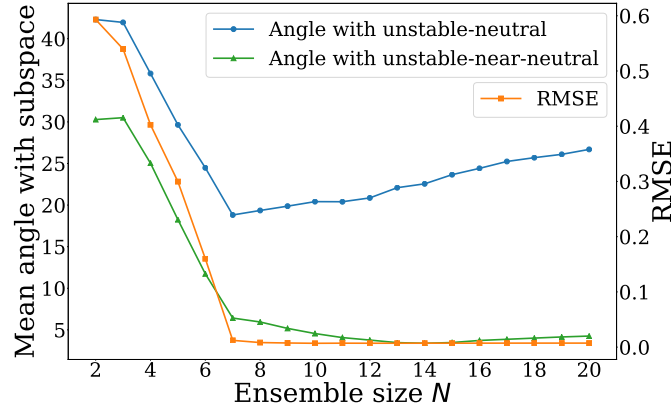

Fig. 7 MAOOAM coupled Ocean-Atmosphere model. Time- and ensemble-averaged angle given by Eq. (37) between the anomalies of the EnKF and the unstable–neutral subspace (left panel, shadow colours, in degree) or the unstable-near–neutral subspace (mid panel, shadow colours, in degree). The time averaged normalised RMSE of the EnKF analysis is also depicted (right panel, shadow colours). All the figures are on the plane $(x, y) = (\Delta t, \sigma^\%)$. The set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \sigma^\% {\rm diag}(\sigma_{\rm md}^1, \ldots, \sigma_{\rm md}^{N_x})$ and $N = 20$. The unstable-near–neutral subspace is defined as follows: it includes the subspace spanned by the unstable and neutral $n_0 = 6$ directions, but also an additional $n_1 = 10$ stable but near–neutral directions with LEs $\lambda_i \in [-5 \times 10^{-3}, 0[\ {\rm day}^{-1}$.

In contrast to what was observed in Fig. 3, we see here that the pattern of the RMSE does not longer resemble the pattern of the angle between the ensemble and unstable-neutral subspace (*cf* left and right panels in Fig. 7). Nevertheless, it bears great similarities with the pattern of the angle between the ensemble and unstable-near–neutral subspace (*cf* mid and right panels in Fig. 7), suggesting that it is now this larger subspace (that includes the n_1 weakly stable modes) that contains most of the error.

This is further confirmed by looking at Fig. 8 that, similarly to Fig. 4, shows the angle (left y-axis) and RMSE of the analysis (right y-axis) against the ensemble size N. The set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \sigma^\% {\rm diag}(\sigma_{\rm md}^1, \ldots, \sigma_{\rm md}^{N_x})$ with $\sigma^\% = 0.08$ and $\Delta t = 1.68$ days.

The critical role of the n_1 weakly stable modes appears now evident when looking at Fig. 8 (*cf* green and blue lines). As opposed to the behaviour of the EnKF applied to the atmospheric model alone, or to the single scale Lorenz96 system used by Bocquet and Carrassi (2017), we see that the ensemble subspace does not longer project much onto the unstable-neutral subspace (blue line with solid circles markers): when $N = n_0 + 1$ the angle reaches its minimum at around 20 degrees (recall that it was about 10 degrees for the atmospheric model, Fig. 4), and it even further increases

Fig. 8 MAOOAM coupled Ocean-Atmosphere model. Time- and ensemble-averaged angle (in degree) between the ensemble anomalies of the EnKF and the unstable-neutral subspace or the unstable-near–neutral subspace as a function of the ensemble size N (left y axis), and the corresponding time-averaged RMSE of the EnKF (right y axis). The set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \sigma^{\%} \mathrm{diag}(\sigma_{\mathrm{md}}^1, \ldots, \sigma_{\mathrm{md}}^{N_x})$ with $\sigma^{\%} = 0.08$ and $\Delta t = 1.68$ days.

when $N > n_0 + 1$, indicating the presence of non-negligible projections outside the $n_0 = 6$ unstable-neutral directions. However, when the angle is computed with respect to the larger unstable-near–neutral subspace (green line with solid triangle markers), we retrieve the match with the RMSE curve (orange line with solid squares markers). A closer inspection further reveals that the angle decreases (the projection grows) fast until $N = n_0 + 1$: the first unstable-neutral $n_0 = 6$ modes still span most of the error. After this initial fast decrease, the angle keeps decreasing at a slower, yet monotonic, rate until approximately $N = 15$ and stays almost constant afterwards. This result demonstrates undoubtedly the importance of the n_1 near–neutral modes. Although possibly asymptotically weakly stable, these directions span a small portion of the filter error that, if included in the ensemble subspace (by properly enlarging the ensemble size) leads to further improvement of the filter performance. This is finally emphasized in Fig. 9, which shows the decrease of the ensemble-averaged angle between an ensemble of $N = 10$ members and the subspaces of increasing dimension beyond the unstable-neutral one. In contrast to Fig. 5, it indicates that the addition of n_1 extra ensemble members may lead to RMSE improvement that is far from negligible. Also, the gap observed around the value $n_0 + n_1 + 1 = 17$ clearly shows that adding more members beyond the range of the near-neutral stability does not bring any benefit. This is a strong indication in favor of a cautious assessment of the ensemble size when working with coupled multi-scale dynamics and performing strongly-coupled DA. Furthermore, as elucidate by Vannitsem and Lucarini (2016) and Tondeur et al. (2020), the near-neutral part of the spectrum in MAOOAM is directly connected to the coupling: including those directions within the ensemble

subspace is paramount to propagate the data information content between ocean and atmosphere. This subspace also proved to be key in producing reliable ensemble forecasts in coupled ocean-atmosphere systems Vannitsem and Duan (2020).
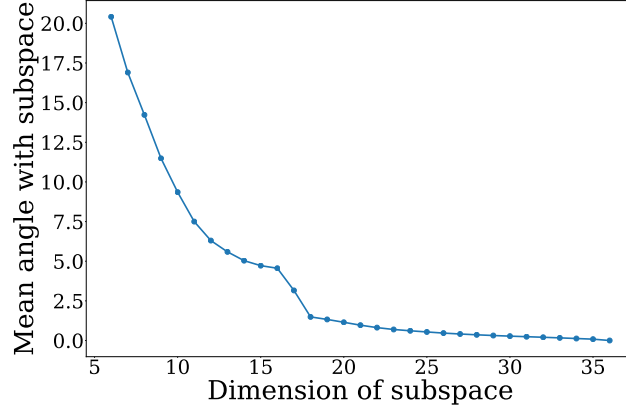
Fig. 9 MAOOAM coupled Ocean-Atmosphere model. Time- and ensemble-averaged angle (in degree) between the ensemble anomalies of the EnKF with $N = 10$ members, and subspaces spanned by the BLVs of increasing dimensions. These subspaces are constructed by starting from the unstable-neutral subspace ($n_0 = 6$) and adding one by one stable directions ordered decreasingly by their LE. The set-up is $\mathbf{H} = \mathbf{I}_{N_x}$, $\mathbf{R} = \sigma\sigma^{\mathsf{T}}$ with $\sigma^{\%} = 0.08$ and $\Delta t = 1.68$ days.

3.2.2 Stochastic dynamics

Results in Sect. 3.1.1 show that the optimal KF and the reduced-rank KF with gain restricted to the leading BLVs are asymptotically equivalent for linear, perfect-deterministic models. Moreover, results in Sect. 3.2.1 demonstrate that in weakly nonlinear error dynamics, with a perfect-deterministic model, these conclusions largely extend to the EnKF. Yet the exact reduced-rank recursion introduced in Sect. 3.1.2 evidenced important differences between the optimal and the reduced-rank formulations in the presence of model error for linear models. In this case, stochastic noise injected in asymptotically stable directions is not entirely damped out and remains finitely bounded. In addition, the upwelling process, albeit inherent to the reduced-rank formulation and not a consequence of model error, will move the constantly injected model noise from unfiltered to filtered directions.

