



Asymptotic behavior of the forecast-assimilation process with unstable dynamics

Dan Crisan, Michael Ghil

► To cite this version:

Dan Crisan, Michael Ghil. Asymptotic behavior of the forecast-assimilation process with unstable dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2023, 33, 10.1063/5.0105590 . insu-04196541

HAL Id: insu-04196541

<https://insu.hal.science/insu-04196541>

Submitted on 5 Sep 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE | FEBRUARY 22 2023

Asymptotic behavior of the forecast–assimilation process with unstable dynamics

Special Collection: [Theory-informed and Data-driven Approaches to Advance Climate Sciences](#)

Dan Crisan   ; Michael Ghil 



Chaos 33, 023139 (2023)

<https://doi.org/10.1063/5.0105590>



View
Online



Export
Citation

CrossMark

Articles You May Be Interested In

Perception of an infrequent assimilation: Labial-to-alveolar assimilation in German

J Acoust Soc Am (May 2008)

Lexical frequency and voice assimilation

J Acoust Soc Am (August 2006)

Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system

Chaos (May 2008)

AIP Advances

Why Publish With Us?



25 DAYS
average time
to 1st decision



740+ DOWNLOADS
average per article



INCLUSIVE
scope

[Learn More](#)

Asymptotic behavior of the forecast–assimilation process with unstable dynamics

Cite as: Chaos 33, 023139 (2023); doi: 10.1063/5.0105590

Submitted: 24 June 2022 · Accepted: 23 January 2023 ·

Published Online: 22 February 2023



View Online



Export Citation



CrossMark

Dan Crisan^{1,a)}  and Michael Ghil^{2,3,b)} 

AFFILIATIONS

¹Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2AZ, United Kingdom

²Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), École Normale Supérieure and PSL University, Paris, France

³Atmospheric and Oceanic Sciences Department, University of California at Los Angeles, Los Angeles, California, USA

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences.

a) Author to whom correspondence should be addressed: d.crisan@imperial.ac.uk

b) Electronic mails: ghil@lmd.ipsl.fr and ghil@atmos.ucla.edu

ABSTRACT

Extensive numerical evidence shows that the assimilation of observations has a stabilizing effect on unstable dynamics, in numerical weather prediction, and elsewhere. In this paper, we apply mathematically rigorous methods to show why this is so. Our stabilization results do not assume a full set of observations and we provide examples where it suffices to observe the model's unstable degrees of freedom.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0105590>

Bjerknes first described weather prediction as an initial-value problem in 1904.¹ As von Neumann and associates started using computers to implement this idea immediately after World War II, it quickly became apparent that the requisite initial data available were incomplete.^{2,3} The appearance of weather satellites in the 1960s led further on to the concept of time-continuous assimilation of remote-sensing data.^{4,5} Nowadays, data assimilation (DA) is being applied across all the areas of the climate sciences and much beyond.^{6–10} Three crucial issues are still insufficiently well understood: (i) standard proofs for the convergence of the DA process rely on the stability of the model dynamics, even in the linear case, while atmospheric and oceanic dynamics are famously unstable and chaotic;^{10–12} (ii) data availability over time appears to successfully compensate for insufficient instantaneous coverage in space;^{2,8,13} and, last but not least, (iii) it appears that observations of a model's unstable manifold are sufficient for the convergence of the time-continuous forecast–assimilation cycle.^{14,15} The present paper uses concepts and methods from the stochastic calculus,¹⁶ random dynamical systems,^{17–19} and nonlinear filtering^{20,21} to achieve several significant steps in clarifying all three of these issues.

I. INTRODUCTION AND MOTIVATION

A. The forecast–assimilation cycle in meteorology

A key metaproblem of data assimilation (DA) in atmospheric, oceanic, and climate problems is to show that sequential filters of various degrees of sophistication are stable and that they converge to solutions with suitable properties that lie sufficiently close to the observations, such as they are, in the case of the *unstable dynamics* that characterizes these problems. Moreover, a full solution to this metaproblem should allow us one to compare, with relative ease, the efficiency and accuracy of several filters.

Heuristically, the motivation for this metaproblem being solvable is the success of practical DA methods in numerical weather prediction (NWP), and in related oceanographic and climate problems, in keeping track of a system's observed state.^{8,22} We outline herein some simple ideas of why sequential filters do have a chance of being stable and convergent, even in the presence of dynamically unstable modes.

As the importance of DA methodology in the climate sciences and an increasing number of other areas, all the way to finance, has been growing rapidly, DA has attracted much attention in relevant

areas of the mathematical sciences.^{20,23–25} As we shall argue further below, this increase of attention has not covered as yet the issues of instability of the basic dynamics that one wishes to track nor that of partial observations. The main point of this paper is to take substantial steps in addressing these two issues with the same degree of rigor as that used so far in addressing the considerably simpler situation of stable dynamics and complete measurements.

To lead up to the full complexity of the setting involved, consider first, for simplicity, a scalar, linear DA problem in continuous time, using “sloppy” notation. The model equation is

$$\dot{x} = Ax + u(t), \quad (1)$$

and the observation equation is

$$y = Hx + v(t), \quad (2)$$

where (u, v) are noises that are white in time t and have “nice” densities—i.e., centered and with finite variance, e.g., Gaussian—while the dot notation $(x, y)^\cdot$ stands for the time derivative. We keep caps (A, H) for the model and observation operators because the scalar case is supposed to be just shorthand for the vector–matrix case. The fully nonlinear and high-dimensional problems arising in the actual applications involve also much more complex noise processes, of course (see, for instance, the work of Nicolis and co-workers in Ref. 26, and further references therein).

The forecast–assimilation (FA) process for \hat{x} , the best linear unbiased estimate of x , obeys^{27,28}

$$\dot{\hat{x}} = A\hat{x} + K(y - H\hat{x}). \quad (3)$$

Here, K is a weight matrix, which equals the Kalman–Bucy optimum^{29,30} in the Gaussian-noise case, and $y - H\hat{x}$ is the *innovation vector* that equals the difference between the actually observed value of x and the one expected by forecasting this value based on past observations.

The intuitive motivation to hope for convergence of this FA cycle to the true evolution—or for its synchronization with the observations³¹—is simply re-writing Eq. (3) as

$$\dot{\hat{x}} = (A - KH)\hat{x} + Ky. \quad (4)$$

This equation exhibits the new, and hopefully stabler, dynamics $(A - KH)$ of the FA process vs the original, pure-evolution dynamics A . It also suggests using random dynamical system (RDS) theory^{17,19} for the FA problem, given the time-dependent forcing by Ky , in which the observations are subject to random errors. An exhaustive presentation of RDS theory is given in the Ludwig Arnold monograph,¹⁷ which makes, however, somewhat difficult reading for the non-specialist. More accessible presentations for DA practitioners can be found in Refs. 18, 32, and 33.

One has to show that the nonlinear, multidimensional—and possibly even infinite-dimensional FA evolution, as in the generalization of Eq. (4) to partial differential equations (PDEs)—is stable^{14,34} even in the presence of dynamic instabilities. For the way that it might still suffice, in the presence of dynamic instabilities, to have $\dim\{y\} \leq \dim\{x\}$, see numerical results for simplified atmospheric and oceanic models in Ref. 22 and in Sec. 2 of the Ghil (1997)³⁵ review paper, for instance. Frank and Zhuk³⁶ did obtain a

mathematically rigorous result along these lines for a deterministic system of nonlinear ordinary differential equations.

We restrict ourselves here to the finite-dimensional case: in the operational practice of numerical weather prediction (NWP), the PDEs governing atmospheric and oceanic flows are discretized in physical space—using finite differences, finite elements, or spectral and pseudo-spectral methods.^{7,10} These days, the number of resulting finite-difference equations in time is very large—up to order of 10^8 – 10^9 —but still finite and will stay so for the foreseeable future.

The nonlinearity of atmospheric and ocean dynamics^{11,37,38} compels us to deal not just with the mean and variance of the estimated state \hat{x} , as in the classical Kalman–Bucy filter,^{27,28} but with the entire probability distribution function (pdf) of the state x , conditioned on the observations z . This pdf may be multimodal or, more generally, non-Gaussian^{21,23} and include the presence of long tails due to extreme events.^{39,40}

To fully describe the pdf of the observed state $x(t)$ given the data $z(t)$, one needs more than this pdf’s mean and variance. Indeed, absent the linearity and Gaussianity assumptions, the system of equations satisfied by the mean and the variance of the pdf of the observed state is no longer closed and we need more than just these two quantities to describe the FA process and compute its evolution. This point of view is, by now, widely shared by the operational NWP community.^{6,41,42}

In this paper, we assume that the model x is a, possibly unstable, stochastic process explicitly defined in Sec. II. We also assume that the model is observed only partially and that the observations arrive continuously in time. The latter assumption is consistent with the already mentioned continuous flow of observations in the satellite era.^{4,5}

In fact, shortly after the advent of meteorological satellites in the late 1960s, Charney *et al.*⁴ formulated the conjecture that a knowledge of the continuous time history of the atmospheric temperature field will allow one to determine the other state variables, in particular, the winds. Ghil and coauthors provided analytical arguments for the correctness of this “Charney conjecture” in two-dimensional (2D) geophysical fluid dynamics (GFD) models⁴³ and documented its usefulness numerically with time-continuous DA of actual remotely sensed temperatures in a fairly realistic NWP-type, three-dimensional (3D) atmospheric model.⁵ Titi and coauthors provided rigorous proofs in a purely deterministic setting for both 2D and 3D models of GFD interest.^{13,44}

Figure 1 here illustrates the FA process’s evolution in the presence of observations that are partial in both their nature—i.e., temperature vs winds—and their spatial coverage—ocean vs land in this simple example. The figure represents DA results using the Kalman–Bucy filter^{27–30} (hereafter KF) for a linear, mid-latitude shallow-water model in one space dimension with a simplified geometry.⁴⁵ In this geometry, there are two data-rich regions of equal size that stand for the North American and Eurasian land masses, alternating with two data-poor regions of the same size that stand for the North Atlantic and North Pacific.

In the particular numerical experiment selected here for illustration purposes, all three model variables—namely, the geopotential height $\phi = gh$ of the free surface, where h is the actual height and g the acceleration of gravity, along with the cartesian velocity components (u, v) —were available at the so-called synoptic times of

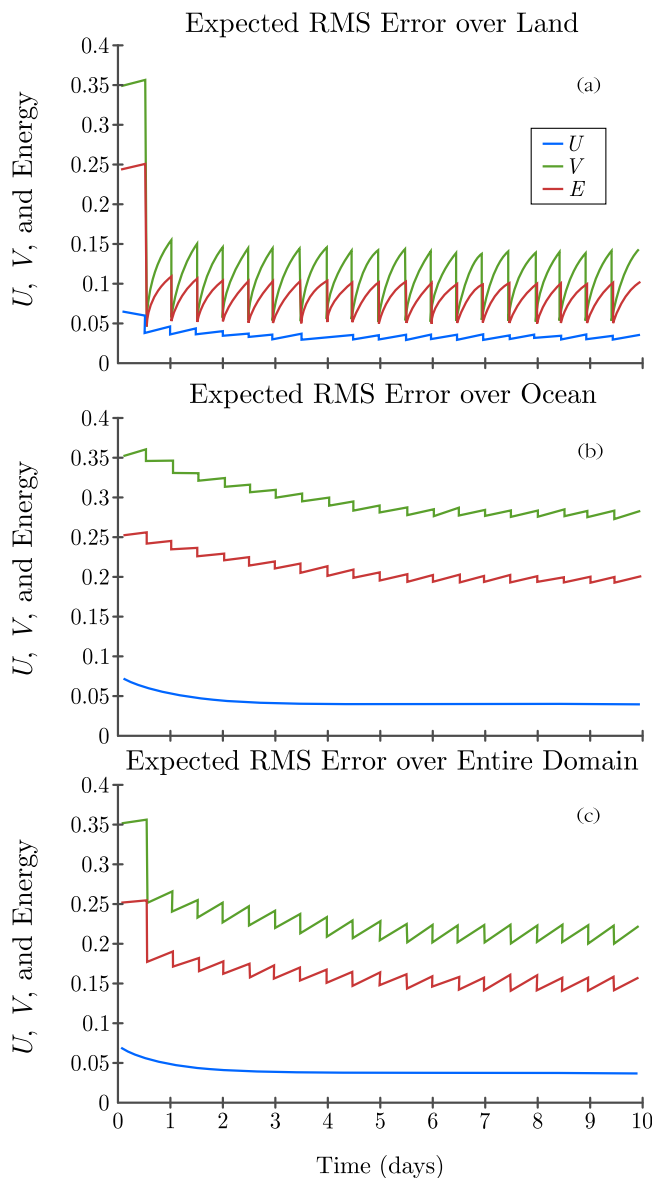


FIG. 1. Typical results of a forecast–assimilation (FA) cycle. (a) Expected root-mean-square (RMS) error over land; (b) expected RMS error over the ocean; and (c) expected RMS error over the entire model domain. Adapted by E. Bach from Ghil *et al.*,⁴⁵ in *Dynamic Meteorology: Data Assimilation Methods*, edited by L. Bengtsson, M. Ghil, and E. Källén (Springer, 1981), pp. 139–224.

noon and midnight GMT over the land areas, while no data at all were available over the ocean areas. In this case, all three curves for $u(t)$, $v(t)$ and $\phi(t)$ —as well as for the energy $E = u^2 + v^2 + \phi^2/\Phi$, where Φ is the equilibrium value of ϕ about which the equations are linearized—have exactly the same behavior.

Over the data-rich land, the error drops sharply at the first observing time, 12 h after the start, and it grows parabolically

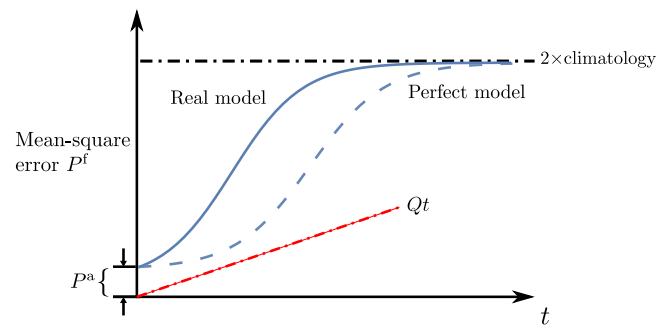


FIG. 2. Schematic diagram of forecast error growth in a numerical weather prediction (NWP) model (see Appendix A for explanations). The perfect-model (dashed blue) and real-model (solid blue) curves here should be compared with the blue curve in the two panels of Fig. 3 there, for $S = 0$ and $S \neq 0$, respectively. The straight red line (dashed-dotted) labeled Qt here represents the linear growth of forecast error variance due to additive white noise, as is the case in Fig. 1(a), between updates and over land, while the Lorenz⁴⁹ and Dalcher and Kalnay⁵⁰ models in the appendix only assume a constant deterministic model error $S \neq 0$. E. Bach kindly provided this figure.

between each observing time and the next one, due to the KF's Gaussian and white-in-time model noise. The parabolic growth of the root-mean-square (RMS) error in this case is due to the additive model noise, as in the case of scalar Brownian motion,⁴⁶ in which it is the variance of the process that is proportional to time t (compare with the dashed red line in Fig. 2). Over land, though, the evolution of the RMS error asymptotes very quickly, in 1–2 model days, to one in which the FA error level oscillates around that of the observational noise, which equals roughly 0.9 in nondimensional units.

Over the data-poor oceans, the RMS error still decreases due to the advection of information from the land areas by the mean westerly winds, $U = -\partial\Phi/\partial y$. However, this error decrease is slower than that over land, the asymptotically periodic behavior is only reached after 4–5 model days, and the mean values of RMS errors stay above the observational noise. Finally, the RMS error behavior over the entire area is essentially a weighted average of the results over land and over ocean.

This type of behavior was modified by the advent of time-continuous satellite data over the oceans, as shown, for instance, by the work of Halem *et al.*,⁴⁷ as part of the DA studies associated with the Global Atmospheric Research Experiment (GARP). While the Ghil *et al.*⁴⁵ model was one-dimensional, linear, stable, and had only a rather small spatial resolution, that of Halem *et al.*⁴⁷ and many others were fully 3D, nonlinear, unstable, and had rather high resolution by the standards prevailing at that point in time.

Figure 5 in Halem *et al.*⁴⁷ (not shown here) clearly indicates the improvement in 6-h forecasts over the Western U.S. from initial states that do use the time-continuous satellite data over the North Pacific vs those that use only the conventional data available over land. The mechanism that advects information by the westerly winds—or misinformation, from ocean to land, in the case of the conventional observing network—is clearly still working, in spite of the unstable and nonlinear dynamics of the forecast model and of the *ad hoc* sequential-estimation method in Ref. 47. The latter was

a successive-correction method^{5,48} in the Halem *et al.*⁴⁷ 3D model, while it was a KF in the Ghil *et al.*⁴⁵ 1D model.

An idealized version of the error components that play a role in the FA process are illustrated in Fig. 2 and discussed further in Appendix A. The major point of the figure and of the associated appendix is that high-end, operational NWP models have various sources of instability, and that NWP would be impossible if DA did not stabilize the FA dynamics and thus achieve real-time low-error forecasts. The main purpose of the present paper is to justify rigorously the numerically observed fact that this is, indeed, the case.

B. A rough sketch of the mathematical formulation

In the present paper's context, the “best estimate” of the model x is its conditional distribution with respect to the observational data $z(t)$ available up to the current time t . We will denote this probability distribution by $\pi(t) \equiv \pi_t$ and the distribution of the model x in the absence of any observational data by $p(t) \equiv p_t$. The pdf p is called the prior distribution and π the posterior distribution, where we dropped for simplicity the dependence on time t . The two distributions π and p can be viewed as dynamical systems that both evolve in the infinite-dimensional space of probability measures \mathcal{P} over the model state space \mathbb{R}^d . Put in simple words, our results refer to the ideal, truly optimal filter and not to any specific approximations thereof, such as the extended Kalman filter (EKF⁸) or ensemble Kalman filter (EnKF^{6,41,42}) or any of the variational methods.^{8,41}

For the comfort of the interested reader coming from the NWP community or from other areas where DA is being used—or its use is being contemplated—we are summarizing in Table I the correspondence of key symbols and terms in this paper vs their counterparts in NWP use. Depending on the model for $x \in \mathbb{R}^d$, the prior system may be unstable: Starting the prior system p from two different initial conditions, the p_0 that results in p_t and $\mu \neq p_0$, and using the same forecast operator for both produces two probability measures, p_t and p_t^μ , that will diverge from each other in time, in the sense that the Wasserstein distance W_2 between the two tends to infinity. The distance W_2 between two probability measures μ and ν is given by

$$W_2(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \right)^{1/2}, \quad (5)$$

where $\Gamma(\mu, \nu)$ denotes the collection of all measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν on the first and second factor, respectively.

TABLE I. Correspondence of symbols and terms in this paper vs their counterparts in NWP.

Notation	Data assimilation language	Stochastic filtering language
p_t	Probability distribution of pure forecast	Prior distribution of the signal
π_t	Probability distribution of the analysis	Conditional distribution of the signal
λ_t^f	Deterministic forecast model M^{f1}	Push-forward operator

This distance is often called in applications the “earth mover’s distance,” following the original motivation of Monge.⁵² Ghil⁵³ originally proposed the idea of using the Wasserstein distance in the context of the climate sciences as a way to generalize the traditional concept of equilibrium climate sensitivity³⁷ in the presence of a time-dependent forcing, such as seasonal or anthropogenic forcing. Further details are given in Sec. 1 of Appendix B and considerably more information on the definitions and methods used herein can be found in the Panaretos and Zemel monograph.⁵⁴

As stated already several times, we concentrate on the prior process’s p_t being unstable, i.e., starting it from the initial condition $\mu \neq p_0$ will lead to divergence of the trajectory p_t^μ from p_t in Wasserstein distance W_2 . We show in Sec. III that, to the contrary, starting the posterior process π_t^μ from the initial condition $\mu \neq p_0$ and evolving it with the same FA operator as that used to evolve π_t will generate a probability measure π_t^μ that will keep the W_2 distance between π_t^μ and π_t bounded in expectation. Moreover, in the linear case, we show that the W_2 distance between π_t^μ and π_t actually tends to zero.

Mathematically, we consider the forecast operator λ_t^f and the FA operator λ_t associated with the FA process, respectively, which are defined by

$$\lambda_t^f : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d), p_t = \lambda_t^f p_0, \quad (6a)$$

$$\lambda_t : \Omega \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d), \pi_t(\omega) = \lambda_t(\omega) \pi_0, \quad (6b)$$

and we show that, under certain conditions,

$$\sup_{t \in [0, \infty)} \mathbb{E}[W_2(\lambda_t \mu, \lambda_t \pi_0)] < \infty, \quad (7)$$

while possibly $\lim_{t \rightarrow \infty} W_2(\lambda_t^f \mu, \lambda_t^f p_0) = \infty$. Moreover, in the linear case, we will show that $\lim_{t \rightarrow \infty} W_2(\lambda_t \mu, \lambda_t \pi_0) = 0$.

Note that, in Eq. (6) above, both the push-forward operators λ_t^f and λ_t are stochastic processes and act on the full probability measures p and π , respectively. In the present framework, we do not limit ourselves just to the mean and variance of the state $x(t)$, as is the case for the linear KF.^{27,28} The distance $W_2(\lambda_t \mu, \lambda_t \pi_0)$ between the RDS $\lambda_t \mu$ starting from μ and the RDS $\lambda_t \pi_0$ starting from π_0 is, of course, random: It depends on the realization of the observation process, and, hence, the expectation in Eq. (7) is taken with respect to the pdf of this process.

We will prove rigorously, in the precise sense described above, that the incorporation of observational data into the FA process does indeed have a stabilizing effect on unstable dynamics, as shown abundantly by NWP practice and suggested, in particular, by the work of Anna Trevisan and her collaborators on assimilation in the unstable manifold (AUS).^{14,15,34} The latter work was one source of inspiration for the present paper.

Another source was RDS theory, according to the arguments presented by one of the authors (MG) at the “Symposium Honoring the Legacy of Anna Trevisan” held in Bologna, Italy in October 2017. A more specific source of inspiration for the rigorous mathematics herein was the paper of Arnold⁵⁵ on stabilization by noise, which was presented to MG by Franco Flandoli during the trimester on “The Mathematics of Climate and the Environment,” held at the Institut Henri Poincaré in Paris in Fall 2019.

It turned out, however, rather quickly that the latter paper's arguments could not be applied directly to the DA problem at hand, since the deterministic component of the process under study there is linear, and we did not see how to extend Arnold's arguments⁵⁵ to fully nonlinear processes. Moreover, the latter arguments are only valid for finite-dimensional systems, while we are dealing here with RDSs that evolve in the infinite-dimensional space of probability measures $\mathcal{P}(\mathbb{R}^d)$.

The layout of the paper is dictated by the intent to bring the two communities—of DA practitioners, on the one hand^{2,8,14,23}—and of the rapidly increasing numbers of applied mathematicians and physicists interested in DA on the other^{21,24,56}—closer together. Hence, after this fairly long introduction, we describe in Sec. II the precise mathematical framework that is used herein. The main rigorous results are outlined in Sec. III and further details on definitions and proofs appear in Appendix B. Conclusions and some thoughts on further work appear in Sec. IV.

II. MATHEMATICAL FRAMEWORK

A. Prior results

The two stabilization theorems presented here, Theorems 3.1 and 3.2, are related to and do use in their proofs certain arguments from the existing stability results in the nonlinear filtering literature, e.g., Refs. 57–65. Such results have not, by and large, matched the objectives of the applied DA community when studying the asymptotic behavior of the FA process.

To be more precise, the applied DA community is interested in results for a forecast cycle that is unstable—as is the case in meteorology and oceanography—and for which applying DA has the mysterious but salutary effect of stabilizing the FA process. In addition, the DA community has to rely, typically, on results where only a small subset of the forecast cycle's degrees of freedom can be observed. Without being comprehensive, of course, we give here a classification of the conditions under which the existing results on filter stability for *nonlinear* forecast processes hold.

- The forecast process is assumed to be ergodic or to have good mixing properties, e.g., Atar,⁶⁶ Atar and Zeitouni,⁶⁷ Budhiraja,⁵⁷ Chigansky and Liptser,⁶⁸ Chigansky, Liptser and Van Handel,⁶⁹ Del Moral and Miclo,⁶¹ and Del Moral, Doucet and Singh.⁶⁰ Please note that ergodicity in these papers is meant in the sense of convergence of the pdf of the push-forward process \mathcal{L}^f as defined in Eq. (6a), not in the pathwise sense of RDSs.^{17,33,70} For example, the stochastic Lorenz model⁷⁰ is ergodic in the pathwise sense of RDSs, but does not have a stable push-forward process \mathcal{L}^f . Note that ergodicity of the pdf of the push-forward process \mathcal{L}^f does imply its stability. For further results on ergodicity of the push-forward process, see, for instance, Ref. 71.
- The forecast cycle is fully observed, e.g., Refs. 64, 72, 73, 69, and 65.
- The forecast process has a drift term, which is the gradient of a convex function, or a perturbation thereof, and its noise term is strictly elliptic, e.g., Refs. 73, 74. This is a very restrictive class of forecast processes, which are expected to be stable (see Ref. 75). Specific drifts belonging to this class are used

in the classical Metropolis-adjusted Langevin algorithm known to converge faster to the diffusion's invariant measure, e.g., Ref. 76.

To try to answer the question raised by the applied community, we no longer insist on proving that the FA process is (exponentially) stable in the sense advocated by the theoretical community. We relax the definition and only require that the FA process initialized from the wrong distribution does not diverge too strongly from the correctly initialized FA process, even when the forecast process does so; this is why the ergodicity assumption for the forecast process is not useful.

We substantially strengthen, however, the stabilization result, in the sense that we want to control the mean and the second moment of the FA process. Again, this is needed for practical reasons. The practitioners want to know that the pointwise estimate of their algorithm of choice does not diverge from the theoretical mean of the FA process. However, they also want to know that their error bars are not too different from the theoretical ones.

Furthermore, the results presented herein have the advantage that the signal $X(t)$ is not required to be fully observable. In particular, the dimension n of the observation $Y(t)$ and the dimension d of the signal $X(t)$, in the notation of Sec. II, do not need to coincide; in operational NWP and many other applications, $d \ll n$.^{2,23}

Heuristically, we need to be able to observe all the “unstable” directions, as suggested by Trevisan and co-workers,^{14,15,34} who gave several fairly realistic examples of this idea working quite well. The connection with our results is provided by some simple illustrative examples in Sec. 4 of Appendix B.

In the linear case, the stability of the FA process is better understood (see, for instance, Refs. 63, 64, 65, and 74). The stability of the associated Riccati equation has been studied by Bishop and Del Moral.^{77,78} The stabilization result obtained in Theorem 3.2 below for the linear signal is weaker than many of the existing results discussed above. For example, no (exponential) rates of convergence are deduced herein. This situation clearly leaves considerable room for proving stronger results, given the hypotheses of Theorem 3.2.

To complete this section, we would like to mention a number of works^{79–83} that explore related properties of the FA process for a discrete-time framework. Bocquet *et al.*⁸⁰ deal with convergence of the Kalman filter covariance matrix for a linear observation operator and a linear and error-free dynamical model. Bocquet and Carrassi⁷⁹ introduced further refinements of the original results, as well as numerical validations in both the linear and the nonlinear case. Gurumoorthy *et al.*⁸³ also studied this problem and showed that the Riccati equation for the Kalman filtering error covariances provides asymptotical bounds on the rank of the forecast and the analysis error covariance matrices; both of these ranks are less than or equal to the number of the forecast process' non-negative Lyapunov exponents. These works are well summarized by Carrassi *et al.*⁸¹ in the recent collective work,⁸⁴ a book that presents some of the latest DA developments in various disciplines of the geosciences. These results are important and could perhaps be exploited to show a similar stabilization of the FA process in the respective frameworks under which they are valid.

B. The present setting

Having presented in Sec. II A a quick review of previous mathematical results on the stability of the FA process, we proceed now by introducing the setting of our two theorems in Sec. III. Let $X = (X^i)_{i=1}^d$ be the solution of the following stochastic differential equation driven by a p -dimensional Brownian motion process $V = (V^j)_{j=1}^p$,

$$X_t = X_0 + \int_0^t f(X_s) ds + \int_0^t \sigma(X_s) dV_s. \quad (8)$$

Here, we assume that $f = (f^i)_{i=1}^d : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma = (\sigma^{ij})_{i=1,\dots,d,j=1,\dots,p} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times p}$ are globally Lipschitz. This will ensure that Eq. (8) has a unique solution.

As stated in Sec. I B, we are interested in tracking the evolution of the full pdf of the prior and posterior processes, namely, p_t and π_t , respectively. To do so, we recall that the process X is a diffusion process with an infinitesimal generator given by

$$A\varphi = \sum_{ij} \frac{1}{2} a_{ij} \partial_i \partial_j \varphi + \sum_i f_i \partial_i \varphi,$$

where $a_{ij} = \sum_k \sigma_{ik} \sigma_{jk}$. The prior distribution of X_t is also called its law in the context of filtering.²¹

For an arbitrary measurable function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ that is integrable with respect to the law of X_t , one has

$$p_t(\varphi) = \mathbb{E}[\varphi(X_t)].$$

By restricting φ further to lie in a suitably chosen space of functions denoted by $\mathcal{D}(A)$,

$$p_t(\varphi) = p_0(\varphi) + \int_0^t p_s(A\varphi) ds. \quad (9)$$

Let $\lambda_t^f : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d)$ be the push-forward operator associated with Eq. (9). In particular, for arbitrary $\mu \in \mathcal{P}(\mathbb{R}^d)$, the map $t \rightarrow \lambda_t^f \mu =: p_t^\mu$ is the solution of (9) with the initial condition μ .

Let Y be an n -dimensional process, where n is the number of observations that are taken to be one-dimensional and have measurement noises modeled by independent Brownian motions,

$$Y_t^i = Y_0^i + \int_0^t h^i(X_s) ds + W_t^i, \quad i = 1, \dots, n, \quad (10)$$

and $h_i(x)$ are the corresponding observation operators, which generalize the observation matrix from the linear framework (see Sec. 3 of Appendix B for details).

Let $\pi = \{\pi_t, t \geq 0\}$ be probability measure-valued process that gives us, at time $t \geq 0$, the conditional distribution of the signal X_t given the observations accumulated up to time t , $\{Y_s, s \in [0, t]\}$. It is this process that we call the FA process in our rigorous mathematical setting. In other words, π_t satisfies

$$\pi_t(\varphi) = \mathbb{E}[\varphi(X_t) | Y_s, s \in [0, t]],$$

where $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ is an arbitrary measurable function that is integrable with respect to the law of π_t . The FA process satisfies the

following stochastic partial differential equation, formulated here in the integral form

$$\begin{aligned} \pi_t(\varphi) &= \pi_0(\varphi) + \int_0^t \pi_s(A\varphi) ds \\ &\quad + \int_0^t (\pi_s(\varphi h^\top) - \pi_s(h^\top) \pi_s(\varphi)) (dY_s - \pi_s(h) ds) \\ &= \pi_0(\varphi) + \int_0^t \pi_s(A\varphi) ds + \int_0^t (\pi_s(\varphi h^\top) - \pi_s(h^\top) \pi_s(\varphi)) dI_s \end{aligned} \quad (11)$$

for any test function $\varphi \in \mathcal{D}(A)$. Here, I is the innovation process, defined as

$$I_t^i = Y_t^i - \int_0^t \pi_s(h_i) ds, \quad i = 1, \dots, n, \quad t \geq 0,$$

which is the rigorous analog of the innovation vector in Eq. (3) of Sec. I A. The innovation process is a Brownian motion (see, e.g., Chap. 3 in Ref. 20). In particular, it is a martingale, and stochastic integrals with respect to the innovation process are easier to manipulate. In particular, upper bounds for the stochastic integrals with respect to martingales are easier to obtain than those for stochastic integrals with respect to general semi-martingales. The observation process is a semi-martingale (it is a Brownian motion plus a drift term), which is harder to handle.