It is of interest thus to compare the differences between the full rank Kalman estimator, and both the standard and the exact reduced-rank KF recursions in a stochastically forced, nonlinear models. As a prototype, we use the model defined by the nonlinear flow of the Lorenz-96 (Lorenz, 1996) system with additive noise,

and $N_x = 40$. Suppose that $t_{k+1} = t_k + \Delta t$, so that the flow map taking all initial conditions to time $+\Delta t$ is defined $\phi_{\Delta t}(\mathbf{x}_k) = \mathbf{x}_{k+1}$. If $\mathbf{x}_k^{\mathrm{t}}$ represents the true physical state at time t_k, we define the dynamical model analogously to Eq. (8a) as a discrete, nonlinear map, $\mathbf{x}_{k+1}^{\mathrm{t}} = \phi_{\Delta t}\left(\mathbf{x}_k^{\mathrm{t}}\right) + \mathbf{w}_k$.

To avoid the interplay and superposition between sampling and model errors, and to be able to focus on the latter alone, instead of the EnKF we use here the EKF (see Sect. 3 and Jazwinski (1970)). The EKF estimates the forecast distribution for $\mathbf{x}_k^{\mathrm{t}}$ via the equation, $\mathbf{x}_{k+1}^{\mathrm{f}} = \phi_{\Delta t}\left(\mathbf{x}_k^{\mathrm{a}}\right)$, taking the analysis mean at time t_k to the forecast mean at time t_{k+1}, and by the linearized forecast equation for the covariance. In the following, the EKF propagates the full-rank covariance equation via the tangent-linear model defined along the mean equation and assimilates observations in all state components. On the other hand, both the standard and the exact reduced-rank formulations restrict the assimilation so that the image space of the gain is equal to the span of the leading r BLVs defined along the tangent-linear model of the mean equation. The difference between the standard and the exact formulation is as follows: the standard formulation only estimates the error covariance in the span of the leading r BLVs while the exact formulation simulates the entire covariance equation, estimating the free forecast covariance in the trailing modes and including its upwelling in the estimate of the uncertainty in the span of leading BLVs (Grudzien et al., 2018b, see proposition 1)

Figure 10, adapted from Grudzien et al. (2018b), illustrates the differences in performance between the above described schemes. On the left, the average RMSE of the (i) *full-rank*, the (ii) *standard reduced-rank* and the *exact reduced-rank EKF* formulations are plotted versus the rank of the reduced-rank gain over 10^5 analyses. The model possesses 13 positive LEs, therefore the reduced-rank EKFs under consideration have at least $r \geq n_0 + 1 = 14$. The system is fully observed with $\mathbf{R} = 0.25\mathbf{I}_{N_x}$. As r approaches $N_x = 40$, the two reduced rank formulations converge to the full rank EKF, with performance considered optimal for a filter in this system. However, for correction rank $r \leq 20$, there are substantial differences in performance between the standard formulation and that which includes the effect of the upwelling of error from the trailing BLVs: the exact formulation reaches both adequate and near-optimal performance with smaller r than the standard recursion.

On the right, Fig. 10 demonstrates the effect of multiplicative inflation on the standard recursion when the correction rank is fixed at $r = 17$. While multiplicative inflation greatly improves the performance of the standard recursion, this performance is actually bounded below by the RMSE of the exact formulation. This example shows that, in addition to sampling error, multiplicative inflation can be used to remedy the inadequacies of the standard reduced-rank formalism which neglects the effects of dynamic upwelling. This dynamic upwelling is a direct byproduct of the estimator being restricted to the span of the leading BLVs. Such an estimator arises when, *e.g.*, the ensemble span aligns with the span of the leading BLVs in the EnKF, as demonstrated in Sect. 3.2.1. The efficacy of dynamic, multiplicative covariance inflation for treating the effect of model error separately from sampling error has also been demonstrated with statistical and optimization methods by, *e.g.*,

Mitchell and Carrassi (2015); Raanes et al. (2015, 2019); Sakov et al. (2018); Fillion et al. (2020). The dynamical upwelling derived in Grudzien et al. (2018b) provides an explanation of one of the mechanisms responsible for the need for covariance inflation.
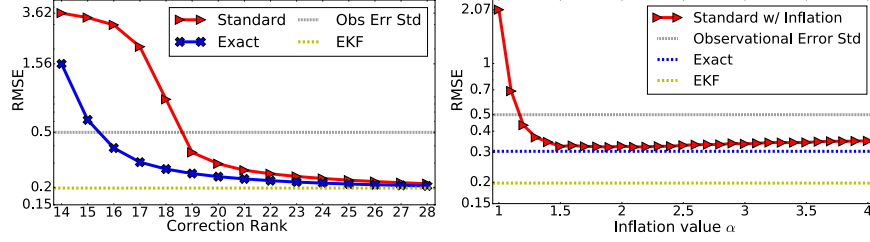
Fig. 10 Left: RMSE of the full-rank, standard reduced-rank and exact reduced-rank EKF. The correction rank of the reduced-rank estimators is varied in the horizontal axis. **Right:** RMSE of the full-rank, standard reduced-rank and exact reduced-rank EKF. Rank is fixed at $r = 17$ and multiplicative inflation in the standard reduced-rank recursion is varied in the horizontal axis

4 Data assimilation for chaotic systems - How chaos became an opportunity

We present here two areas where the knowledge about the chaotic nature and properties of the model dynamics, have been used pro-actively to achieve a better track of the system of interest and to reduce the forecast error. There have been two ways how this has been accomplished. The former has to do with the design of methods to inform where and when additional fewer data would lead to a major improvement in the analysis and forecast skill. This is usually referred to as "adaptive observation" or "target observation", and is the content of Sect. 4.1. The second has to do with incorporating, within the DA process itself, the information about the unstable subspace, with the goal of achieving a computational economy while maximising the error reduction. This gave raise to a family of DA methods known as "assimilation in the unstable subspace", summarised in Sect. 4.2.

Our treatments here are intentionally succinct and have the character of a review, but interested readers will find all the appropriate references to the original works.

4.1 Targeting observations using the unstable subspace

Historically, one very important development in this type of dynamical analysis of the climate predictability / DA problem arose from the question of how to generate adaptive observation systems. By "adaptive observations" is meant here observations whose locations, time and type is chosen such that their impact on the state estimate or on the forecast skill is the largest. It was well understood at the time that the growth of forecast errors was confined to a lower dimensional subspace of rapidly growing perturbations (Trevisan and Legnani, 1995). The Fronts and Atlantic Storm Track (FASTEX) program (Snyder, 1996), a multinational collaboration to investigate the growth and development of frontal cyclones, in particular was motivated by these dynamical approaches to generate adaptive observation schemes that would target regions of rapid forecast error growth. Other similar international efforts have followed, some of them including actual field campaign of measurements: the THORPEX (Fourrié et al., 2006) and the Winter Storm Reconnaissance programs (Szunyogh et al., 2002; Hamill et al., 2013). Two main approaches were considered, the *forced singular vectors* (Buizza et al., 1993; Palmer et al., 1998), and the *bred vectors* (Toth and Kalnay, 1993, 1997), to identify the sensitivity areas of high forecast uncertainty.

Forced singular vectors are generated by the right singular vectors of the forward-in-time, tangent-linear model resolvent $\mathbf{M}$. It can be seen from the earlier discussions in Sect. 2 that the singular vectors can be interpreted as a finite-time approximation of the FLVs along a model forecast. They therefore indicate region where error will grow. On the other hand, the bred vectors are an ensemble-based approach to identify sensitivity regions. Particularly, the breeding scheme simulates how the modes of fast growing error are maintained and propagated through the successive use of short range forecasts in weather prediction. The bred vectors are formed by initializing small perturbations of a control trajectory and forecasting these in parallel along the control. By successive rescaling of the perturbations amplitude back to a small value, this mimics the evolution of small perturbations under the tangent-linear model, and the span of these perturbations generically converges to the leading BLVs. Both of these approaches represent an early and intuitive way to utilizing the ergodic theory of chaotic dynamical systems in designing an effective adaptive observation scheme by targeting an unstable subspace in some form.

Trevisan and Uboldi (2004a); Uboldi and Trevisan (2006) utilized the bred vectors / BLV analysis of Toth and Kalnay (1993, 1997) and explicitly linked the methodology to Lyapunov stability theory. Importantly it was recognised that, instead of mimicking the error growth of the unforced (free) forecast, it was important to track the errors that develop within the DA cycle. This led to a modified version of the breeding approach known as Breeding on the Data Assimilation System (BDAS) Uboldi et al. (2005); Carrassi et al. (2007). The BDAS scheme for adaptive observations is based on the principle of the support of the forecast error lying primarily in the span of the BLVs, with the bred vectors acting as a proxy for the explicit decomposition. In practice the locations of few adaptive observations were selected to be in the areas where the leading BDAS modes attained their local maxima. In experiments with

an atmospheric quasi-geostrophic model, BDAS was used successfully to locate one additional observation at each analysis time of a 3DVar cycle, leading to a dramatic improvement of the analysis skill, compared to cases where either a fixed or a randomly located adaptive observation was assimilated Carrassi et al. (2007).

4.2 Assimilation in the unstable subspace

Data assimilation has long been studied with the trade-off between accuracy and numerical efficiency as a goal. To this end a natural choice has been that of devising reduced-order schemes that focus the observational constraint on smaller, albeit crucial, part of the full dynamics. One of the most celebrated among those approaches is the assimilation in the unstable subspace (AUS) where the DA procedure is explicitly designed to track and control the unstable manifold of the dynamics, usually of much smaller dimension of the full phase space, thus aiming to a reduction in computational cost (Palatella et al., 2013).

We have seen in Sect. 2 that the full phase space of a chaotic dissipative dynamical system can be seen as split in a (usually much smaller) unstable–neutral subspace and a stable one (Kuptsov and Parlitz, 2012). For instance, Carrassi et al. (2007) have shown how a quasi-geostrophic atmospheric model of $O(10^5)$ degrees of freedom possesses an unstable–neutral subspace of dimension as small as $n_0 = 24$.

We have furthermore seen that, in deterministic chaotic systems and under the linear regime of error evolution, the uncertainty in the state estimate converges to zero outside of the unstable-neutral subspace. This phenomenon was at the core of the idea of AUS, whereby the unstable–neutral subspace (or a suitable numerical approximation of it) is explicitly used in the DA scheme to parametrise the description (both temporally and spatially) of the uncertainty in the state estimate (Trevisan and Uboldi, 2004b; Uboldi and Trevisan, 2006; Carrassi et al., 2008a).

The AUS concept has been since then applied to different model scenarios and embedded within either KF-like or variational methods (Palatella et al., 2013). Carrassi et al. (2008b) plugged AUS into a 3DVar cycle in such a way that the observations in the proximity of the leading BDAS mode's maximum were assimilated by imposing that the analysis increment follows the BDAS mode. This implied that the larger the estimated error growth with BDAS, the larger the analysis correction. The combined 3DVar-AUS was extraordinarily more accurate than the 3DVar when the same amount of data were assimilated. AUS was subsequently generalised and embedded into 4DVar (4DVar-AUS, Trevisan et al. (2010)), and in an EKF, (EKF-AUS, Trevisan and Palatella (2011)). The forecast error covariance was projected so as to confine the analysis correction to the unstable-neutral subspace. Remarkably both reduced-rank formulations, 4DVar-AUS and EKF-AUS, showed superior skills than their full-rank counterparts.

AUS relied on the assumption that errors evolve linearly. Going beyond this, Palatella and Trevisan (2015) presented an original way of mixing the contributions from the various Lyapunov vectors such that a quadratic expansions of the error is

considered. This improved the performance of the standard EKF-AUS particularly in regimes of increasing nonlinearities. It remains however to be seen to which extent AUS concept could be used within fully nonlinear DA schemes. This is the subject of Sect. 5.1 and of the references therein. Although AUS has been largely used in a perfect model scenario, Palatella and Grasso (2018) proposed a suitable modification that allows for incorporating parametric model errors. As proved in Grudzien et al. (2018b) and detailed in Sect. 3.1.2 and 3.2.2 the use of AUS in chaotic systems forced by additive stochastic noise would necessarily require the additional inclusion of the asymptotically weakly stable modes.

A key caveat in all of the aforementioned applications of AUS is that one needs to compute in real time the BLVs to be used in the analysis. Therefore, while AUS proved capable to improve accuracy, it did not accomplishes a computational economy, unless the BLVs were all pre-computed and stored. Note however that the latter is not just a technological challenge given that the BLVs depends on the system's state and vice-versa if BLVs are to be used in the analysis update. Thus it is not straightforward to decouple their online estimation and use within the analysis. At the same time however, as we have seen in Sect. 3, AUS concept proved to be very powerful to understand, design and interpret the functioning of the KF and EnKF in chaotic systems.

5 Forward looking

5.1 AUS in non-Gaussian filter?

In this subsection we attempt to improve the performance of the (bootstrap) particle filter (PF, Farchi and Bocquet, 2018) by AUS. The underlying hypothesis is that observational components in the stable subspace contribute little in the way of precision (since nearby orbits within the stable subspace converge), but a lot of noise. Therefore, the investigation will explore whether discarding observational information outside of the unstable subspace can mitigate the acute collapse of weights experienced by PFs in high-dimensional systems, manifesting the "curse of dimensionality". If so, this could be used to reduce the number of particles required, which scales exponentially with some measure of the system size (Snyder et al., 2008). A secondary objective is to investigate the effectiveness of a few different methods of targeting observing systems to the unstable subspace, potentially also reducing costs.

We perform synthetic DA experiments with the Lorenz-96 system. The state size is set to $N_x = 10$ and there is no dynamical noise ($\mathbf{Q} = \mathbf{0}$). Four different observation configurations targeting the unstable subspace are tested. For each of them, observations are taken 0.2 apart in time, with independent noise of variance 1.5. Each experiment lasts for 10^5 analysis cycles. The RMSE averages of each method are tabulated for a range of ensemble sizes and observation operator ranks, N_y, and plotted as curves in Fig. 11. The plotted scores represent the lowest obtained

among a large number of tuning settings, selected for optimality at each point. For the PF the tuning parameters are: (i) the threshold for resampling, which is triggered if the threshold is larger than the effective ensemble size, $\|\mathbf{w}\|^{-2}$, where $\mathbf{w}$ is the vector of weights, and (ii) the bandwidth (scaling) of the regularizing post-resample jitter, whose covariance is computed from the weighted ensemble. For comparison the (symmetric square-root) EnKF (Hunt et al., 2004) is also tested. Its tuning parameters are (i) the post-analysis inflation factor and (ii) whether or not to apply random, covariance-preserving rotations (Sakov and Oke, 2008b).