More on stochastic partial differential equations and their difficulties can be found, for instance, in Ref. 85 or Ref. 86.

In order to analyze the asymptotic behavior of the FA process, we can recast the solution of (11) as an RDS. More precisely, there exists a measurable map $\lambda : [0, \infty) \times \mathcal{P}(\mathbb{R}^d) \times \Omega \rightarrow \mathcal{P}(\mathbb{R}^d)$, $(t, \mu, \omega) \mapsto \lambda(t, \omega)\mu$ such that $\lambda(0, \omega) = I$, namely, the identity map on $\mathcal{P}(\mathbb{R}^d)$, and

$$\lambda(t+s, \omega) = \lambda(t, \vartheta_s \omega) \circ \lambda(s, \omega) \quad (13)$$

for all $t, s \in [0, \infty)$ and for all $\omega \in \Omega$. In (13) the symbol \circ means map composition. A family of maps $\lambda(t, \omega)$ satisfying (13) is called a cocycle, and (13) is the cocycle property. The map $\{\vartheta_t : \Omega \rightarrow \Omega\}, t \in [0, \infty)$ is a family of measure-preserving transformations of a probability space (Ω, \mathcal{F}, P) termed the shift operators (see, for instance, Sec. 2.5 in Ref. 16 for further details on the shift operators).

Using this map, the solution of (11) can be expressed as

$$\pi_t(\omega) = \lambda(t, \omega)\pi^0. \quad (14)$$

Moreover, for an arbitrary $\mu \in \mathcal{P}(\mathbb{R}^d)$, the process $\pi^\mu = \{\pi_t^\mu, t \geq 0\}$, defined as

$$\pi_t^\mu(\omega) := \lambda(t, \omega)\mu,$$

is the solution of the SPDE (11) with initial condition μ . Finally, the map $\lambda(t, \omega)$ is a continuous map when we endow $\mathcal{P}(\mathbb{R}^d)$ (or, rather, the set of probability measures with second moment) with the topology induced by the Wasserstein metric.

Since the FA process is infinite-dimensional, as explained in Sec. I B, its RDS characterization is not immediate. RDS theory is well developed for finite-dimensional processes.¹⁷ A subclass of these questions is settled in a fairly satisfactory manner by the theory of

stochastic flows (see, for instance, Refs. 87 and 88). A stochastic flow needs jointly continuous dependence of the solutions of the stochastic differential equation under consideration on time and on the initial state, except for a set of measure zero. This often does not hold for infinite dimensions. Some infinite-dimensional systems do generate a stochastic flow, others do not. For further definitions of possibly infinite-dimensional RDSs, as well as for other related results, we refer to Refs. 19 and 89. The RDS characterization of the FA process is discussed in Sec. 5 of Appendix B.

In this paper, we show that, despite the possible divergence of the prior distributions, that is,

$$\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t) = \infty,$$

the FA process has a stabilizing effect, in the sense that it keeps the distance $W_2(\pi_t^\mu, \pi_t)$ uniformly bounded in expectation. Moreover, in the linear case, it makes the distance $W_2(\pi_t^\mu, \pi_t)$ vanish asymptotically. The main results of the paper are Theorems 3.1 and 3.2 below.

III. MAIN RESULTS

We introduce now the Wasserstein metric on $\mathcal{P}^2(\mathbb{R}^d)$ defined by (5) on the set of all probability measures on the collection of Borel sets $\mathcal{B}(\mathbb{R}^d)$ that have a finite second moment. Recall that the set $\Gamma(\mu, \nu)$ in (5) denotes the collection of all measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν on the first and second factor, respectively; it is called the set of all couplings of the measures μ and ν .

The Wasserstein metric is equivalently defined by

$$W_2(\mu, \nu) = (\inf \mathbb{E}[|X - Y|^2])^{1/2}, \quad (15)$$

where $\mathbb{E}[Z]$ denotes the expected value of a random variable or vector Z and the infimum is taken over all joint distributions of the random variables X and Y with marginals μ and ν , respectively.

The main results of the paper are:

Theorem 3.1: For nonlinear coefficients f , h , and σ and measures π_0 and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ that satisfy the conditions stated in Appendix B 2, there exists a bound $R = R(\pi_0, \mu)$ such that

$$\sup_{t \geq 0} \mathbb{E}[W_2(\pi_t^\mu, \pi_t)] \leq R. \quad (16)$$

The complete proof of this theorem is given in Appendix B 2. For the benefit of the curious but hasty reader, we provide here a brief sketch of the argument. First, we give a bound on the difference $\hat{\pi}_t^\mu - \hat{\pi}_t$ between the mean of the FA process initialized from μ and the original FA process, initialized from π_0 . This is done in two steps: For arbitrary $\delta > 0$, we deduce that there exists a constant c_δ independent of k such that

$$\sup_{t \in [k\delta, (k+1)\delta]} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] \leq c_\delta \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{x}_{k\delta}|].$$

This inequality appears as Eq. (B10) in Appendix B.

Next, we show that there exists R_δ such that $\sup_{k \geq 0} \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{x}_{k\delta}|] \leq R_\delta$. These two inequalities give us a uniform bound, over $t \in [0, \infty)$, of the difference $\mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|]$. Finally, the uniform bound, for all positive times, of $\mathbb{E}[W_2(\pi_t^\mu, \pi_t)]$ comes by means of Lemma 2.1 in the appendix from the bound on the difference

$\mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|]$ and that of the covariance matrices of the FA process initialized from μ and, respectively, the original FA process, initialized from π_0 .

Theorem 3.2: For linear coefficients f , h , and σ and measures π_0 and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ that satisfy the conditions stated in Appendix B 3, we have the much stronger result that

$$\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t) = 0. \quad (17)$$

The complete proof of Theorem 3.2 is given in Appendix B 3. Again, we provide here a brief sketch of the argument. We define $\pi_t^0 := N(\hat{x}_t^0, P_t^0)$ to be a suitably chosen probability measure process that serves as a “reference.” Then, we show that π_t^μ gets asymptotically close to the reference process π_t^0 , regardless of the initial condition, and since this holds true for $\mu = \pi_0$ too, we immediately deduce that $\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t) = 0$.

The reference process π^0 is convenient to work with: its centered version $N(0, P_t^0)$ converges weakly, as well as in Wasserstein distance, to $\pi^\infty = N(0, P_\infty)$. Using the equivalent definition of the Wasserstein distance (15), we deduce that $\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t^0) = 0$, if and only if the following three properties hold true:

- $\lim_{t \rightarrow \infty} |\hat{\pi}_t^\mu - \hat{x}_t^0| = 0$;
- $\lim_{t \rightarrow \infty} |P_{\pi_t^\mu} - P_t^0| = 0$; and
- $\lim_{t \rightarrow \infty} |\pi_t^\mu(\varphi_t) - \pi_t^0(\varphi_t)| = 0$ for any bounded uniformly continuous function φ , where φ_t is the same function shifted by the mean \hat{x}_t^0 , that is, $\varphi_t(x) := \varphi(x + \hat{x}_t^0)$, $x \in \mathbb{R}^d$.

These three properties are then shown to hold, thus completing the proof.

Remark 3.1: In the linear case of Theorem 3.2, it is fairly easy to verify that the Kalman–Bucy filter^{27,28} satisfies the assumptions of the theorem. For the nonlinear case of Theorem 3.1, it is the subject of future research to find numerical criteria that guarantee the required assumptions. This is a challenging problem as there are many suboptimal approximations of an optimal filter and the verification of the corresponding hypotheses may prove more difficult.

Remark 3.2: Stability properties of suboptimal filters—e.g., the extended Kalman–Bucy filters (EKFs),^{8,90} as opposed to the truly optimal filter studied herein—have also been studied under the assumption of uniformly stable and fully observable signals. The stability constraint for EKFs has been removed in Refs. 77, 91, 92 (see also Ref. 93 for a study of the stability of the mean-squared filtering error).

IV. CONCLUSIONS AND FURTHER WORK

A. Summary

The main results of this work are given by Theorems 1 and 2 in Sec. III. Essentially,

- for nonlinear dynamics or observations—including unstable dynamics of the prior process and subject to certain technical but plausible assumptions—the supremum of the expectation of the Wasserstein distance W_2 between the true posterior solution and a solution of the FA process with the wrong initial conditions remains bounded at all future times; and

- (ii) for linear dynamics and observations—including unstable dynamics of the prior process and subject to certain technical but plausible assumptions—the Wasserstein distance W_2 between the two posterior distributions tends to zero.

B. Discussion

These results, to the best of our knowledge, are the first to address the stability of the posterior FA process given an unstable prior process. In the linear case, convergence in W_2 of the posterior processes starting from correct and incorrect initial data, p_0 and μ , has been demonstrated (Theorem 3.2) and the applicability to partial observations has been illustrated in Appendix B 4. In particular, observing the unstable components of the prior process, as originally proposed by Trevisan and her collaborators,^{14,15,34,94,95} seems to be an excellent idea. The results in the nonlinear case only guarantee W_2 -boundedness of the difference between the two posterior processes starting from distinct initial probability measures. A considerable amount of practical DA work also indicates that the FA process can track the correct solution,^{2,8,23,56} in particular, when using observations from the unstable subspace.⁶

The first results herein toward a more realistic mathematical treatment of the unstable-dynamics case open the door to a whole slew of additional results, both theoretical and practical.

a. Deterministic and stochastic EnKF. Operational DA in NWP relies these days more and more on the EnKF.^{6,41,42} However, in practice, most operational EnKF algorithms randomize only over the NWP model's initial states and not over observations, too, as done herein.

The operational NWP literature on DA distinguishes, in fact, between the *deterministic* EnKF, which only takes into account random errors in the observations via the covariance matrix \mathbf{R} of observational errors, and the *stochastic* EnKF, which explicitly simulates random errors in the observations. An excellent review of the EnKF for atmospheric DA in general appears in Ref. 96, with particular attention to this issue in its Sec. 2b. Lawson and Hansen⁹⁷ give interesting examples of the two versions of EnKF being applied to relatively simple examples of atmospheric and oceanic flows in one and two spatial dimensions, and Hoteit *et al.*⁹⁸ discuss some of the problems that might arise in the stochastic EnKF by introducing these observational random errors into the FA process. van Leeuwen⁹⁹ has proposed recently a self-consistent way of applying the stochastic EnKF.

Given the novel convergence results obtained herein in the presence of a random observation process and some of the renaissance interest in the NWP literature, it might be worthwhile revisiting the usefulness of the stochastic EnKF. In particular, retaining random perturbations in the observations might obviate the need for artificial inflation of the ensemble's rapidly lost variance for the deterministic EnKF.

b. Multiple models and model error. In practice, in NWP and elsewhere, prediction can be served by more than one model. The models can differ by their spatial resolution, by the

physical processes taken into consideration and by the numerical discretization of the PDEs governing them. Multi-model DA is discussed in some detail by Bach and Ghil,¹⁰⁰ including the issue of model error growth in this situation. It would be of considerable interest to extend the rigorous results herein to such a broader setting.

Typically, given the fact that DA is more expensive than straight forecasting,^{2,6,8,20,23} it is natural to use lower-resolution models for the FA process than for the forecasting. In the setup of Sec. II B herein, doing so corresponds to distinct prior processes p_i and posterior processes π_i and appropriate consideration of such issues would be quite worthwhile.

c. Parameter estimation and the synchronization point of view. In Sec. I A, we have mentioned already the view of the FA process as the synchronization of the forecast model with the observed process.^{101,102} This point of view has been used as a unifying principle between DA and supermodeling, namely, the use of ensembles of models that do not only serve for *a posteriori* averaging of their results but learn from each other in the process of a simulation or prediction run.³¹ This learning is clearly related to the estimation of imperfectly known model parameters.^{29,30}

In this broader perspective, one could try to demonstrate, given suitable hypotheses and observations, (i) the convergence of a single model's parameter estimation process and (ii) the convergence of a supermodel to the observed process.

d. Practical examples. We presented in Appendix B 4, a simple linear model and two observation schemes to illustrate the fact that stabilization by the FA process does not require observing all of a model's degrees of freedom and that observing just the unstable ones suffices. In future work, we aim to apply these results to realistic models, for example, models that numerically approximate the PDEs of geophysical fluid dynamics.

e. Particle filters. The results presented in this paper are theoretical in nature. In practice, the FA process cannot be computed exactly: numerical approximations are required to estimate the posterior distribution of the signal, given the data. Among these numerical approximations, particle filters have the crucial property of being theoretically justified in the sense that the numerical error can be controlled by the computational effort. Moreover, they are asymptotically consistent, i.e., as the number of data points used increases, the sequence of estimates converges in probability.

A *particle filter* is a sequential Monte Carlo method in which the posterior distribution is approximated using a set of *particles*, yielding a random measures of the form $\sum_{\ell} w_{\ell}^{\delta} \delta(x_{\ell}^{\delta})$, where δ is

the Dirac delta function, $w_{\ell}^1, w_{\ell}^2, \dots$ are the *weights* of the particles and $x_{\ell}^1, x_{\ell}^2, \dots$ are their corresponding positions,²⁰ centered around the state vector x_t . The approximations evolve in time, by following the time evolution induced by the prior model and are corrected by the observations to keep them close to the evolution of the FA process.^{20,25}

Particle filters have been very successful in many applications, including engineering, economics, and finance (see, for instance, Ref. 103, and the references therein). In recent years, applications of particle filtering to DA problems for planetary flows have flourished. For in-depth reviews of the most recent efforts in this direction, see Refs. 56 and 104. Such applications require enhancements of the classical particle filters in order to enable them to tackle the so-called curse of dimensionality, e.g., by relying on optimal transport ideas,⁵⁶ tempering,^{105–108} localization,^{56,109} model reduction,^{106,110} jittering,¹⁰⁵ nudging,¹⁰⁷ and judicious proposal densities.¹¹¹ Some of these approaches have been tested in operational NWP weather prediction systems.¹⁰⁹ The suitability of particle filters for high-dimensional problems has been studied in Ref. 105 and tested in Refs. 106 and 107. For example, in Ref. 107, the method is used for the stochastic incompressible two-dimensional Euler model with forcing and damping, while in Ref. 106, it is tested for a two-dimensional quasi-geostrophic model.

The theoretical results herein pave the way for the stability analysis of particle filters under the same assumptions. More precisely, one can attempt to show that particle filters have numerical errors that can be controlled *uniformly* in time. Again, we will be guided in pursuing such results by existing ones in nonlinear filtering, stochastic analysis, and applied probability.^{59,60,112–115} The bound will be in expectation, as in Theorem 3.1 of this paper. Coupled with Remark 2.5 in Sec. 2 of Appendix B, such results will offer theoretical validation to applying particle filters for *long-run* DA problems.

DEDICATION

This paper is dedicated to the memory of Anna Trevisan and to her contributions to data assimilation.

ACKNOWLEDGMENTS

It is a pleasure to thank Eviatar Bach for several useful discussions, for providing Figs. 1 and 2 and a draft of Appendix A, as well as for comments on the near-final manuscript. Alberto Carrassi and Pierre del Moral also read the near-final manuscript and made constructive suggestions. M.G. acknowledges, however belatedly, the Nelder Fellowship of the Imperial College's Mathematics Department, <http://www.imperial.ac.uk/mathematics/research/opportunities/nelder-visiting-fellowships/fellows/professor-michael-ghil/> that put the two authors in closer contact in Spring 2014. Both authors are pleased to acknowledge the Institut Henri Poincaré's trimester on "The Mathematics of Climate and the Environment" in Fall 2019, <http://www.ihp.fr/en/CEB/T3-2019>, which supported the real start of the collaboration leading to this paper. It is particularly gratifying to thank Franco Flandoli for an extended conversation during this trimester that set the two authors on the right path for the use of the tools in Ref. 19. The comments of two reviewers—one from the DA community and the other from the mathematical one—have further improved the paper. D.C. was partially supported by EU project STUOD—DLV-856408. The present paper is TiPES contribution # 206; this project has received funding from the EU Horizon 2020 research and innovation program under Grant Agreement No. 820970, and it helps support the work of M.G. Work on this paper has also been supported by the EIT Climate-KIC (Grant No.