The top-left panel Fig. 11 shows that the error decreases monotonically in N_y when the observation operator are rows of the identity. The top-right panel shows, by contrast, that when observing the BLVs, $\mathbf{L}_k^{\mathrm{b}}$, the PF with largest ensemble levels off at near-optimal performance with as few as $N_y = n_0 = 4$ observations, only improving marginally thereafter. The same trait can also be noted for lower N, albeit less clearly. Also recall that a similar "levelling off" of RMSE around n_0 occurred in Fig. 4 of Sect. 3.2.1, whose x-axis is N (rather than N_y, as here). This result demonstrates that targeting observations to the directions of dynamical growth of the uncertainty is efficacious. Interestingly, the performance of any given DA method is nearly independent of the observing system (*i.e.* panel) for $N_y = 10$. This makes sense considering that all of $\mathbf{I}$, $\mathbf{L}_k^{\mathrm{b}}$, $\mathbf{V}_{k+1}$, and $\mathbf{U}_k$ are orthogonal, *i.e.* equal up to a rotation.

In all panels of Fig. 11, the RMSE score of the PF (as well as that of the EnKF), for any ensemble size, never degrades by the inclusion of more observations (except for what we adjudge to be noise). This feature was also found in experiments with even smaller ensembles than shown, and experiments with larger observation errors. Thus it seems that the curse of dimensionality is not mitigated by discarding observations. Hence there is little to gain by eliminating observation components outside of the unstable-neutral subspace, apart for the potential computational efficiency in the case that unstable-neutral has been pre-computed.

It should be noted that this finding runs counter to that of Maclean and Vleck (2019); Beeson and Namachchivaya (2020); Potthast et al. (2019), all of which report success in mitigating weight collapse by reducing the observations to the leading components of $\mathbf{L}_k^{\mathrm{b}}$ or some related matrix. Beeson and Namachchivaya (2020) also tested observation operators given by time-local backward and forward Lyapunov vectors, defined through the singular value decomposition of the resolvent: $\mathbf{M}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^{\top}$. They reported better targeting results with $\mathbf{V}_{k+1}$ than with $\mathbf{U}_k$, which in turn was more effective than $\mathbf{L}_k^{\mathrm{b}}$. Our results similarly show $\mathbf{U}_k$ (bottom-right panel) to be slightly more effective for targeting observations than $\mathbf{L}_k^{\mathrm{b}}$ (top-right panel), albeit only for intermediate ensemble sizes. Both, however, are more effective than $\mathbf{V}_{k+1}$ (bottom-left). Moreover, also for $\mathbf{V}_{k+1}$ and $\mathbf{U}_k$, we never observe lower RMSE scores when using fewer observation components. There are some differences in the experimental setting; notably our system is deterministic, and the jitter we apply is restricted to the ensemble subspace. It is unclear if these differences can account for the disparity in conclusions.

Why does PF performance not suffer from the inclusion of further, visibly redundant, observations (contrary to our initial expectation)? Consider the likelihood (*i.e.*
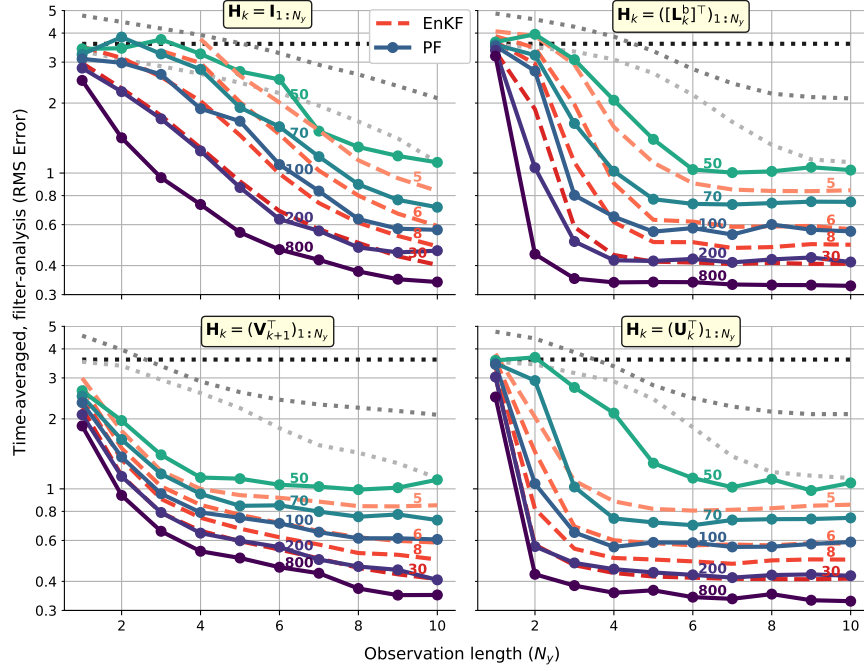
Fig. 11 Benchmarks of filter accuracy (RMSE) from synthetic DA experiments on the Lorenz-96 system, plotted as functions of the observation dimension (N_y). Specifically, the observation operator consists of the N_y leading rows of the transpose of the identity matrix (*top-left*), the BLV matrix, $\mathbf{L}_k^b$, computed by recursive QR decompositions (*top-right*), the 1-cycle forward singular vectors, $\mathbf{V}_{k+1}$ (*bottom-left*), and 1-cycle backward singular vectors, $\mathbf{U}_k$ (*bottom-right*), all of which are defined via the fundamental matrix of the orbit of the (supposedly unknown) truth. The number of members/particles used for the ensemble Kalman filter (EnKF, *dashed*) and Particle filter (PF, *solid*) is tagged above each line. No further improvement is obtained by increasing the ensemble size beyond $N = 30$ for the EnKF and $N = 800$ for the PF, for which the PF achieves better accuracy than the EnKF, as expected for nonlinear systems. The greyscale, dotted lines, included for context, show the performance of baseline methods, whose analysis estimates, $\mathbf{x}^a$, are given by $\overline{\mathbf{x}}$ for Climatology (*black*), $\overline{\mathbf{x}} + \mathbf{K}(\overline{\mathbf{C}})\, [\mathbf{y} - \overline{\mathbf{x}}]$ for Optimal interp. (*dark grey*), $\mathbf{x}^f + \mathbf{K}(c\mathbf{I})\, [\mathbf{y} - \mathbf{x}^f]$ for 3D-Var (*light grey*). Here, $\overline{\mathbf{x}}$ and $\overline{\mathbf{C}}$ are the mean and covariance of the (invariant measure of the) system dynamics, $\mathbf{K}(\mathbf{C}) = \mathbf{C}\mathbf{H}^\top (\mathbf{H}\mathbf{C}\mathbf{H}^\top + \mathbf{R})^{-1}$ is a gain matrix, $\mathbf{x}^f$ is the forecast of the previous $\mathbf{x}^a$, and c is a scaling factor subject to tuning. The plots show that targeting observations to the directions of dynamical growth of the uncertainty is efficacious and that, for this task, $\mathbf{L}_k^b$ and $\mathbf{U}_k$ are similarly effective, and superior to $\mathbf{V}_{k+1}$. Moreover, the RMSE performance of any method, also for the PF (which is our focus), never degrades with the inclusion of more observations.

weighting) of particle $n \in [1, \ldots, N]$, at a given (implicit) time:

$$p(\mathbf{y}|\mathbf{x}_n) = \phi\big(\|\mathbf{y} - \mathbf{H}\mathbf{x}_n\|_{\mathbf{R}}^2\big), \tag{38}$$

where the norm is defined as $\|\mathbf{y}\|_{\mathbf{R}}^2 = \mathbf{y}^\top \mathbf{R}^{-1} \mathbf{y}$, and ϕ is a radial density such that Eq. (38) represents an elliptical distribution, for example a Gaussian. Suppose for

the sake of simplicity that both the observation operator and its error covariance are identity, *i.e.* $\mathbf{H} = \mathbf{R} = \mathbf{I}$, and let $\mathbf{\Pi}$ be the matrix of orthogonal projection onto the unstable-neutral subspace. Now, assuming the PF controls the error in the unstable-neutral subspace, it seems reasonable to assume, given the results for linear systems of Sect. 3, that the particles will converge onto the unstable-neutral subspace: $(\mathbf{I} - \mathbf{\Pi})(\mathbf{x}_n - \mathbf{x}) \to \mathbf{0}$, at least as a first-order approximation. Supposing this, and by decomposing the norm into two orthogonal components, it can be shown that