190733); EIT Climate-KIC is supported by the European Institute of Innovation & Technology (EIT), a body of the European Union.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Dan Crisan: Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Michael Ghil:** Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

APPENDIX A. FORECAST ERROR GROWTH IN NWP

In order to better understand the nature and role of forecast error growth in the FA process, we consider here three different models for error growth in NWP, namely, those of Leith,¹¹⁶ Lorenz,⁴⁹ and of Dalcher and Kalnay.⁵⁰

The forecast error model of Leith¹¹⁶ is

$$\dot{V} = \alpha V + S, \quad (\text{A1})$$

where V is the mean-square error, S is the systematic model error, and t is the lead time. The V in this appendix should not be confused with the Brownian motion process in Eq. (8) of Sec. II. The forecast error growth in the Leith model is given by

$$V(t) = \left(V_0 + \frac{S}{\alpha} \right) e^{\alpha t} - \frac{S}{\alpha}, \quad (\text{A2})$$

where $V_0 = V(0)$ is the initial error. Note that short-time forecast errors grow exponentially and that the systematic model error acts to increase the coefficient of this growth.

Leith's forecast error model can only apply for short-time error growth, since it does not saturate. Note that under certain statistical assumptions, the mean-square error saturation value of a single forecast will be $2C$, where C is the climatological variance. In a real NWP model with N variables, the scalar C will be equal to the mean trace of the climatological covariance matrix \mathbf{C} , where \mathbf{C} has dimension $N \times N$. For ensemble forecasts, the saturation value becomes $(1 + 1/m)C$, where m is the ensemble size, see Ref. 117.

Lorenz's model of forecast error growth⁴⁹ is

$$\dot{E} = aE(E_\infty - E), \quad (\text{A3})$$

where E is the root-mean-square error and E_∞ is its saturation value. To compare this model directly to models based on mean-square error, like Eq. (A1), we can change variables to $V = E^2$, and get that

$$\dot{V} = 2aV^{1/2}V(1 - (V/V_\infty)^{1/2}). \quad (\text{A4})$$

Lorenz's model includes a nonlinear saturation term but does not incorporate systematic model error S .

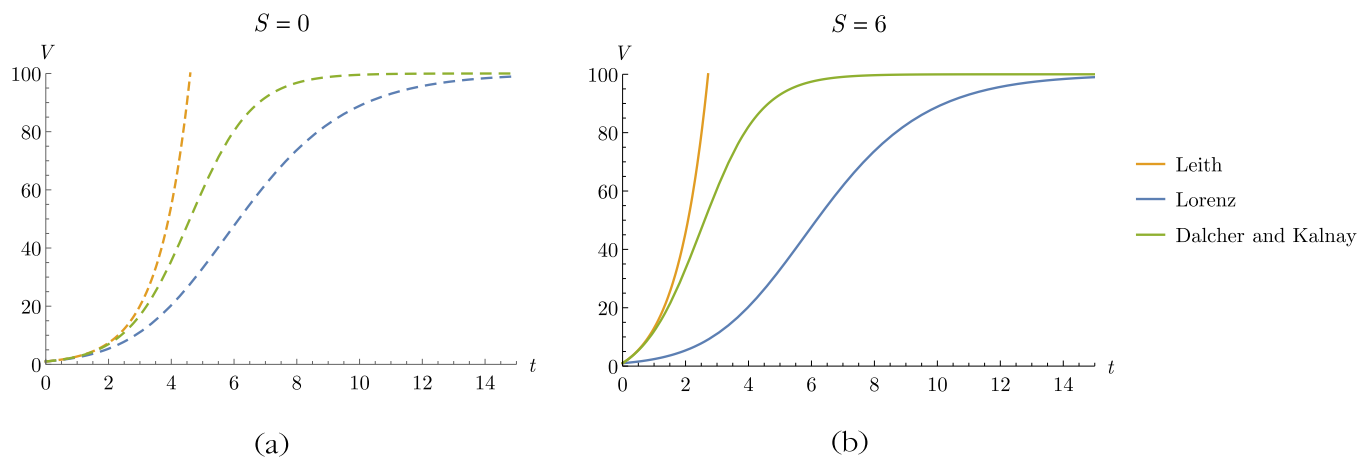


FIG. 3. Comparison between the model error growth $V(t)$ in the Leith,¹¹⁶ Lorenz,⁴⁹ and DK⁵⁰ models. The systematic model error is (a) $S = 0$, dashed curves; and (b) $S = 6$, solid curves. See legend for color identification.

The error model proposed by Dalcher and Kalnay⁵⁰ (henceforth DK) combines the key features of the Leith¹¹⁷ and Lorenz⁴⁹ models,

$$\dot{V} = (\alpha V + S)(1 - V/V_\infty). \quad (\text{A5})$$

It, thus, includes both saturation V_∞ and systematic model error S . For short-time error growth, we can take $V_\infty \rightarrow \infty$, recovering Leith's model. For $S = 0$, the model is similar to that of Lorenz, but with V_∞ having unit power in the saturation term, rather than $1/2$.

To compare the three models graphically, we set $\alpha = 1$, $V_0 = 1$, and $V_\infty = 100$. To match the Lorenz model's short-term error growth to that of the other two models, we set $a = \alpha / (2V_\infty^{1/2})$.

Figure 3 shows a comparison between the three error models, both with a perfect model for which $S = 0$ and with an imperfect model with $S = 6$. The Leith curve is the same in both cases, since it does not account for S . For the perfect model case, all three curves experience similar short-term exponential error growth. However, the Leith and DK models grow faster in the imperfect case than in the perfect case. In both cases, the Leith model diverges from the other two curves in the medium range due to its lack of saturation. The Lorenz and DK curves both saturate to V_∞ , although DK saturates more quickly due to the difference in the functional forms of the saturation terms in the two models.

DK actually used their model's three parameters, (α, S, V_∞) , to match their error growth curve to the archived real-time performance for the years 1980–1981 of the European Centre for Medium-range Weather Forecasts (ECMWF) model. Their Fig. 9 shows separate, near-perfect fits out to 10 days, for boreal winter, when the weather is more active, and boreal summer, when it is less so. Stroe and Royer¹¹⁸ subsequently generalized the DK model, by introducing the power V^p , with $p \neq 2$, in the saturation term, and took the limit $p \rightarrow \infty$ to obtain

$$\dot{V} = -aV \log(V/V_\infty). \quad (\text{A6})$$

These authors found that Eq. (A6) gave better fits for extended-range, 45-day experimental weather forecasts than either Eq. (A4) or (A5).

Simmons *et al.*¹¹⁹ also obtained rather good fits to operational NWP model performance with Lorenz's quadratic error growth model, but with a smaller error growth exponent than in the DK paper.⁵⁰ Trevisan *et al.*,¹²⁰ though, showed—by using an intermediate, quasigeostrophic two-layer model on the β -plane¹²¹—that only very small initial errors in such a model obey Lorenz's quadratic error growth model and that the error growth curve, in general, depends significantly on the magnitude of the initial errors.

Savijärvi¹²² combined features of the Lorenz and DK models in the study of the (then) U.S. National Meteorological Center's (NMC's) Medium-Range Forecast (MRF) Model's 0–10-day forecasts for 1988–1993. Growth parameters, as well as model and analysis errors for this data set, were estimated using the quadratic error growth assumption. Savijärvi showed that both the MRF model error and analysis error nearly halved during the six years under study but, at the same time, the growth parameters nearly doubled, since smaller errors grow faster.

Model error growth is, thus, a complex topic with much more to be said about; see, for instance, the line of inquiry developed by Nicolis and co-workers, which includes transient bimodality of the error's pdf.²⁶ The topic's quick review in this appendix suffices, though, to show the presence of error growth-generating instabilities in high-end, operational NWP models. The ground covered here in Secs. II and III and in Appendix B shows that DA can overcome these instabilities, in theory as well as in practice.

APPENDIX B. RIGOROUS DEFINITIONS AND PROOFS

1. The Wasserstein topology

In this appendix, we present the reason for showing the stabilizing effect of the FA process on unstable dynamics with respect to the Wasserstein topology—i.e., the topology generated by the

Wasserstein distance—and not with respect to the more popular weak topology. Before doing so, we provide a little more information on the Wasserstein distance for the benefit of NWP practitioners who might not yet be familiar with it.

Gaspard Monge, an artillery officer in Napoleon's armies, as well as one of the founders of France's Ecole Polytechnique, introduced it as early as the 1780s,⁵² and Leonid V. Kantorovich¹²³ used it during World War II in optimizing the transport of resources within the Soviet Union. The contemporary developments of this distance and of its applications are largely due to Dobrushin,¹²⁴ who coined the name Wasserstein¹²⁵ distance for it, and to Villani.¹²⁶

In the climate sciences, Ghil⁵³ illustrated the use of the Wasserstein distance for measuring the parameter sensitivity of simple models with time-dependent forcing, thus providing a link between nonautonomous dynamical systems theory¹⁸ and optimal transport.¹²⁶ Robin *et al.*¹²⁷ then used this distance to compute the difference between the snapshot attractors of the Lorenz¹²⁸ model for different time-dependent forcings, while Vissio *et al.*¹²⁹ used it to help intercomparing climate models and evaluating their performance against given benchmarks in the Coupled Model Intercomparison Project that is part of the Intergovernmental Panel on Climate Change process.

Returning to the main purpose of this appendix, let us introduce first some notation. For a measure $\mu \in \mathcal{P}^2(\mathbb{R}^d)$, we use the following notation:

(i) **means:** $\hat{\mu} = (\hat{\mu}^i)_{i \in 1, \dots, d}$ is the mean vector of μ , i.e.,

$$\hat{\mu}^i = \int_{\mathbb{R}^d} x_i \mu(dx), \quad i \in 1, \dots, d; \quad |\hat{\mu}| = \left(\sum_{i=1}^d (\hat{\mu}^i)^2 \right)^{1/2}; \quad (\text{B1})$$

(ii) **second moments:** μ^2 is the sum of the second moments of μ , i.e.,

$$\mu^2 = \sum_{i=1}^d \int_{\mathbb{R}^d} x_i^2 \mu(dx); \quad (\text{B2})$$

(iii) **covariance matrix:** $P_\mu = (P_\mu^{ij})_{i,j \in 1, \dots, d}$ is the covariance matrix of μ , i.e.,

$$P_\mu^{ij} = \int_{\mathbb{R}^d} (x_i - \hat{\mu}^i)(x_j - \hat{\mu}^j) \mu(dx), \quad i, j \in 1, \dots, d, \quad (\text{B3a})$$

$$|P_\mu| = \left(\sum_{i,j=1}^d (P_\mu^{ij})^2 \right)^{1/2}. \quad (\text{B3b})$$

Next, we recall the definition of the weak topology on the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ and note that, of course, $\mathcal{P}^2(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$. We can thus consider also weak convergence of probability measures belonging to the smaller space $\mathcal{P}^2(\mathbb{R}^d)$:

Definition B.1 (Weak topology): A sequence of probability measures $(\mu_n)_n \in \mathcal{P}(\mathbb{R}^d)$, converges *weakly* to $\mu \in \mathcal{P}(\mathbb{R}^d)$ if and only if $(\mu_n(\varphi))_n$ converges to $\mu(\varphi)$ as $n \rightarrow \infty$ for all $\varphi \in C_b(\mathbb{R}^d)$. The *weak topology* on the space $\mathcal{P}(\mathbb{R}^d)$ is defined to be the weakest topology such that for all $f \in C_b(\mathbb{R}^d)$, the function $\mu \mapsto \mu(f)$ is continuous.

The weak convergence of $(\mu_n)_n$ to μ is denoted $\mu_n \Rightarrow \mu$. A set of probability measures $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^d)$ is relatively compact in the weak topology if and only if for all $\varepsilon > 0$ there exists K_ε such that $\mu(K_\varepsilon) \geq 1 - \varepsilon$ for all $\mu \in \mathcal{A}$. If $\mathcal{A} \subset \mathcal{P}^2(\mathbb{R}^d)$, one can show that the set \mathcal{A} will be relatively compact in the weak topology if the means and the covariance matrices of the probability measures in the set \mathcal{A} are uniformly bounded.

The set \mathcal{A} can, however, still be relatively compact, even if the means of the probability measures in it do not remain bounded. For example, if we choose $\mu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_n$, then the sequence μ_n is relatively compact—in fact, $\mu_n \Rightarrow \delta_0$ —but the corresponding sequence of means $\hat{\mu}_n$ is not bounded, since $\mu_n(\varphi) = n$. On the other hand, if we choose $\mu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_n$, then the sequence μ_n is relatively compact, the means $\hat{\mu} \equiv 1$ form a trivially bounded sequence, $\mu_n(\varphi) = 1$, but the second moments are not, as $\mu_n^2 = n$. This state of affairs is not satisfactory for our purposes.

The Wasserstein topology, though, adds the convergence of the first and second moments to the weak convergence of the measures. To be precise, we have $\lim_{n \rightarrow \infty} W_2(\mu_n, \mu) = 0$ for $\mu_n, \mu \in \mathcal{P}^2(\mathbb{R}^d)$, if and only if μ_n converges to μ in the weak topology and the first and second moments converge as well, using the notation of Eqs. (B1)–(B3) above. Moreover, a set of probability measures $\mathcal{A} \in \mathcal{P}^2(\mathbb{R}^d)$ is relatively compact in the topology given by the Wasserstein distance if and only if

$$\lim_{R \rightarrow \infty} \sup_{\mu \in \mathcal{A}} \int_{|x| > R} |x|^2 \mu(dx) = 0. \quad (\text{B4})$$

Finally, we have the following lemma which follows immediately from the alternative definition (15) of the Wasserstein distance W_2 :

Lemma 2.1: There exists a constant $C = C(d)$ such that, for any $\mu, \nu \in \mathcal{P}^2(\mathbb{R}^d)$,

$$(W_2(\mu, \nu))^2 \leq C(\mu^2 + \nu^2),$$

$$W_2(\mu, \nu) \leq C(|P_\mu|^{1/2} + |P_\nu|^{1/2} + |\hat{\mu} - \hat{\nu}|).$$

2. Assumptions and proof of Theorem 3.1

To start, we formulate here the set of assumptions on the coefficients of the signal and observation Eqs. (8)+(10) under which Theorem 3.1 holds:

- We assume that the coefficients f and h can be decomposed into a linear part and a *bounded* nonlinear part. In other words, we will assume that

$$f = F\mathcal{J} + \tilde{f}, \quad h = H\mathcal{J} + \tilde{h}, \quad (\text{B5})$$

where

- (i) $\mathcal{J}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the identity function defined as $\mathcal{J}(x) = x$ for any $x \in \mathbb{R}^d$;
- (ii) $F \in \mathbb{R}^{d \times d}$, $H \in \mathbb{R}^{d \times n}$ are given matrices; and
- (iii) $\tilde{f}: \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\tilde{h}: \mathbb{R}^d \rightarrow \mathbb{R}^n$ are bounded measurable functions that incorporate the nonlinear parts of the coefficients of the system (8) and (10).