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}_n\|^2 = \|\mathbf{\Pi}(\mathbf{y} - \mathbf{H}\mathbf{x}_n)\|^2 + \|(\mathbf{I} - \mathbf{\Pi})\mathbf{v}\|^2, \qquad (39)$$

where $\mathbf{v}$ is the observation noise. Thus, projecting the observations onto the unstable-neutral subspace, which would eliminate the last term of Eq. (39), merely reduces the data mismatch by a *constant* (with respect to n, the particle index). Thus

$$p(\mathbf{\Pi}\mathbf{y}|\mathbf{x}_n) = c\, p(\mathbf{y}|\mathbf{x}_n), \qquad (40)$$

for some $c > 0$ which is rendered inconsequential by the subsequent weight normalization. Hence, the inclusion of the remaining observation (outside the unstable-neutral subspace), does not cause a higher variance in the weights, nor the associated tendency to collapse/degeneracy.

The essence of the above reasoning is that the PF prior has zero support, *i.e.* uncertainty, in the stable subspace, and therefore ignores observational components, *including the noise*, in that subspace. This should be contrasted with the situation for dynamical noise (model error), which is "up-welled" from the stable to the unstable subspace, as per Sect. 3.1.2. This distinction exemplifies the difference between uncertainty addition (by dynamical noise) and subtraction (by likelihood updates).

For nonlinear dynamical systems, as illustrated in Sect. 3.2, the particles do not neatly align with the unstable-neutral subspace. Capturing this nonlinear aspect of the flow is generally seen as an advantage of the PF. It might be argued, though, that the observation noise is likely large compared to the spread of the prior particles in the stable subspace, and therefore the observations should be reduced by discarding the corresponding components. However, as highlighted above, the observational noise is constant in the particle index, and therefore its amplitude does not constitute a particular source of weight variability. Instead, the weight variability originates in the prior, which was assumed to have low variability in this scenario.

The analysis resulting in Eq. (40) assumed that $\mathbf{H} = \mathbf{R} = \mathbf{I}$ followed by $\mathbf{H} = \mathbf{\Pi}$. In the case of the observation operator consisting of the leading $N_y \geq n_0$ rows of the transposed BLV matrix, $\mathbf{H}_k = (\mathbf{L}_k^{\mathrm{b}})_{1:N_y}^\top$, the same conclusion can be derived along similar lines. In the general case, for any $\mathbf{H}$, it is not obvious how to reduce the observations so as to only measure the unstable-neutral subspace. To accomplish this, both Maclean and Vleck (2019) and Beeson and Namachchivaya (2020) apply the pseudo-inverse $\mathbf{H}^+$ before their reduction. This can be costly if $\mathbf{H}$ is large. A more practical approach is to reduce the observations as $\hat{\mathbf{y}} = (\mathbf{H}\hat{\mathbf{Q}})^+\mathbf{y}$, or $\hat{\mathbf{y}} = \mathbf{\Pi}\mathbf{y}$ with $\mathbf{\Pi} = (\mathbf{H}\hat{\mathbf{Q}})(\mathbf{H}\hat{\mathbf{Q}})^+$, with $\hat{\mathbf{Q}} = (\mathbf{L}_k^{\mathrm{b}})_{1:N_y}^\top$ In any case, re-doing the same derivation, Eq. (40) again follows, including the same implications for the weights.

In summary, discarding observational information outside of the unstable sub-space does not yield improvements in the PF because it already embodies this "flow-dependent" information. A similar conclusion was also drawn for the iterative ensemble Kalman smoother by Bocquet and Carrassi (2017). Thus, while AUS is a powerful explanatory and diagnostics tool, it is not obvious if it can be used to combat the curse of dimensionality for PFs.

Lastly, this section adds some clarification to the influential paper by Snyder et al. (2008), whose conclusion is sometimes taken to be that the required ensemble size for PFs scales exponentially with *observation* size. Our results rather indicate that the performance of a well-tuned PF will not deteriorate with the inclusion of more observations (even if they are redundant). In other words, that the required ensemble size depends on the rank of the *state* space, or more precisely for chaotic dynamics, the rank of the unstable-neutral subspace. Snyder et al. (2015) points out that the "effective dimension" may be limited by the observation size if this is smaller than the state size. However, in case $N_y < n_0$, no filtering system using flow-dependent priors will be able to achieve satisfactory performance, because the system is not sufficiently observed. Of course, the question of observability is complicated by considering time-dependent observation networks, while localization can also be applied to alleviate the curse of dimensionality (Farchi and Bocquet, 2018).

5.2 Data assimilation and random attractors

AUS and its theoretical extensions provide a framework to interpret the asymptotic inferences of the EnKF. Provided that the forecast anomalies can be considered to be perturbations of the true physical state, and if their evolution can be approximated by the tangent-linear model along the true trajectory, the dynamics of the EnKF anomalies can be decomposed along the Oseledec spaces of the true physical state. The stability and accuracy of the EnKF is largely determined by the ability of the ensemble-based gain to correct the growth of forecast errors in the unstable-neutral subspace, with respect to the uniform-complete observability of these modes. Additive noise complicates the description of the asymptotic forecast uncertainty as uncorrected forecast errors in the span of the stable BLVs may be bounded but impractically large. The upwelling of such errors into the ensemble span furthermore necessitates covariance inflation to rectify the systematic underestimation of the forecast uncertainty in the standard KF-AUS recursion.

Model stochasticity arises systematically in multiscale climate dynamics where there are large scale-separations between resolved and unresolved dynamic processes. In the asymptotic limit of scale-separation, unresolved dynamics can be reduced to additive Gaussian noise due to the Central Limit Theorem; finite scale-separation in reality leads to non-Markovian memory terms in addition to additive stochastic forcing in the exact model reduction of a multiscale model, as in Mori-Zwansig formalism (Gottwald et al., 2015). Several mathematically rigorous frameworks have been developed to model and simulate the effect of small-scale dynamics

on the large-scale dynamics with stochastic parameterization, including averaging methods, perturbation methods and combinations of the two — see, *e.g.*, the survey of approaches by Demaeyer and Vannitsem (2018).

The theory of random dynamical systems offers a novel means of analysis of the DA cycle for multi-scale chaotic systems with large scale separations and model reduction error. Characterizing the asymptotic Bayesian posterior in terms of the properties of a random, nonlinear and ergodic attractor is a natural step forward in the philosopy of AUS. Recent work suggests that the support of the posterior measure of the DA cycle can be asymptotically bounded by the support of the SRB measure in deterministic, nonlinear dynamical systems (Oljača et al., 2018). While this is an intuitively appealing result, the existence of an SRB measure in deterministic dynamics usually requires a hyperbolicity assumption which may not be appropriate in weather and climate, *e.g.* (Vannitsem and Lucarini, 2016). However, many of the theoretical challenges in showing the existence of SRB measures in deterministic dynamics are relaxed in a random dynamical systems setting. Indeed, the Pesin entropy formula holds under very general assumptions for stochastic flows of diffeomorphisms (Liu and Qian, 2006)[theorem 3.1 and discussion on page 127], establishing the link once again between the observed instability in the dynamics and the statistical properties of the invariant measure.

For such a dynamical interpretation of the DA cycle to be credible, the correct specification of random models in stochastic-physical systems is a primary concern; stochastically reduced models should be specified to preserve conservation laws and the original model's dynamics (Cotter et al., 2019). In addition to the correct stochastic model specification, important differences in the statistical properties of model forecasts of stochastic dynamical systems have been observed due to the discretization errors of certain low-order schemes. For example, Frank and Gottwald (2018) develop an order 2.0 Taylor scheme to correct the bias in the drift term induced by the Euler-Maruyama scheme in their study system. Grudzien et al. (2020) likewise find that the bias due to discritization error of the Euler-Maruyama scheme can be sufficient to cause filter divergence of the EnKF in the stochastically forced Lorenz-96 model. Grudzien et al. (2020) emphasize the important role of efficient, weak numerical schemes for the simulation of ensemble-based forecasts. Unlike strong convergent schemes, numerical schemes that converge in the weak sense can make reductions in the complexity of simulation by emphasizing the accuracy of the convergence of the ensemble to the forecast distribution rather than the accuracy of any individual ensemble member.

6 Summary and Conclusion

Chaos is ubiquitous in natural, physical and laboratory systems. Scientists have long coped with this whenever attempting to model, predict or control such systems. Combining and confronting models with data is common in science and *data assimilation* (DA) is the term coined in the context of numerical weather prediction

science to encompass the methods that perform such a combination. The outputs of DA is the improved representation of the system under study, and an estimate of the associated uncertainty.

Inevitably, the sensitivity to initial conditions of chaotic systems, including state-dependence of the directions of error growth, is a challenge for DA. On the other hand, the energy dissipation typical of real systems implies a "dimension reduction" in that errors are confined within a subspace of the full system's phase space.

The DA process requires furnishing a prior distribution, whose specification is a severe difficulty in high dimension. Gaussian methods reduce the complexity to that of estimating the prior mean and covariance. Yet, with the exceptionally high-dimensions of geophysical problems, a proper estimate, and storage of the prior covariance matrices is still challenging. Thus, suitable reduced-rank formulations should be used to lower the computational load while maintaining a good description of the errors about the full system.

In chaotic dissipative systems this goal can be achieved by monitoring the unstable-neutral subspace of the model dynamics and performing DA only within that subspace. This is the idea at the basis of the class of methods known as *assimilation in the unstable subspace* (Palatella et al., 2013), that were developed mainly in the years between 2004 to 2015; we reviewed them in Sect. 4. Despite reducing the problem size to that of the unstable-neutral subspace (of dimension $n_0 \ll N_x$), AUS methods proved aptly skillful, very close to those of their more costly full-rank competitors.

However, the reduction of cost achieved at the analysis steps is offset by the additional cost of monitoring and tracking the unstable-neutral subspace. This requires computing the Lyapunov vectors (usually the backward Lyapunov vectors), which implies computing the tangent linear model (or alternatively evolving an ensemble of bred perturbations mimicking the evolution under the tangent linear model) and a repeated QR matrix decomposition. For specific purposes, one can opt for tracking only a few dominant unstable modes. This was the case for early adaptive (targeted) observation results (Carrassi et al., 2007) or when AUS was used to complement a classical DA method (Carrassi et al., 2008b).

In parallel to the early developments of AUS, several studies with ensemble DA methods in chaotic dissipative systems were suggesting that a number of their key properties were related to the model instabilities, including the rank and span of the ensemble-based forecast error covariance, as well as the skill of the analysis (Sakov and Oke, 2008b; Carrassi et al., 2009; Ng et al., 2011; Bocquet and Sakov, 2014). These results indicate that the ensemble anomalies automatically align with the unstable-neutral subspace, thus resulting in the analysis to be confined to it.

To put this mechanism on a more rigorous theoretical footing, a stream of works have studied the relation between the unstable-neutral subspace in linear systems using the Kalman filter (KF) and Kalman smoother (KS) (Gurumoorthy et al., 2017; Bocquet et al., 2017; Bocquet and Carrassi, 2017). These works, reviewed in Sect. 3.1, have provided analytic proofs that the span of the error covariance matrices of the KF and KS tends asymptotically to the unstable–neutral subspace, independent of the initial condition (*i.e.* no matter the number of the ensemble members, provided

it exceeds the size of the unstable-neutral subspace). For stochastic systems with additive noise it was proved that asymptotically weakly stable modes, that one might discard in deterministic systems, must be included and analytic bounds for the error were provided (Grudzien et al., 2018a).

How do these results hold for nonlinear systems? In the case of chaotic deterministic systems, this was studied in Bocquet and Carrassi (2017) and further in Sect. 3.2.1 of this chapter. It was numerically showed that an ensemble comprising at least as many members as the size of the unstable-neutral subspace plus one ($N \geq n_0 + 1$) is needed to achieve satisfactorily skill with the ensemble Kalman filter (EnKF). Section 3.2.1 also considered the case of a coupled multiscale system with a quasi-degenerate spectrum of Lyapunov exponents. This originates in the presence of many close-to-zero exponents that are related to the coupling mechanisms (Vannitsem and Lucarini, 2016). It is shown that their full inclusion in the ensemble design is needed to reduce the EnKF analysis error to a satisfactorily low level.

The case of nonlinear stochastic chaotic systems with additive noise was studied in Grudzien et al. (2018b) and reviewed in Sect. 3.2.2. The section explains the role of the weakly stable modes already identified in linear systems, but also discovered the *upwelling mechanism* for which uncertainty is upwelled from unfiltered (stable) modes to filtered (unstable) ones. The upwelling phenomenon is not exclusive of nonlinear systems and it is in fact present in linear systems too. It provides an additional rationale to the use of multiplicative inflation, otherwise known by numerically evidence to be needed for a proper functioning of reduced-rank filters even in perfect model scenarios (see *e.g.* Raanes et al., 2019).

An outlook at how this research may evolve is given in Sect. 5. In particular in Sect. 5.1 we provide original results on the use of the AUS approach, *i.e.* exploiting the unstable-neutral subspace, in particle filters (PFs), a fully non-Gaussian DA method. Results indicate that targeting observations within the unstable-neutral subspace is very effective. However, by analogy with what was proved for the EnKF in Sect. 3.2.1, adding observations along the stable modes does not deteriorate the analysis. In the particle filter too, the particles automatically align along the unstable-neutral subspace so that the contribution from observations in its complement stable subspace is negligible. Our results shed also new insight on the scaling of the particle numbers needed to reach convergence. It is shown to depend on the size of the unstable-neutral subspace rather than the observation vector size. In Sect.5.2 we surveyed how the novel concept of random attractors could offer new ways to further exploit the idea behind AUS on stochastic multi-scale systems with large scale separation. Moreover, we also described how to amend numerical integration schemes when doing ensemble DA on such systems, as a trade-off between accuracy and computational cost.

It is important to recall that all of the results with the EnKF and PF that we have presented are obtained without the use of *localization* (see *e.g.* Carrassi et al. (2018), their chapter 4.4, and Farchi and Bocquet (2018) for localisation in the EnKF and PF, respectively). While we are well aware of the dramatic positive impact of localization on the filters' skill, we intentionally did not use it as it artificially

changes the ensemble-covariance rank and span, thus making it impossible to link them exclusively to the model instabilities.

The use of the time-dependent unstable-neutral subspace to represent uncertainty in dynamical systems is still very appealing and potentially prone to success in a wider area than explored so far. For instance, in a recent work by Bocquet et al. (2020) model error arise from parameter mispecification and the EnKF is applied to estimate simultaneously the (chaotic) model state and N_p parameters. The EnKF is used in the state-augmentation formulation and the standard persistence model is adopted for the parameter dynamics. It was shown that the bound for the necessary ensemble size becomes $N \geq n_0 + N_p + 1$: N_p additional members are required to infer the N_p parameters. While the linear one-to-one relation between the number of parameters and that of the additional members is a consequence of the choice of a persistence model and will change if a different parameters dynamics is in place, this result further highlights how much the design of the EnKF is, and can be tied to the properties of the dynamical model.