- The covariance matrices of the processes π^μ and π , respectively, are uniformly bounded in expectation:

$$\sup_{t \geq 0} \mathbb{E}[|P_t^\mu|^8] = C^\mu < \infty, \quad \sup_{t \geq 0} \mathbb{E}[|P_t^{\pi_0}|^8] = C^{\pi_0} < \infty. \quad (\text{B6})$$

- The matrix-valued process $Q_s^\mu := F - P_s^\mu H^\top H - \lambda_s^\mu H$ is exponentially stable in expectation, where λ_s^μ is the matrix-valued process defined as $\lambda_s^\mu := \pi_s^\mu (\mathcal{J} - \hat{\pi}^\mu) \tilde{h}^\top$ for $s \geq 0$. In other words, if $\psi_{s,t}$ is the solution of the linear matrix ordinary differential equation

$$d_t \psi_{s,t} = Q_t^\mu \psi_{s,t}, \quad \psi_{s,s} = I,$$

where I is the identity matrix, then there exists some constant $c > 0$ such that

$$\mathbb{E}[|\psi_{s,t}|^2] \leq e^{-c(t-s)}. \quad (\text{B7})$$

Moreover, we assume that there exists $\delta > 0$ such that

$$\int_{k\delta}^{(k+1)\delta} \mathbb{E}[|\psi_{k\delta,s}^{-1}|^4] ds \leq C^{inv}, \quad (\text{B8})$$

where C^{inv} is a constant independent of k .

The proof of Theorem 3.1 requires the bound of the difference $\hat{\pi}_t^\mu - \hat{\pi}_t$ between the mean of the FA process initialized from μ and the original FA process, initialized from π_0 . We deduce from (11) that, for $t \in [k\delta, (k+1)\delta]$,

$$\begin{aligned} \hat{\pi}_{tk\delta}^\mu &= \hat{\pi}_{k\delta}^\mu + \int_{k\delta}^t \pi_s^\mu(f) ds + \int_{k\delta}^t \mathcal{K}_s^\mu(dY_s - \pi_s^\mu(h) ds), \\ \hat{\pi}_t^\mu &= \hat{\pi}_{k\delta}^\mu + \int_{k\delta}^t \pi_s^\mu(f) ds + \int_{k\delta}^t \mathcal{K}_s^\mu(dI_s + (\pi_s(h) - \pi_s^\mu(h)) ds), \\ \hat{\pi}_t &= \hat{\pi}_{k\delta} + \int_{k\delta}^t \pi_s(f) ds + \int_{k\delta}^t \mathcal{K}_s dI_s. \end{aligned}$$

Here,

$$\begin{aligned} \mathcal{K}_s^\mu &:= \pi_s^\mu(\mathcal{J}h^\top) - \pi_s^\mu(\mathcal{J})\pi_s^\mu(h^\top) = P_s^\mu H^\top + \lambda_s^\mu, \\ \mathcal{K}_s &:= \pi_s(\mathcal{J}h^\top) - \pi_s(\mathcal{J})\pi_s(h^\top) = P_s^{\pi_0} H^\top + \lambda_s. \end{aligned}$$

P_s^μ and $P_s^{\pi_0}$ are the covariance matrices of π_s^μ and of π , respectively, and $\lambda_s^\mu := \pi_s^\mu((\mathcal{J} - \hat{\pi}^\mu)\tilde{h}^\top)$ and $\lambda_s := \pi_s((\mathcal{J} - \hat{\pi})\tilde{h}^\top)$, respectively. It follows that

$$\begin{aligned} \hat{\pi}_t^\mu - \hat{\pi}_t &= (\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}) + \int_{k\delta}^t (\pi_s^\mu - \pi_s)(f - \mathcal{K}_s^\mu h) ds \\ &\quad + \int_{k\delta}^t (\mathcal{K}_s^\mu - \mathcal{K}_s) dI_s \\ &= (\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}) + \int_{k\delta}^t Q_s^\mu (\hat{\pi}_s^\mu - \hat{\pi}_s) ds \\ &\quad + \int_{k\delta}^t (P_s^\mu - P_s) H^\top dI_s + \mathcal{K}_{k\delta:t}^\mu, \end{aligned} \quad (\text{B9})$$

where $\mathcal{K}_{k\delta:t}^\mu$ is a process that contains the nonlinearities in the evolutions $\hat{\pi}^\mu$ and $\hat{\pi}$,

$$\begin{aligned} \mathcal{K}_{k\delta:t}^\mu &= \int_{k\delta}^t (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{K}_s^\mu \tilde{h}) ds + \int_{k\delta}^t (\lambda_s^\mu - \lambda_s) dI_s, \\ t &\in [k\delta, (k+1)\delta]. \end{aligned}$$

Replacing the observation process Y_t by the innovation process I_t in the evolution equations for $\hat{\pi}_t^\mu$ and $\hat{\pi}_t$, respectively, is important: Unlike Y_t , I_t is a standard Brownian motion, which enables us to use classical stochastic calculus properties to bound the moments of the stochastic integrals appearing in Eq. (B9).

Using an argument based on the Grönwall inequality, one deduces that there exists a constant c_δ independent of k such that

$$\sup_{t \in [k\delta, (k+1)\delta]} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] \leq c_\delta \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}|]. \quad (\text{B10})$$

We will show that there exists R_δ such that

$$\sup_{k \geq 0} \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}|] \leq R_\delta. \quad (\text{B11})$$

From (B10) and (B11), one can then deduce that

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] &\leq \sup_{k \geq 0} \sup_{t \in [k\delta, (k+1)\delta]} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] \\ &\leq c_\delta \sup_{k \geq 0} \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}|] \leq c_\delta R_\delta < \infty. \end{aligned}$$

Finally, from Lemma 2.1 and (B6), it follows that

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E}[W_2(\pi_t^\mu, \pi_t)] &\leq \sup_{t \geq 0} \mathbb{E}[\sqrt{|P_t^\mu|}] + \sup_{t \geq 0} \mathbb{E}[\sqrt{|P_t^{\pi_0}|}] \\ &\quad + \sup_{t \geq 0} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] \\ &\leq \left(\sup_{t \geq 0} \mathbb{E}[|P_t^\mu|^8] \right)^{\frac{1}{16}} + \left(\sup_{t \geq 0} \mathbb{E}[|P_t^{\pi_0}|^8] \right)^{\frac{1}{16}} \\ &\quad + \sup_{t \geq 0} \mathbb{E}[|\hat{\pi}_t^\mu - \hat{\pi}_t|] \\ &\leq (C^\mu)^{\frac{1}{16}} + (C^{\pi_0})^{\frac{1}{16}} + c_\delta R_\delta < \infty, \end{aligned}$$

which yields our claim.

To complete the proof, the validity of (B11) remains to be shown. From (B9), one obtains a form of Duhamel's principle for the difference between the values of the mean $\hat{\pi}$ of the FA process at the steps $(k+1)\delta$ and $k\delta$,

$$\begin{aligned} \hat{\pi}_{(k+1)\delta}^\mu - \hat{\pi}_{k\delta}^\mu &= \psi_{k\delta, (k+1)\delta} (\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}) \\ &\quad + \psi_{k\delta, (k+1)\delta} \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (P_s^\mu - P_s) H^\top dI_s \\ &\quad + \psi_{k\delta, (k+1)\delta} \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} d\mathcal{K}_s^\mu. \end{aligned} \quad (\text{B12})$$

We analyze next each of the three terms on the right-hand side of (B12). For the first term, we use (B7) to derive the inequality

$$\mathbb{E}[|\psi_{k\delta, (k+1)\delta} (\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta})|] \leq e^{-c_\delta} \mathbb{E}[|\hat{\pi}_{k\delta}^\mu - \hat{\pi}_{k\delta}|]. \quad (\text{B13})$$

For the second term, we use the so-called Itô's integral isometry property, cf. Karatzas and Shreve,¹⁶ to obtain

$$\begin{aligned} & \mathbb{E} \left[\left| \psi_{k\delta, (k+1)\delta} \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (P_s^\mu - P_s) H^\top dI_s \right|^2 \right] \\ & \leq \mathbb{E} \left[\left| \psi_{k\delta, (k+1)\delta} \right|^2 \right] \mathbb{E} \left[\left| \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (P_s^\mu - P_s) H^\top dI_s \right|^2 \right] \\ & \leq e^{-2c\delta} \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |\psi_{k\delta, s}^{-1} (P_s^\mu - P_s) H^\top|^2 ds \right] \\ & \leq e^{-2c\delta} \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |\psi_{k\delta, s}^{-1}|^4 ds \right] \\ & \quad + e^{-2c\delta} \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |(P_s^\mu - P_s) H^\top|^4 ds \right] \leq C^l e^{-2c\delta}, \quad (\text{B14}) \end{aligned}$$

where $C^l = C^{inv} + \delta |H|^4 \sqrt{C^\mu} + \delta |H|^4 \sqrt{C^{\pi 0}}$.

Next, (B6) implies that there exists positive constants $C^{\pi\mu}$, C^π , $C^{\lambda\mu}$, and C^λ such that

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E}[|\mathcal{X}_t^\mu|^8] &= C^{\pi\mu} < \infty, \quad \sup_{t \geq 0} \mathbb{E}[|\mathcal{X}_t|^\pi] = C^\pi < \infty, \\ \sup_{t \geq 0} \mathbb{E}[|\lambda_t^\mu|^8] &= C^{\lambda\mu} < \infty, \quad \sup_{t \geq 0} \mathbb{E}[|\lambda_t|^\lambda] = C^\lambda < \infty. \quad (\text{B15}) \end{aligned}$$

For the third term, one gets

$$\begin{aligned} & \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} dz_s^\mu \\ &= \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) ds \\ & \quad + \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) (\lambda_s^\mu - \lambda_s) dI_s, \quad (\text{B16}) \end{aligned}$$

and one proceeds to bound separately the two terms in (B16). For the first term in (B16), we get

$$\begin{aligned} & \mathbb{E} \left[\left| \psi_{k\delta, (k+1)\delta} \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) ds \right|^2 \right] \\ & \leq e^{-c\delta} \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |\psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h})|^2 ds \right] \\ & \leq e^{-c\delta} \left(\mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |\psi_{k\delta, s}^{-1}|^2 ds \right] \right. \\ & \quad \left. + \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} |(\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h})|^2 ds \right] \right) \\ & \leq e^{-c\delta} \left(C^{inv} + \delta |\tilde{f}|^2 + |\tilde{h}|^2 (C^{\pi\mu})^{\frac{1}{4}} \delta \right). \quad (\text{B17}) \end{aligned}$$

For the second term in (B16), we get that

$$\begin{aligned} & \mathbb{E} \left[\left| \psi_{k\delta, (k+1)\delta} \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) (\lambda_s^\mu - \lambda_s) dI_s \right|^2 \right] \\ & \leq e^{-c\delta} \mathbb{E} \left[\left| \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) (\lambda_s^\mu - \lambda_s) dI_s \right|^2 \right] \\ & \leq e^{-c\delta} \int_{k\delta}^{(k+1)\delta} \mathbb{E} \left[|\psi_{k\delta, s}^{-1} (\pi_s^\mu - \pi_s) (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) (\lambda_s^\mu - \lambda_s)|^2 \right] ds \\ & \leq e^{-c\delta} \int_{k\delta}^{(k+1)\delta} \mathbb{E} [|\psi_{k\delta, s}^{-1}|^4] ds \\ & \quad + e^{-c\delta} \int_{k\delta}^{(k+1)\delta} \mathbb{E} [|\pi_s^\mu - \pi_s| (\tilde{f} - \mathcal{X}_s^\mu \tilde{h}) (\lambda_s^\mu - \lambda_s)|^4] ds \\ & \leq e^{-c\delta} \left(C^{inv} + \delta |\tilde{f}|^4 (\sqrt{C^{\lambda\mu}} + \sqrt{C^\lambda}) + \delta |\tilde{h}|^4 (C^{\pi\mu} + C^{\pi\epsilon}) \right). \quad (\text{B18}) \end{aligned}$$

From (B16)–(B18), it follows that there exists a constant

$$C^{\pi\mu} = C^{\pi\mu} (\delta, C^{inv}, \tilde{f}, C^{\lambda\mu}, C^\lambda, \tilde{h}, C^{\pi\mu}, C^{\pi\epsilon}),$$

independent of k such that

$$\mathbb{E} \left[\left| \int_{k\delta}^{(k+1)\delta} \psi_{k\delta, s}^{-1} dz_s^\mu \right|^2 \right] \leq C^{\pi\mu} e^{-c\delta}. \quad (\text{B19})$$

Finally, from (B12), (B13), (B14), and (B19), we deduce that

$$\mathbb{E} \left[\left| \hat{\pi}_{(k+1)\delta}^\mu - \hat{x}_{(k+1)\delta} \right|^2 \right] \leq e^{-c\delta} \left(\mathbb{E} [|\hat{\pi}_{k\delta}^\mu - \hat{x}_{k\delta}|] + C^l + C^{\pi\mu} \right). \quad (\text{B20})$$

Choose now $R = R(\delta) = \max \left(|\hat{\mu} - \hat{\pi}_0|, \frac{C^l + C^{\pi\mu}}{c\delta} \right)$ and use induction to prove (B11). From the definition of R , we deduce that $|\hat{\mu} - \hat{\pi}_0| \leq R$. Next, assume that $\mathbb{E} [|\hat{\pi}_{k\delta}^\mu - \hat{x}_{k\delta}|] \leq R$. From this and (B20), one can obtain that

$$\begin{aligned} \mathbb{E} [|\hat{\pi}_{(k+1)\delta}^\mu - \hat{x}_{(k+1)\delta}|] & \leq e^{-c\delta} (R + C^l + C^{\pi\mu}) \\ & \leq e^{-c\delta} R \left(1 + \frac{C^l + C^{\pi\mu}}{R} \right) \\ & \leq e^{-c\delta} R (1 + c\delta) \leq R. \end{aligned}$$

It follows that $\mathbb{E} [|\hat{\pi}_{k\delta}^\mu - \hat{x}_{k\delta}|] \leq R$ holds true for any $k \geq 0$ and so does (B11). The proof of Theorem 3.1 is now complete. \square

Remark 2.3: The proof of Theorem 3.1 relies on the application of Duhamel's principle, as in (B12). The Duhamel principle can be applied on an arbitrary interval $[0, t]$ to deduce that

$$\begin{aligned} \hat{\pi}_t^\mu - \hat{\pi}_{k\delta}^\mu &= \psi_{0,t} (\hat{\pi}_0^\mu - \hat{x}_0) + \psi_{0,t} \int_0^t \psi_{0,s}^{-1} (P_s^\mu - P_s) H^\top dI_s \\ & \quad + \psi_{0,t} \int_0^t \psi_{0,s}^{-1} dz_s^\mu. \quad (\text{B21}) \end{aligned}$$

For a deterministic ψ , we can rewrite (B21) as

$$\hat{\pi}_t^\mu - \hat{\pi}_{k\delta}^\mu = \psi_{0,t}(\hat{\pi}_0^\mu - \hat{x}_0) + \int_0^t \psi_{s,t}(P_s^\mu - P_s)H^\top dI_s + \int_0^t \psi_{s,t}dz_s^\mu. \quad (\text{B22})$$

In this case, the bound of the stochastic terms in (B22) would follow directly from (B7) or, more precisely, from the deterministic exponential decay of $\psi_{s,t}$. However, in our case, ψ is *not* deterministic. Moreover, the process $\psi_{s,t}$ is not adapted with respect to the filtration generated by the Brownian motion I . To be more precise, it does not depend on $\{I_r, r \in [0, s]\}$ but on $\{I_r, r \in [0, t]\}$. Hence, the stochastic integrals in (B22) do not make sense as standard Itô integrals. One can interpret them using a more general definition of Skorohod integration, but the control of such resulting integrals is no longer immediately obvious. A bound using Malliavin calculus may be possible (see the monograph of Nualart¹³⁰ for details of the methodology).

In another approach, one could attempt to keep out of the stochastic integrals the part that is not adapted, i.e., $\psi_{0,t}$, and only use (B21) but not (B22). To do so, one would need to bound the exponential blowup of the stochastic integrals in (B21) as $t \rightarrow \infty$ over the entire positive half line $[0, \infty)$. Instead, we limit ourselves to apply Duhamel's principle on intervals of the form $[k\delta, (k+1)\delta]$, where δ is small enough to be able to rely on the bound given by (B8).