Future developments along these lines will unavoidably have to tackle the obstacle of the computing cost of the unstable-neutral subspace. We speculate that recent progress in the area of machine learning (Goodfellow et al., 2016) may help. Neural network surrogate models of chaotic systems have shown capabilities to reproduce the spectrum of the asymptotic Lyapunov exponents fairly well (Pathak et al., 2017; Brajard et al., 2020). It is matter of future investigations to explore the possibilities of machine learning algorithms that learn about the time-dependent instabilities from offline long model simulations and then assist the model in the prediction mode by providing a proxy of the unstable-subspace at each analysis time.

References

L. Y. Adrianova. *Introduction to linear systems of differential equations*. American Mathematical Soc., 1995.

M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. SIAM, 2016.

L. Barreira and Y. B. Pesin. *Lyapunov Exponents and Smooth Ergodic Theory*. Student Mathematical Library. American Mathematical Society, 2002. ISBN 9780821829219.

R. Beeson and N. S. Namachchivaya. Particle filtering for chaotic dynamical systems using future right-singular vectors. *Nonlinear Dynamics*, June 2020. doi: 10.1007/s11071-020-05727-y.

G. Benettin, L. Galgani, A. Giorgilli, and J. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory. *Meccanica*, 15(1):9–20, 1980.

M. Bocquet. Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750, 2011.

M. Bocquet and A. Carrassi. Four-dimensional ensemble variational data assimilation and the unstable subspace. *Tellus A*, 69(1):1304504, 2017.

M. Bocquet and P. Sakov. An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.*, 140:1521–1535, 2014.

M. Bocquet, K. S. Gurumoorthy, A. Apte, A. Carrassi, C. Grudzien, and C. K. R. T. Jones. Degenerate Kalman filter error covariances and their convergence onto the unstable subspace. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1): 304–333, 2017.

M. Bocquet, A. Farchi, and Q. Malartic. Online learning of both state and dynamics using ensemble kalman filters. *Foundation of Data Science*, 0:00–00, 2020. doi: 10.3934/fods.2020015. Accepted for publication.

J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. *Journal of Computational Science*, 44:101171, 2020.

R. Buizza, J. Tribbia, F. Molteni, and T. N. Palmer. Computation of optimal unstable structures for a numerical weather prediction model. *Tellus A*, 45(5):388–407, 1993.

A. Carrassi, A. Trevisan, and F. Uboldi. Adaptive observations and assimilation in the unstable subspace by breeding on the data-assimilation system. *Tellus A*, 59 (1):101–113, 2007.

A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi. Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos*, 18:023112, 2008a.

A. Carrassi, A. Trevisan, L. Descamps, O. Talagrand, and F. Uboldi. Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: a comparison with the EnKF. *Nonlinear Processes in Geophysics*, 15:503–521, 2008b.

A. Carrassi, S. Vannitsem, D. Zupanski, and M. Zupanski. The maximum likelihood ensemble filter performances in chaotic systems. *Tellus A*, 61:587–600, 2009.

A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences-an overview on methods, issues and perspectives. *WIREs Clim Change*, e535., 2018. doi: 10.1002/wcc.535.

C. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko. Numerically modeling stochastic lie transport in fluid dynamics. *Multiscale Modeling & Simulation*, 17 (1):192–232, 2019.

L. De Cruz, J. Demaeyer, and S. Vannitsem. The modular arbitrary-order ocean-atmosphere model: MAOOAM v1.0. *Geoscientific Model Development*, 9(8):2793–2808, 2016. doi: 10.5194/gmd-9-2793-2016.

J. Demaeyer and L. De Cruz. Climdyn/qgs: qgs version 0.2.0 release, July 2020.
    URL `https://doi.org/10.5281/zenodo.3941877`.

J. Demaeyer and S. Vannitsem. Stochastic parameterization of subgrid-scale pro-
    cesses: A review of recent physically based approaches. In *Advances in Nonlinear
    Geosciences*, pages 55–85. Springer, 2018.

J. Demaeyer, L. De Cruz, and S. Vannitsem. qgs: A flexible python framework
    of reduced-order multiscale climate models. *Journal of Open Source Software*,
    Submitted. URL `https://raw.githubusercontent.com/openjournals/`
    `joss-papers/joss.02597/joss.02597/10.21105.joss.02597.pdf`.

L. Dieci and E. S. Van Vleck. Lyapunov spectral intervals: theory and computation.
    SIAM Journal on Numerical Analysis, 40(2):516–542, 2002.

L. Dieci and E. S. Van Vleck. Lyapunov and sacker–sell spectral intervals. *Journal
    of dynamics and differential equations*, 19(2):265–293, 2007.

G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag
    Berlin Heildelberg, second edition, 2009a.

G. Evensen. *Data assimilation: the ensemble Kalman filter*. Springer Science &
    Business Media, 2009b.

A. Farchi and M. Bocquet. Review article: Comparison of local particle filters and
    new implementations. *Nonlinear Processes in Geophysics*, 25(4):765–807, Nov.
    2018. doi: 10.5194/npg-25-765-2018.

A. Fillion, M. Bocquet, S. Gratton, S. Gürol, and P. Sakov. An iterative ensemble
    kalman smoother in presence of additive model error. *SIAM/ASA Journal on
    Uncertainty Quantification*, 8(1):198–228, 2020. doi: 10.1137/19M1244147.

N. Fourrié, D. Marchal, F. Rabier, B. Chapnik, and G. Desroziers. Impact study
    of the 2003 north atlantic thorpex regional campaign. *Quarterly Journal of the
    Royal Meteorological Society*, 132(615):275–295, 2006.

J. Frank and G. A. Gottwald. A note on statistical consistency of numerical integrators
    for multiscale dynamics. *Multiscale Modeling & Simulation*, 16(2):1017–1033,
    2018.

G. Froyland, T. Hüls, G. P. Morriss, and T. M. Watson. Computing covariant
    lyapunov vectors, oseledets vectors, and dichotomy projectors: A comparative
    numerical study. *Physica D: Nonlinear Phenomena*, 247(1):18–39, 2013.

M. Ghil and P. Malanotte-Rizzoli. Data assimilation in meteorology and oceanog-
    raphy. In *Advances in geophysics*, volume 33, pages 141–266. Elsevier, 1991.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

G. A. Gottwald, D. Crommelin, and C. Franzke. Stochastic climate theory. *Nonlinear
    and Stochastic Climate Dynamics*, pages 209–240, 2015.

C. Grudzien, A. Carrassi, and M. Bocquet. Asymptotic forecast uncertainty and the
    unstable subspace in the presence of additive model error. *SIAM/ASA Journal on
    Uncertainty Quantification*, 6(4):1335–1363, 2018a.

C. Grudzien, A. Carrassi, and M. Bocquet. Chaotic dynamics and the role of
    covariance inflation for reduced rank Kalman filters with model error. *Nonlinear
    Processes in Geophysics*, 25(3):633–648, 2018b.

C. Grudzien, M. Bocquet, and A. Carrassi. On the numerical integration of the lorenz-96 model, with scalar additive noise, for benchmark twin experiments. *Geoscientific Model Development*, 13(4):1903–1924, 2020.

K. S. Gurumoorthy, C. Grudzien, A. Apte, A. Carrassi, and C. K. R. T. Jones. Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution. *SIAM Journal on Control and Optimization*, 55(2): 741–759, 2017.

T. M. Hamill, F. Yang, C. Cardinali, and S. J. Majumdar. Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Monthly weather review*, 141(6):2058–2065, 2013.

B. R. Hunt, E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin. Four-dimensional ensemble Kalman filtering. *Tellus A*, 56(4):273–277, 2004.

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New-York, 1970.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82:35–45, 1960.

E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 2003.

P. V. Kuptsov and U. Parlitz. Theory and computation of covariant lyapunov vectors. *Journal of Nonlinear Science*, 22(5):727–762, 2012.

B. Legras and R. Vautard. A guide to lyapunov vectors. In *ECMWF Workshop on Predictability*, pages 135–146, Reading, United-Kingdom, 1996. ECMWF.

P. D. Liu and M. Qian. *Smooth ergodic theory of random dynamical systems*. Springer, 2006.

E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

E. N. Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

J. Maclean and E. S. V. Vleck. Particle filters for data assimilation based on reduced order data models, 2019.

L. Mitchell and A. Carrassi. Accounting for model error due to unresolved scales within ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1417–1428, 2015.

G. H. C. Ng, D. McLaughlin, D. Entekhabi, and A. Ahanin. The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*, 63(5):958–977, 2011.

T. N. Nipen, I. A. Seierstad, C. Lussana, J. Kristiansen, and O. Hov. Adopting Citizen Observations in Operational Weather Prediction. *Bulletin of the American Meteorological Society*, 101(1):E43–E57, 10 2019. ISSN 0003-0007. doi: 10. 1175/BAMS-D-18-0237.1.

L. Oljača, J. Bröcker, and T. Kuna. Almost sure error bounds for data assimilation in dissipative systems with unbounded observation noise. *SIAM Journal on Applied Dynamical Systems*, 17(4):2882–2914, 2018. doi: 10.1137/17M1162305.

L. Palatella and F. Grasso. The ekf-aus-nl algorithm implemented without the linear tangent model and in presence of parametric model error. *SoftwareX*, 7:28 – 33, 2018. ISSN 2352-7110.

L. Palatella and A. Trevisan. Interaction of Lyapunov vectors in the formulation of the nonlinear extension of the Kalman filter. *Physical Review E*, 91:042905, 2015.

L. Palatella, A. Carrassi, and A. Trevisan. Lyapunov vectors and assimilation in the unstable subspace: theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46:254020, 2013.

T. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, 55(4):633–653, 1998.

J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott. Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(12):121102, 2017.

S. Penny, E. Bach, K. Bhargava, C.-C. Chang, C. Da, L. Sun, and T. Yoshida. Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, 11(6):1803–1829, 2019.

S. G. Penny. Mathematical foundations of hybrid data assimilation from a synchronization perspective. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(12):126801, 2017.

S. G. Penny and T. M. Hamill. Coupled data assimilation for integrated earth system analysis and prediction. *Bulletin of the American Meteorological Society*, 98(7): ES169–ES172, 2017.

A. Pikovsky and A. Politi. *Lyapunov exponents: a tool to explore complex dynamics*. Cambridge University Press, 2016.

H. Poincaré. *Les méthodes nouvelles de la mécanique céleste. Tome III*. GAUTHIER-VILLARS, 1899.

R. Potthast, A. Walter, and A. Rhodin. A localized adaptive particle filter within an operational NWP framework. *Monthly Weather Review*, 147(1):345–362, 2019.

P. N. Raanes, A. Carrassi, and L. Bertino. Extending the square root method to account for additive forecast noise in ensemble methods. *Monthly Weather Review*, 143(10):3857–3873, 2015.

P. N. Raanes, M. Bocquet, and A. Carrassi. Adaptive covariance inflation in the ensemble kalman filter by gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, 145(718):53–75, 2019.

B. B. Reinhold and R. T. Pierrehumbert. Dynamics of weather regimes: Quasi-stationary waves and blocking. *Monthly Weather Review*, 110(9):1105–1145, 1982.

D. Ruelle. Ergodic theory of differentiable dynamical systems. *Inst. Hautes Études Sci. Publ. Math.*, 50(50):27–58, 1979. ISSN 0073-8301.

P. Sakov and L. Bertino. Relation between two common localisation methods for the enkf. *Computational Geosciences*, 15(2):225–237, 2011.

P. Sakov and P. R. Oke. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371, 2008a.

P. Sakov and P. R. Oke. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3):1042–1053, 2008b.

P. Sakov, J. M. Haussaire, and M. Bocquet. An iterative ensemble Kalman filter in presence of additive model error. *Quarterly Journal of the Royal Meteorological Society*, 2018.

I. Shimada and T. Nagashima. A numerical approach to ergodic problem of dissipative dynamical systems. *Progress of Theoretical Physics*, 61(6):1605–1616, 1979.

C. Snyder. Summary of an informal workshop on adaptive observations and fastex. *Bulletin of the American Meteorological Society*, 77(5):953–961, 1996.

C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008.

C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761, 2015.

I. Szunyogh, Z. Toth, A. V. Zimin, S. J. Majumdar, and A. Persson. Propagation of the effect of targeted observations: The 2000 winter storm reconnaissance program. *Monthly weather review*, 130(5):1144–1165, 2002.

P. D. Thompson. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, 9(3):275–295, 1957. doi: 10.1111/j.2153-3490.1957.tb01885.x.

M. Tondeur, A. Carrassi, S. Vannitsem, and M. Bocquet. On temporal scale separation in coupled data assimilation with the ensemble kalman filter. *Journal of Statistical Physics*, 179:1161–1185, 2020.

Z. Toth and E. Kalnay. Ensemble forecasting at nmc: The generation of perturbations. *Bulletin of the american meteorological society*, 74(12):2317–2330, 1993.

Z. Toth and E. Kalnay. Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125(12):3297–3319, 1997.

A. Trevisan and R. Legnani. Transient error growth and local predictability: A study in the lorenz system. *Tellus A*, 47(1):103–117, 1995.

A. Trevisan and L. Palatella. On the Kalman filter error covariance collapse into the unstable subspace. *Nonlinear Processes in Geophysics*, 18(2):243–250, 2011.

A. Trevisan and F. Uboldi. Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle. *Journal of the Atmospheric Sciences*, 61:103–113, 2004a.

A. Trevisan and F. Uboldi. Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system. *Journal of the Atmospheric Sciences*, 61(1):103–113, 2004b.

A. Trevisan, M. D'Isidoro, and O. Talagrand. Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. *Q. J. R. Meteorol. Soc.*, 136:487–496, 2010.

F. Uboldi and A. Trevisan. Detecting unstable structures and controlling error growth by assimilation of standard and adaptive observations in a primitive equation ocean model. *Nonlinear Processes in Geophysics*, 16:67–81, 2006.

F. Uboldi, A. Trevisan, and A. Carrassi. Developing a dynamically based assimilation method for targeted and standard observations. *Nonlinear Processes in Geophysics*, 12(1):149–156, 2005. doi: 10.5194/npg-12-149-2005.

G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press, 2 edition, 2017. doi: 10.1017/9781107588417.

S. Vannitsem. Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(3):032101, 2017.

S. Vannitsem and W. Duan. On the use of near-neutral backward lyapunov vectors to get reliable ensemble forecasts in coupled ocean–atmosphere systems. *Climate Dynamics*, 55:1125–1139, 2020. doi: 10.1007/s00382-020-05313-3.

S. Vannitsem and V. Lucarini. Statistical and dynamical properties of covariant lyapunov vectors in a coupled atmosphere-ocean model—multiscale effects, geometric degeneracy, and error dynamics. *Journal of Physics A: Mathematical and Theoretical*, 49(22):224001, 2016.

S. Vannitsem, J. Demaeyer, L. De Cruz, and M. Ghil. Low-frequency variability and heat transport in a low-order nonlinear coupled ocean–atmosphere model. *Physica D: Nonlinear Phenomena*, 309:71 – 85, 2015. ISSN 0167-2789.