Finally, note that, using again an argument based on the Grönwall inequality, one obtains that, for any $T > 0$, there exists a constant c_T such that the expected difference between the means of the FA process started from different initial distributions μ and π_0 satisfies

$$\sup_{t \in [0, T]} \mathbb{E} [|\hat{\pi}_t^\mu - \hat{\pi}_t|] \leq c_T. \quad (\text{B23})$$

Thus, we only need to check that there exists a time horizon T for which

$$\sup_{t \in [T, \infty)} \mathbb{E} [|\hat{\pi}_t^\mu - \hat{\pi}_t|] \leq c_T. \quad (\text{B24})$$

The bound of the eighth moments of the covariance matrices P_t^μ and $P_t^{\pi_0}$, respectively, in (B6), as well as of the fourth moments of $\psi_{k\delta:s}^{-1}$, suffices for the result above, but it is perhaps not necessary. We leave it for future research to find the optimal bounds.

Remark 2.4: In the linear and Gaussian case, \tilde{f} and \tilde{h} are constants and $\lambda_s^\mu = 0$. Therefore,

$$z^\mu = \int_0^t (\tilde{f} - P_s^\mu H^\top \tilde{h}) ds;$$

in other words, z^μ is a process with bounded variation. Moreover, P_s^μ is deterministic. In this case, all the technical difficulties described above vanish and Theorem 3.1 holds true with a much simpler proof. In this case, though, a much stronger result holds, namely, Theorem 3.2, which is proved in Sec. 3 of this appendix.

Remark 2.5: Theorem 3.1 gives us only an upper bound on the expected value of the distance between π_t^μ and π_t . From it, we can derive the following alternative control: Let us denote by $T(t, R)$ the average time spent by the process π_t^μ outside the ball $B(\pi_t, R)$,

that is

$$T(t, R) = \frac{1}{t} \int_0^t I_{\{W_2(\pi_t^\mu, \pi_t^0) > R\}} ds.$$

Then, for any $\varepsilon > 0$, there exists a constant $R = R(\mu, \varepsilon)$ independent of t such that

$$\sup_{t \geq 0} \mathbb{E} [T(t, R)] < \varepsilon.$$

We note that we cannot expect $T(t, R)$ to decrease to zero as t tends to ∞ , since $\hat{\pi}_t^\mu - \hat{x}_t$ can be viewed as a d -dimensional Ornstein–Uhlenbeck process with random coefficients and perturbed by the random residue process z .

Remark 2.6: The concept of stabilization described in Theorem 3.1 is much weaker than the one described in many classical stabilization results, e.g., such as those covered by Budhiraja,⁵⁷ Picard,⁶⁴ or Van Handel.⁶⁵ On the other hand, it has the advantage that the signal $X(t)$ is not required to be fully observable. In particular, the dimension n of the observation $Y(t)$ and the dimension d of the signal $X(t)$, in the notation of Sec. II, do not need to coincide. Heuristically, we need to be able to observe all the “unstable” directions. See Sec. 4 in this Appendix for some simple illustrative examples in this direction. This feature of our results is very advantageous in practical applications as, in many situations, the dimension of the observation space is considerably smaller than that of the state space, $n \ll d$.^{2,23}

Remark 2.7: One can deduce a criterion for the bound (B6) to hold. The derivation of such a criterion can be carried out in terms of the centralized third moments of π_t^μ , along the lines of the arguments in Sec. 6.2 of Bain and Crisan,²⁰ and it is the subject of future research.

3. Assumptions and proof of Theorem 3.2

The assumptions and the arguments in this section are based on the framework of Ocone and Pardoux⁶³ and, in part, on their results. In particular, we impose the following set of assumptions on the coefficients of the signal and observation Eqs. (8) and (10) under which Theorem 3.2 holds:

- We assume that the coefficients f and h are linear. In other words, we will assume that

$$f = F\mathcal{J} + \tilde{f}, \quad h = H\mathcal{J} + \tilde{h}, \quad (\text{B25})$$

where

- $\mathcal{J}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the identity function defined as $\mathcal{J}(x) = x$ for any $x \in \mathbb{R}^d$;
 - $F \in \mathbb{R}^{d \times d}$, $H \in \mathbb{R}^{d \times n}$ are given matrices;
 - \tilde{f} and \tilde{h} are d -dimensional and n -dimensional vectors, respectively; and
 - the function σ is a constant $d \times d$ -matrix.
- We assume that the measure μ has finite second moments—and, thus, it belongs to the Wasserstein space, as discussed in Appendix A—and that it is absolutely continuous with respect to π_0 . We denote by θ_μ the density of μ with respect to π_0 , which is integrable with respect to π_0 . Using a standard probabilistic result (see, for example, Problem 3.20 in Karatzas and Shreve¹⁶), it follows that there exists a random variable $\Upsilon_\mu \geq 0$ such that

$\lim_{t \rightarrow \infty} \mathbb{E} [\theta_\mu(X_0) | \mathcal{Y}_t] = \Upsilon_\mu$. Moreover, both $\mathbb{E} [\theta_\mu(X_0) | \mathcal{Y}_t]$ and Υ_μ have a pathwise representation in terms of the observation path Y_t . We assume that, for the path for which we do the analysis, $\lim_{t \rightarrow \infty} \mathbb{E} [\theta_\mu(X_0) | \mathcal{Y}_t] = \Upsilon_\mu > 0$. Similarly, we will be assuming that both $\lim_{t \rightarrow \infty} \mathbb{E} [\theta_\mu(X_0) X_0 | \mathcal{Y}_t]$ and $\lim_{t \rightarrow \infty} \mathbb{E} [\theta_\mu(X_0) X_0 X_0^* | \mathcal{Y}_t]$ exist, based on the same standard probabilistic result.

- We define $\pi_t^0 := N(\hat{x}_t^0, P_t^0)$, where \hat{x} satisfies the stochastic differential equation

$$\begin{aligned} d\hat{x}_t^0 &= (F\hat{x}_t^0 + \tilde{f}) dt + P_t H^\top (dY_t - (H\hat{x}_t^0 + \tilde{h}) dt) \\ &= (Q_t \hat{x}_t^0 + \tilde{f} - P_t H^\top \tilde{h}) dt + P_t H_t^\top dY_t, \quad \hat{x}_0 = 0, \end{aligned} \quad (\text{B26})$$

and P^0 satisfies the deterministic matrix Riccati equation

$$\begin{aligned} \frac{dP_t^0}{dt} &= \sigma_t \sigma_t^\top + F P_t^0 + P_t^0 F^\top - P_t^0 H^\top H P_t^0 \\ &= F Q_t + Q_t F^\top + \sigma_t \sigma_t^\top + P_t H^\top H P_t, \end{aligned} \quad (\text{B27})$$

with $P_0 = 0 \in \mathbb{R}^{d \times d}$ being the matrix with null entries. We assume, furthermore, that there exists a unique solution $P_\infty \geq 0$ to the algebraic Riccati equation

$$\sigma \sigma^\top + F P + P F^\top - P H^\top H P = 0, \quad (\text{B28})$$

and that $Q_t := F - P_t^0 H^\top H$ is asymptotically stable. Using this unique non-negative solution, we define $Q_\infty := F - P_\infty H^\top H$.⁶⁵

Following from Remark 2.1 and Lemma 2.2 in Ocone and Pardoux,⁶³ which, in turn, uses the classical result of Theorem 3.7 in Kwakernaak and Sivan,¹³¹ we make the following remark:

Remark 2.8: If (F, H) is detectable and (F, σ_0) is stabilizable, then (B28) has a unique non-negative definite solution P_∞ and one has both $P_t \rightarrow P_\infty$ and $Q_t \rightarrow Q_\infty$ for any initial condition $P_0 \geq 0$, with Q_t asymptotically stable. See Chap. 1 in Kwakernaak and Sivan¹³¹ for the definition of detectability and stabilizability. In particular, the eigenvalues of Q_∞ have negative real parts and for any $0 < a < \inf \{-\operatorname{Re} \lambda; \text{ over } \lambda \text{ being an eigenvalue of } Q_\infty\}$, there is a constant K_a such that

$$\|P_t - P_\infty\| \leq K_a e^{-at}.$$

The last fact can be proved by observing that

$$\begin{aligned} \frac{d}{dt} (P_t - P_\infty) &= [F - 1/2 (P_t + P_\infty) H^\top H] (P_t - P_\infty) \\ &\quad + (P_t - P_\infty) [F - 1/2 (P_t^R + P_\infty) H^\top H]^\top \end{aligned}$$

and carrying out an analysis similar to that in Theorem 2.3 of Ref. 63.

We will show below that π^μ gets asymptotically close to the “reference process” π^0 , regardless of the initial condition, and since this holds true for $\mu = \pi_0$ too, we immediately deduce that

$$\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t) = 0.$$

Note that π coincides with π^0 if the initial condition π_0 is the Dirac delta distribution at 0, that is, $\pi_0 = \delta_0$. The reference process π^0 is convenient to work with: From Remark 2.8 we deduce that the centered version of $\pi^{0, -\hat{x}_t^0} := N(0, P_t^0)$ converges weakly, as well as

in Wasserstein distance, to $\pi^\infty = N(0, P_\infty)$. Using the equivalent definition of the Wasserstein distance (15), it follows that

$$\begin{aligned} \lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t^0) = 0 &\iff \lim_{t \rightarrow \infty} W_2\left(\pi_t^{\mu, -\hat{x}_t^0}, \pi_t^{0, -\hat{x}_t^0}\right) = 0 \\ &\iff \lim_{t \rightarrow \infty} W_2\left(\pi_t^{\mu, -\hat{x}_t^0}, \pi_t^\infty\right) = 0. \end{aligned}$$

To prove the last limit, it suffices to show that $\pi_t^{\mu, -\hat{x}_t^0}$ converges to π^∞ in the weak topology and the first and second moments of $\pi_t^{\mu, -\hat{x}_t^0}$ converge to the first and second moments of π^∞ , respectively. Equivalently, it suffices to show that:

- $\pi_t^{\mu, -\hat{x}_t^0}$ converges to π^∞ in the weak topology;
- $\lim_{t \rightarrow \infty} |\hat{\pi}_t^\mu - \hat{x}_t^0| = 0$, i.e., the distance between the mean of π_t^μ and that of π_t^0 tends to 0; and
- $\lim_{t \rightarrow \infty} |P_{\pi_t^\mu} - P_t^0| = 0$, i.e., the distance between the covariance matrix of π_t^μ and that of π_t^0 tends to 0.

Since the set of bounded uniformly continuous functions are convergence determining, to justify that $\pi_t^{\mu, -\hat{x}_t^0}$ converges to π^∞ in the weak topology, it suffices to show that $\lim_{t \rightarrow \infty} |\pi_t^{\mu, -\hat{x}_t^0}(\varphi) - \pi^\infty(\varphi)| = 0$ for any φ that is a bounded and uniformly continuous function. To recap, we have that $\lim_{t \rightarrow \infty} W_2(\pi_t^\mu, \pi_t^0) = 0$, if and only if the following three properties hold true:

- $\lim_{t \rightarrow \infty} |\hat{\pi}_t^\mu - \hat{x}_t^0| = 0$;
- $\lim_{t \rightarrow \infty} |P_{\pi_t^\mu} - P_t^0| = 0$; and
- $\lim_{t \rightarrow \infty} |\pi_t^\mu(\varphi_t) - \pi_t^0(\varphi_t)| = 0$ for any bounded uniformly continuous function φ , where φ_t is the same function shifted by the mean \hat{x}_t^0 , that is $\varphi_t(x) := \varphi(x + \hat{x}_t^0)$, $x \in \mathbb{R}^d$.

Note that the reference process π^0 is not unique: we can replace it by any other process $\tilde{\pi}^0$ with the property that $\lim_{t \rightarrow \infty} W_2(\tilde{\pi}_t^0, \pi_t^0) = 0$.

The proof of Theorem 3.2 is largely based on the arguments in the proof of Theorem 2.6 in Ref. 63. Using the linearity of Eq. (8), we can decompose the signal as follows:

$$\begin{aligned} X_t &= e^{Ft} X_0 + \tilde{X}_t, \\ \tilde{X}_t &:= \int_0^t (F\tilde{X}_s + \tilde{f}) ds + \int_0^t \sigma dV_s = \int_0^t e^{F(t-s)} (\tilde{f} ds + \sigma dV_s), \\ Y_t &= \int_0^t (He^{Fs} X_0 + H\tilde{X}_s + \tilde{h}) ds + W_t, \end{aligned}$$

and we introduce the new measure $\bar{\mathbb{P}}$ defined by

$$\left. \frac{d\bar{\mathbb{P}}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \exp \left[\int_0^t -\langle He^{Fs} X_0, dW_s \rangle - \frac{1}{2} \int_0^t |He^{Fs} X_0|^2 ds \right], \quad t \geq 0.$$

By Girsanov's theorem, under this measure $\bar{\mathbb{P}}$, the process

$$\bar{W}_t := \int_0^t He^{Bs} X_0 ds + W_t, \quad t \leq T$$

is a Brownian motion, and X_0 is independent of $(V_t, \bar{W}_t)_{t \leq T}$. Following the proof of Proposition 3.13 in Ref. 20, the law of \bar{X}_0 remains

unchanged under $\bar{\mathbb{P}}$. Let

$$\begin{aligned} L_t &:= \frac{d\mathbb{P}}{d\bar{\mathbb{P}}} \Big|_{\mathcal{F}_t} = \exp \left[\int_0^t \langle H e^{B_s} X_0, d\bar{W}_s \rangle - \frac{1}{2} \int_0^t |H e^{B_s} X_0|^2 ds \right] \\ &= \exp \left(\langle X_0, W^\circ \rangle - \frac{1}{2} \langle M_t X_0, X_0 \rangle \right), \end{aligned}$$

where

$$(W_t^\circ, M_t) = \left(e^{F^* t} \int_0^t H^* d\bar{W}_s, \int_0^t e^{F^* s} H^* H e^{F s} ds \right).$$

Let $\psi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a function that is integrable with respect to the joint law of (X_0, \tilde{X}_t) . Then,

$$\mathbb{E} [\psi(X_0, \tilde{X}_t) | \mathcal{Y}_t] = \frac{\bar{\mathbb{E}} [\psi(X_0, \tilde{X}_t) L_t | \mathcal{Y}_t]}{\bar{\mathbb{E}} [L_t | \mathcal{Y}_t]}.$$

From the above formula for \tilde{X}_t and the fact that X_0 and (\bar{W}, Y) are $\bar{\mathbb{P}}$ -independent, we deduce that

$$\begin{aligned} \mathbb{E} [\psi(X_0, \tilde{X}_t) | \mathcal{Y}_t] &= \frac{\int_{\mathbb{R}^d} e^{-1/2(M_t x, x)} \bar{\mathbb{E}} [\psi(x, \tilde{X}_t) \exp(\langle x, W^\circ \rangle) | \mathcal{Y}_t] \pi_0(dx)}{\int_{\mathbb{R}^d} e^{-1/2(M_t x, x)} \bar{\mathbb{E}} [\exp(\langle x, W^\circ \rangle) | \mathcal{Y}_t] \pi_0(dx)}. \quad (\text{B29}) \end{aligned}$$

In particular, for any function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that the random variable $\varphi(X_t) = \varphi(e^{Ft} X_0 + \tilde{X}_t)$ is integrable, we deduce from (B29) that

$$\begin{aligned} \mathbb{E} [\varphi(X_t) | \mathcal{Y}_t] &= \frac{\int_{\mathbb{R}^d} e^{-1/2(M_t x, x)} \bar{\mathbb{E}} [\varphi_x^1(\tilde{X}_t, W^\circ) | \mathcal{Y}_t] \pi_0(dx)}{\int_{\mathbb{R}^d} e^{-1/2(M_t x, x)} \bar{\mathbb{E}} [\varphi_x^2(\tilde{X}_t, W^\circ) | \mathcal{Y}_t] \pi_0(dx)}; \\ \varphi_x^1(a, b) &= \varphi(e^{Ft} x + a) \exp\langle x, b \rangle, \quad a \in \mathbb{R}^d, b \in \mathbb{R}^n; \quad \text{and} \\ \varphi_x^2(a, b) &= \exp\langle x, b \rangle, \quad a \in \mathbb{R}^d, b \in \mathbb{R}^n. \end{aligned}$$

The pair (\tilde{X}, W°) satisfies a linear system of stochastic differential equations driven by the $(d+m)$ -Brownian motion (V, W°) with initial condition $(0, 0)$. Moreover, under $\bar{\mathbb{P}}$, the process Y satisfies

$$Y_t = \int_0^t (H \tilde{X}_s + \tilde{h}) ds + \bar{W}_t. \quad (\text{B30})$$

It follows that we can express $\bar{\mathbb{E}} [\varphi_x^i(\tilde{X}_t, W_t^\circ) | \mathcal{Y}_t], i = 1, 2$, as integrals with respect to a Gaussian distribution η with mean vector $(\hat{x}_t^0, \hat{x}_t^W)^*$ and covariance matrix

$$C_t = \begin{pmatrix} P_t^0 & S_t \\ S_t^* & R_t \end{pmatrix}$$

that satisfy the equations satisfied by a Kalman–Bucy filter with the signal equations identical to those for the pair (\tilde{X}, W_t°) and the observation equation identical to (B30).

More precisely, \hat{x}_t^0 is the solution of Eq. (B26), P_t^0 is the solution of Eq. (B27), \hat{x}_t^W is the solution of the equation

$$d\hat{x}_t^W = (e^{Ft} + S_t)^* H^* (dY_t - (H \hat{x}_t^0 + \tilde{h}) dt), \quad \hat{x}_0^W = 0,$$

and the pair (S_t, Q_t) solve the equations

$$\frac{dS_t}{dt} = FS_t - P_t^0 H^* H (e^{Ft} + S_t), \quad S_0 = 0,$$

$$\frac{dR_t}{dt} = -e^{F^* t} H^* H S_t - S_t^* H^* H e^{Ft} - S_t^* H^* H S_t, \quad R_0 = 0;$$

see formulas (B4)–(B10) in Ref. 63 for details. One can check that

$$\frac{dS_t^\circ}{dt} = (F - P_t^0 H^* H) S_t^\circ = Q_t S_t^\circ,$$

where $S_t^\circ := e^{Ft} + S_t$ and, since Q_t is asymptotically stable, we conclude that

$$\lim_{t \rightarrow \infty} \|S_t^\circ\| = 0. \quad (\text{B31})$$

In fact, similar to Remark 2.8, the convergence in (B31) has an exponential decay rate to 0.

We deduce that

$$\bar{\mathbb{E}} [\varphi_x^2(\tilde{X}_t, W^\circ) | \mathcal{Y}_t] = \bar{\mathbb{E}} [\exp\langle x, W^\circ \rangle | \mathcal{Y}_t] = e^{\langle x, \hat{x}_t^W \rangle + 1/2(Q_t x, x)}, \quad (\text{B32})$$

$$\begin{aligned} \bar{\mathbb{E}} [\varphi_x^1(\tilde{X}_t, W^\circ) | \mathcal{Y}_t] &= \bar{\mathbb{E}} [\varphi(e^{Ft} x + \tilde{X}_t) \exp\langle x, W^\circ \rangle | \mathcal{Y}_t] \\ &= e^{\langle x, \hat{x}_t^W \rangle + 1/2(Q_t x, x)} \int \varphi(S_t^\circ x + a) \pi_t^0(da). \end{aligned}$$

It follows that

$$\pi_t(\varphi) = \mathbb{E} [\varphi(X_t) | \mathcal{Y}_t] = \frac{\int_{\mathbb{R}^d} \Xi_t(x) \int \varphi(S_t^\circ x + a) \pi_t^0(da) \pi_0(dx)}{\int_{\mathbb{R}^d} \Xi_t(x) \pi_0(dx)},$$

where $\Xi_t(x) := e^{-1/2(M_t x, x) + \langle x, \hat{x}_t^W \rangle + 1/2(Q_t x, x)}$. Moreover, by choosing a function $\psi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ in (B29) that is independent in the second component, $\psi(x, y) = \varphi(x)$, $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, we obtain from (B32) that

$$\mathbb{E} [\varphi(X_0) | \mathcal{Y}_t] = \frac{\int_{\mathbb{R}^d} \Xi_t(x) \varphi(x) \pi_0(dx)}{\int_{\mathbb{R}^d} \Xi_t(x) \pi_0(dx)} = \frac{\pi_0(\Xi_t \varphi)}{\pi_0(\Xi_t)}.$$

We identify the FA process $\lambda : [0, \infty) \times \mathcal{P}(\mathbb{R}^d) \times \Omega \rightarrow \mathcal{P}(\mathbb{R}^d)$ by the formula

$$\begin{aligned} \pi_t^\mu(\varphi) &= \lambda(t, \omega) \mu(\varphi) \\ &:= \frac{\int_{\mathbb{R}^d} \Xi_t(x) \int \varphi(S_t^\circ x + a) \pi_t^0(da) \mu(dx)}{\int_{\mathbb{R}^d} \Xi_t(x) \mu(dx)} \\ &= \frac{1}{C^{\pi_t^\mu}} \int_{\mathbb{R}^d} \Xi_t(x) \int \varphi(S_t^\circ x + a) \pi_t^0(da) \theta_\mu(x) \pi_0(dx), \end{aligned} \quad (\text{B33})$$

where $C^{\pi_t^\mu}$ is the normalization constant

$$\begin{aligned} C^{\pi_t^\mu} &:= \int_{\mathbb{R}^d} \Xi_t(x) \theta_\mu(x) \pi_0(dx) = \pi_0(\Xi_t \theta_\mu) \\ &= \mathbb{E} [\theta_\mu(X_0) | \mathcal{Y}_t] \pi_0(\Xi_t). \end{aligned}$$

To obtain representation (B33), we used the fact that μ is absolutely continuous with respect to π_0 and that θ_μ is the density of μ with respect to π_0 .

Convergence of the first moments. Observe that

$$\begin{aligned}\hat{\pi}_t^\mu &= \frac{1}{C^{\pi_t^\mu}} \int_{\mathbb{R}^d} \Xi(x) \int (S_t^\circ x + a) \pi_t^0(da) \theta_\mu(x) \pi_0(dx) \\ &= \hat{x}_t^0 + \frac{S_t^\circ \int_{\mathbb{R}^d} x \Xi_t(x) \theta_\mu(x) \pi_0(dx)}{\int_{\mathbb{R}^d} \Xi_t(x) \theta_\mu(x) \pi_0(dx)} \\ &= \hat{x}_t^0 + \frac{S_t^\circ (\mathbb{E}[\theta_\mu(X_0) X_0 | \mathcal{F}_t] \pi_0(\Xi_t))}{(\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t] \pi_0(\Xi_t))}.\end{aligned}\quad (\text{B34})$$

It follows that

$$|\hat{\pi}_t^\mu - \hat{x}_t^0| \leq \frac{\|S_t^\circ\| \|\mathbb{E}[\theta_\mu(X_0) X_0 | \mathcal{F}_t]\|}{\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]}.$$

Since the processes $t \rightarrow \mathbb{E}[\theta_\mu(X_0) X_0 | \mathcal{F}_t]$ and $t \rightarrow \mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]$ converge and the second limit is positive, we have, by using (B31),

$$\lim_{t \rightarrow \infty} |\hat{\pi}_t^\mu - \hat{x}_t^0| = \lim_{t \rightarrow \infty} \|S_t^\circ\| \frac{\lim_{t \rightarrow \infty} \|\mathbb{E}[\theta_\mu(X_0) X_0 | \mathcal{F}_t]\|}{\lim_{t \rightarrow \infty} \|\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]\|} = 0. \quad (\text{B35})$$

Again, similar to Remark 2.8, the convergence in (B35) is exponentially fast. Since $\lim_{t \rightarrow \infty} e^{-\varepsilon t} |\pi_t^\mu| = \lim_{t \rightarrow \infty} e^{-\varepsilon t} |\hat{x}_t^0| = 0$, for any $\varepsilon > 0$, it follows that, for any $i, j = 1, \dots, d$,

$$\lim_{t \rightarrow \infty} |(\pi_t^\mu)^i (\pi_t^\mu)^j - (\hat{x}_t^0)^i (\hat{x}_t^0)^j| = 0.$$

The last limit is used in the proof of the convergence of the covariance matrix below.

Convergence of the covariance matrix. Choose

$$\varphi^{ij} : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \varphi^{ij}(x^1, \dots, x^d) := x^i x^j - (\pi_t^\mu)^i (\pi_t^\mu)^j.$$

We have that

$$\begin{aligned}(P_{\pi_t^\mu})^{ij} - (P_t^0)^{ij} &= \pi_t^\mu(\varphi^{ij}) - (P_t^0)^{ij} \\ &= \frac{1}{C^{\pi_t^\mu}} \int_{\mathbb{R}^d} \Xi(x) \int \varphi^{ij}((S_t^\circ x + a) \pi_t^0(da) \\ &\quad \times \theta_\mu(x) \pi_0(dx) - (P_t^0)^{ij} \\ &= \sum_{k,l} \frac{(S_t^\circ)^{il} (S_t^\circ)^{jk} (\mathbb{E}[\theta_\mu(X_0) X_0^l X_0^k | \mathcal{F}_t])}{\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]} \\ &\quad + \sum_l \frac{(S_t^\circ)^{il} (\mathbb{E}[\theta_\mu(X_0) X_0^l | \mathcal{F}_t])}{\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]} (\hat{x}_t^0)^j \\ &\quad + \sum_l \frac{(S_t^\circ)^{jl} (\mathbb{E}[\theta_\mu(X_0) X_0^l | \mathcal{F}_t])}{\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]} (\hat{x}_t^0)^i \\ &\quad + (\hat{x}_t^0)^i (\hat{x}_t^0)^j - (\pi_t^\mu)^i (\pi_t^\mu)^j,\end{aligned}$$

which gives the required convergence to 0.

Convergence for bounded uniformly continuous test functions. We have that

$$|\pi_t^\mu(\varphi_t) - \pi_t^0(\varphi_t)| \leq \frac{1}{C^{\pi_t^\mu}} \int_{\mathbb{R}^d} \Xi(x) \int |\varphi_t((S_t^\circ x + a) - \varphi_t(a)| \times \pi_t^0(da) \theta_\mu(x) \pi_0(dx).$$

Decomposing next the integral in the numerator into the sum of the integral over the region $|(S_t^\circ x)| < \delta$ and the integral over the region $|(S_t^\circ x)| \geq \delta$ yields

$$\begin{aligned}|\pi_t^\mu(\varphi_t) - \pi_t^0(\varphi_t)| &\leq \sup_{|y-y'| < \delta} |\varphi(y) - \varphi(y')| \\ &\quad + 2\|\varphi\|_\infty \frac{(\mathbb{E}[\theta_\mu(X_0) \mathbf{1}_{\{|(S_t^\circ x_0)| \geq \delta\}} | \mathcal{F}_t])}{\mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]} \\ &\leq \sup_{|y-y'| < \delta} |\varphi(y) - \varphi(y')| \\ &\quad + 2 \frac{\|\varphi\|_\infty (S_t^\circ) \mathbb{E}[\theta_\mu(X_0) | X_0|^2 | \mathcal{F}_t]}{\delta^2 \mathbb{E}[\theta_\mu(X_0) | \mathcal{F}_t]}.\end{aligned}$$

As above, it follows that

$$\limsup_{t \rightarrow \infty} |\pi_t^\mu(\varphi_t) - \pi_t^0(\varphi_t)| \leq \sup_{|y-y'| < \delta} |\varphi(y) - \varphi(y')|$$

and then using the uniform continuity of φ , we deduce the result and thus complete the proof. \square

4. A simple example

In this subsection, we provide an illustrative example of unstable dynamics and partial observations that still yield convergence of the FA process (see also Refs. 14 and 95). We choose a two-dimensional linear signal (X^1, X^2) with $\lambda_1 < 0$ and $\lambda_2 > 0$,

$$X_t^i = x_0^i + \int_0^t \lambda_i X_s^i ds + \int_0^t \sigma^i dW_s^i, \quad i = 1, 2,$$

so that the law of (X^1, X^2) is not stable. More precisely, if we choose two systems (X^1, X^2) , $(\tilde{X}^1, \tilde{X}^2)$ starting from (x_0^1, x_0^2) and $(\tilde{x}_0^1, \tilde{x}_0^2)$, respectively, then their corresponding expected values drift away from each other. In particular,

$$\lim_{t \rightarrow \infty} |E[X_t^2] - E[\tilde{X}_t^2]| = \lim_{t \rightarrow \infty} e^{\lambda_2 t} |x_0^2 - \tilde{x}_0^2| = \infty.$$

As a result, the Wasserstein distance W_2 between p_t and \tilde{p}_t tends to ∞ .

In fact, one can easily show that the first component is stable, while the second one is unstable. However, we can stabilize the system by observing *only* the second component, namely, the unstable one. For example, we can choose a one-dimensional observation process of the form

$$dY_t = hX_t^2 dt + dW_t,$$

which will guarantee that

$$\lim_{t \rightarrow \infty} d_W(\pi_t, \tilde{\pi}_t) = 0.$$

One can justify this either by checking that the assumptions of Theorem 3.2 are satisfied or by explicit calculation. Using the second approach, one notices that the probability measures π_t and $\tilde{\pi}_t$ are both Gaussian and shows that (i) the distance between the corresponding means converges to 0 and (ii) that the covariance matrices of π_t and $\tilde{\pi}_t$ coincide. The latter two matrices are given by $\text{diag}(q_t^{11}, q_t^{22})$. They are diagonal and

$$\begin{aligned} \frac{d}{dt} \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix} \\ &+ \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ &- \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & h^2 \end{pmatrix} \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix}, \\ \frac{d}{dt} \begin{pmatrix} q_t^{11} & 0 \\ 0 & q_t^{22} \end{pmatrix} &= \begin{pmatrix} 1 + \lambda_1 p_t^{22} & 0 \\ 0 & 1 + 2\lambda_2 p_t^{11} - h^2 (p_t^{11})^2 \end{pmatrix}, \end{aligned}$$

which implies that

$$\lim_{t \rightarrow \infty} \begin{pmatrix} p_t^{11} \\ p_t^{22} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2\lambda_1} \\ \frac{\sqrt{(\lambda_2)^2 + h^2} - \lambda_2}{h^2} \end{pmatrix},$$

and the limit is valid independent of the initial condition. This result together with the Gaussianity property of the processes π_t and $\tilde{\pi}_t$ implies the convergence in Wasserstein distance.

In the reverse situation, in which we would observe the stable component and not the unstable one, the posterior distribution would not be stable.

In the above example, one can still apply existing results to the one-dimensional unstable system and, after coupling it to the unobserved component, obtain the stability of the pair. This works because the observation process depends only on the observed component. In multidimensional, nonlinear applications with very large dimensions d and n , though, it would be quite difficult to do this as it may not be possible to identify the unstable components *a priori*.

The result presented in this paper does not assume that one needs to identify a system's unstable components and observe those. For example, if the observation process has the form

$$dY_t = h(X_t^1 + X_t^2) dt + dW_t,$$

the FA process will still stabilize the system. More precisely, it is still the case that both π_t and $\tilde{\pi}_t$ are Gaussian and that

$$\lim_{t \rightarrow \infty} d_W(\pi_t, \tilde{\pi}_t) = 0.$$

5. The forecast-assimilation (FA) process as a random dynamical system (RDS)

In this subsection, we give a brief justification of the fact that the solution of (11) can be recast as an RDS evolving in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$. This result is related to similar approaches in the nonlinear filtering literature (see, for instance, Refs. 57, 88, 64, 63, etc. The cornerstone of the argument is that the

solution of equation (11) can be expressed as

$$\pi_t^\mu = \lambda(t, \omega) \mu = \frac{1}{C^{\pi_0}} \rho_{0:t}(\mu), \quad (\text{B36})$$

where $\rho_{t_0:t_1}(\mu)$ is a two-parameter measure-valued process defined as

$$\rho_{t_0:t_1}(\mu)(\varphi) = \int_{\mathbb{R}^d} \tilde{\mathbb{E}}_{t_0,x}[\tilde{Z}_{t_0:t_1} \varphi(\tilde{X}_{t_0:t_1}) | \mathcal{Y}] \mu(dx), \quad (\text{B37})$$

where φ is an arbitrary bounded Borel measurable map and C^μ is the normalization constant $C^\mu = \rho_{t_0:t_1}(\mu)(1)$. In (B37), we have the following definitions:

(i) the process $\tilde{Z}_{t_0:t_1} = \{\tilde{Z}_{t_0:t_1}, t \geq 0\}$ is defined by

$$\tilde{Z}_{t_0:t_1} = \exp \left(\sum_{i=1}^m \int_{t_0}^{t_1} h^i(\tilde{X}_{t_0,s}) dY_s^i - \frac{1}{2} \sum_{i=1}^m \int_{t_0}^{t_1} h^i(\tilde{X}_{t_0,s})^2 ds \right); \quad (\text{B38})$$

- (ii) the process $\tilde{X}_{t_0,(\cdot)} = \{\tilde{X}_{t_0,t_1}, t_0, t_1 \geq 0\}$ is a stochastic process independent of Y satisfying the signal Eq. (8) on $[t_0, \infty)$; and
(iii) $\tilde{\mathbb{E}}_{t_0,x}$ is the expectation with respect to a probability measure $\mathbb{P}_{t_0,x}$ under which Y is a Brownian motion independent of \tilde{X} and $\tilde{X}_{t_0,t_0} \equiv x$.

The independence of \tilde{X} and Y under $\mathbb{P}_{t_0,x}$ in formula (B36) enables us to show that $\rho_{t_0:t_1}$ is an RDS, which will immediately imply that $\pi_{t_0:t_1}$ is one, too. To justify this, we introduce $\Theta^{Y(\cdot)}(t_0, t_1)$ to be the following two-parameter family of random variables:

$$\begin{aligned} \Theta^{Y(\cdot)}(t_0, t_1) &\triangleq \exp \left(h(\tilde{X}_{t_0,t_1})^\top y_{t_1} - h(\tilde{X}_{t_0,t_0})^\top y_{t_0} + \tilde{I}_{t_0,t_1}^{Y(\cdot)} \right. \\ &\quad \left. - \frac{1}{2} \sum_{i=1}^m \int_{t_0}^{t_1} h^i(\tilde{X}_{t_0,s})^2 ds \right), \end{aligned} \quad (\text{B39})$$

where $\tilde{I}_{t_0,t_1}^{Y(\cdot)}$ is a version of the stochastic integral $\int_{t_0}^{t_1} y_s^\top dh(\tilde{X}_{t_0,s})$ and $y_{(\cdot)}$ is a continuous path, $y_{(\cdot)} \in C_{\mathbb{R}^m}[0, \infty)$. The argument of the exponent in the definition of $\Theta^{Y(\cdot)}(t_0, t_1)$ is recognizable as a formal integration by parts of the argument of the exponential in (B37).

Let $\rho_{t_0,t_1}^{Y(\cdot)}(\mu)$ and $\pi_{t_0,t_1}^{Y(\cdot)}(\mu)$ be the following two-parameter measure-valued processes:

$$\rho_{t_0,t_1}^{Y(\cdot)}(\mu)(\varphi) = \int_{\mathbb{R}^d} \tilde{\mathbb{E}}_{t_0,x}[\varphi(\tilde{X}_{t_0,t_1}) \Theta^{Y(\cdot)}(t_0, t_1) | \mathcal{Y}] \mu(dx), \quad (\text{B40a})$$

$$\pi_{t_0,t_1}^{Y(\cdot)}(\mu)(\varphi) = \frac{\rho_{t_0,t_1}^{Y(\cdot)}(\mu)(\varphi)}{\rho_{t_0,t_1}^{Y(\cdot)}(\mu)(1)}. \quad (\text{B40b})$$

Then, $\rho_{t_0,t_1}^{Y(\cdot)}(\mu)$ and $\pi_{t_0,t_1}^{Y(\cdot)}(\mu)$ are versions of $\rho_{t_0,t_1}(\mu)$ and $\pi_{t_0,t_1}(\mu)$.

Since $\rho_{t_0,t_1}^{Y(\cdot)}, \pi_{t_0,t_1}^{Y(\cdot)}$ can be recast as time-inhomogeneous dynamical systems, we can use them as a basis for defining $\rho_{t_0,t_1}(\mu)$ and $\pi_{t_0,t_1}(\mu)$. As a result, it is indeed the case that $\rho_{t_0,t_1}(\mu)$ and $\pi_{t_0,t_1}(\mu)$ can be viewed as RDSs. Moreover, one can show that

$\rho_{t_0, \cdot}(\mu) = \{\rho_{t_0, t_1}(\mu), t_1 \geq t_0\}$ satisfies the evolution equation

$$\rho_{t_0, t_1}(\mu)(\varphi) = \mu(\varphi) + \int_{t_0}^{t_1} \rho_{t_0, s}(\mu)(A\varphi) ds + \int_{t_0}^{t_1} \rho_{t_0, s}(\mu)(\varphi h^\top) dY_s \quad (\text{B41})$$

for any $\varphi \in \mathcal{D}(A)$ and, in particular, that

$$\begin{aligned} \rho_{t_0, t_1}(\mu)(1) &= 1 + \int_{t_0}^{t_1} \rho_{t_0, s}(\mu)(h^\top) dY_s \\ &= 1 + \int_{t_0}^{t_1} \rho_{t_0, s}(\mu)(1) \pi_{t_0, s}(\mu)(h^\top) dY_s. \end{aligned} \quad (\text{B42})$$

From (B41) and (B42), one deduces that the ratio $\rho_{t_0, t_1}(\mu)(\varphi)/\rho_{t_0, t_1}(\mu)(1)$ satisfies (11) with initial condition $\pi_{t_0, t_0}(\mu) = \mu$ and, therefore, by the uniqueness of the solution of (11), we obtain that the solution of (11) can indeed be recast as an RDS.

Finally, notice that $\pi_{0, t}(\pi_0) = \pi_{t_0, t}(\pi_{t_0})$ is indeed the FA process we consider throughout this paper. For the particular case of $\mu = \pi_{t_0}$, formula (B36) is known as the Kallianpur–Striebel formula⁵⁸ and it is deduced directly from the definition of the conditional expectation. For arbitrary μ , (B36) serves as definition for the RDS, which is then shown to be the solution of the evolution Eq. (11) starting from μ at time t_0 .

Remark 2.9: The map $\pi_{t_0, t_1}^{(\cdot)} : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d)$ is a continuous map when we endow $\mathcal{P}(\mathbb{R}^d)$ or, rather, the set of probability measures with finite second moment, with the topology induced by the Wasserstein metric.

REFERENCES

- ¹V. Bjerknes, “Das problem der wettvorhersage, betrachtet vom standpunkte der mechanik und der physik,” *Meteorol. Z.* **21**, 1–7 (1904).
- ²L. Bengtsson, M. Ghil, and E. Källén, *Dynamic Meteorology: Data Assimilation Methods* (Springer, 1981).
- ³R. A. Panofsky, “Objective weather-map analysis,” *J. Meteorol.* **6**(6), 386–392 (1949).
- ⁴J. G. Charney, M. Halem, and R. Jastrow, “Use of incomplete historical data to infer the present state of the atmosphere,” *J. Atmos. Sci.* **26**, 1160–1163 (1969).
- ⁵M. Ghil, M. Halem, and R. Atlas, “Time-continuous assimilation of remote-sounding data and its effect on weather forecasting,” *Mon. Weather Rev.* **107**, 140–171 (1979).
- ⁶A. Carrassi, C. Grudzien, M. Bocquet, J. Demayer, P. Raanes, and S. Vannitsem, “Data assimilation for chaotic systems,” in *Data Assimilation for Atmospheric, Oceanic and Hydrological Applications*, edited by S.-K. Park and X. Liang (Springer Science & Business Media, 2020).
- ⁷B. Cushman-Roisin and J.-M. Beckers, *Introduction to Geophysical Fluid Dynamics: Physical and Numerical Aspects*, 2nd ed. (Academic Press, 2011), pp. 875.
- ⁸M. Ghil and P. Malanotte-Rizzoli, “Data assimilation in meteorology and oceanography,” *Adv. Geophys.* **33**, 141–266 (1991).
- ⁹G. A. Gottwald and S. Reich, “Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations,” *Chaos* **31**(10), 101103 (2021).
- ¹⁰E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, 2003).
- ¹¹M. Ghil and S. Childress, *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics* (Springer Science+Business Media, Berlin/, 1987). Reissued in pdf, 2012.
- ¹²E. N. Lorenz, “Deterministic nonperiodic flow,” *J. Atmos. Sci.* **20**, 130–141 (1963).
- ¹³A. Farhat, E. Lunasin, and E. S. Titi, “A data assimilation algorithm: The paradigm of the 3D Leray- α model of turbulence,” in *Partial Differential*

Equations Arising from Physics and Geometry (Cambridge University Press, 2019), pp. 253–273.

- ¹⁴A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi, “Data assimilation as a non-linear dynamical systems problem: Stability and convergence of the prediction-assimilation system,” *Chaos* **18**(2), 023112 (2008).
- ¹⁵A. Trevisan and F. Uboldi, “Assimilation of standard and targeted observations within the unstable subspace of the observation–analysis–forecast cycle system,” *J. Atmos. Sci.* **61**(1), 103–113 (2004).
- ¹⁶I. Karatzas and S. E. Shreve, “Brownian motion and stochastic calculus,” in *Graduate Texts in Mathematics* (Springer-Verlag, New York, 1988), Vol. 113.
- ¹⁷L. Arnold, *Random Dynamical Systems* (Springer-Verlag, New York, 1998).
- ¹⁸T. Caraballo and X. Han, *Applied Nonautonomous and Random Dynamical Systems: Applied Dynamical Systems* (Springer Science+Business Media, 2017).
- ¹⁹H. Crauel and F. Flandoli, “Attractors for random dynamical systems,” *Probab. Theory Relat. Fields* **100**(3), 365–393 (1994).
- ²⁰A. Bain and D. Crisan, “Fundamentals of stochastic filtering,” in *Stochastic Modelling and Applied Probability* (Springer, New York, 2009), Vol. 60.
- ²¹*The Oxford Handbook of Nonlinear Filtering*, edited by D. Crisan and B. Rozovskii (Oxford University Press, Oxford, 2011).
- ²²M. Ghil and R. Todling, “Tracking atmospheric instabilities with the Kalman filter. Part II: Two-layer results,” *Mon. Weather Rev.* **124**, 2340–2352 (1996).
- ²³M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications* (SIAM, Philadelphia, PA, 2016).
- ²⁴K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction* (Springer, Cham, 2015).
- ²⁵S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation* (Cambridge University Press, New York, 2015).
- ²⁶C. Nicolis, R. A. P. Perdigao, and S. Vannitsem, “Dynamics of prediction errors under the combined effect of initial condition and model errors,” *J. Atmos. Sci.* **66**(3), 766–778 (2009).
- ²⁷R. Kalman, “A new approach to linear filtering and prediction problems,” *ASME J. Basic Eng.* **82D**, 35–45 (1960).
- ²⁸R. Kalman and R. Bucy, “New results in linear filtering and prediction theory,” *ASME J. Basic Eng.* **83D**, 95–108 (1961).
- ²⁹A. Gelb, *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974).
- ³⁰A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Dover Publications, 1970). Reprinted by Courier Corp. in 2007.
- ³¹G. S. Duane, C. Grabow, F. Seltén, and M. Ghil, “Introduction to focus issue: Synchronization in large networks and continuous media–data, models, and supermodels,” *Chaos* **27**(12), 126601 (2017).
- ³²G. D. Charó, M. D. Chekroun, D. Sciamarella, and M. Ghil, “Topological effects of noise on nonlinear dynamics,” *arXiv:2010.09611v5* (2021).
- ³³M. Ghil, M. D. Chekroun, and E. Simonnet, “Climate dynamics and fluid mechanics: Natural variability and related uncertainties,” *Physica D* **237**(14–17), 2111–2126 (2008).
- ³⁴A. 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**(647), 487–496 (2010).
- ³⁵M. Ghil, “Advances in sequential estimation for atmospheric and oceanic flows,” *J. Meteorol. Soc. Japan. Ser. II* **75**(1B), 289–304 (1997).
- ³⁶J. Frank and S. Zhuk, “A detectability criterion and data assimilation for nonlinear differential equations,” *Nonlinearity* **31**(11), 5235 (2018).
- ³⁷M. Ghil and V. Lucarini, “The physics of climate variability and climate change,” *Rev. Mod. Phys.* **92**(3), 1–77 (2020).
- ³⁸E. N. Lorenz, *The Nature and Theory of the General Circulation of the Atmosphere* (World Meteorological Organization Geneva, 1967), Vol. 218.
- ³⁹*Extreme Events: Observations, Modeling, and Economics*, edited by M. Chavez, M. Ghil, and J. Urrutia-Fucugauchi (John Wiley & Sons, 2015), Vol. 214.
- ⁴⁰M. Ghil, P. Yiou, S. Hallegatte, B. D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H. W. Rust, P. Shebalin, M. Vrac, A. Witt, and I. Zaliapin, “Extreme events: Dynamics, statistics and prediction,” *Nonlinear Process. Geophys.* **18**(3), 295–350 (2011).
- ⁴¹M. Bocquet, C. A. Pires, and L. Wu, “Beyond Gaussian statistical modeling in geophysical data assimilation,” *Mon. Weather Rev.* **138**(8), 2997–3023 (2010).

- ⁴²P. J. Van Leeuwen, "Particle filtering in geophysical systems," *Mon. Weather Rev.* **137**(12), 4089–4114 (2009).
- ⁴³M. Ghil, B. Shkoller, and V. Yangarber, "A balanced diagnostic system compatible with a barotropic prognostic model," *Mon. Weather Rev.* **105**(10), 1223–1238 (1977).
- ⁴⁴A. Farhat, M. S. Jolly, and E. S. Titi, "Continuous data assimilation for the 2D Bénard convection through velocity measurements alone," *Physica D* **303**, 59–66 (2015).
- ⁴⁵M. Ghil, S. E. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, "Applications of estimation theory to numerical weather prediction," in *Dynamic Meteorology: Data Assimilation Methods*, edited by L. Bengtsson, M. Ghil, and E. Källén (Springer, 1981), pp. 139–224.
- ⁴⁶N. Wax, *Selected Papers on Noise and Stochastic Processes* (Courier Dover Publications, 1954).
- ⁴⁷M. Halem, E. Kalnay, W. E. Baker, and R. Atlas, "An assessment of the FGGE satellite observing system during SOP-1," *Bull. Am. Meteorol. Soc.* **63**(4), 407–426 (1982).
- ⁴⁸G. P. Cressman, "An operational objective analysis system," *Mon. Weather Rev.* **87**(10), 367–374 (1959).
- ⁴⁹E. N. Lorenz, "Atmospheric predictability experiments with a large numerical model," *Tellus* **34**(6), 505–513 (1982).
- ⁵⁰A. Dalcher and E. Kalnay, "Error growth and predictability in operational ECMWF forecasts," *Tellus A* **39A**(5), 474–491 (1987).
- ⁵¹K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc, "Unified notation for data assimilation: Operational, sequential and variational, in *Special Issue on Data Assimilation in Meteorology and Oceanography: Theory and Practice*," *J. Meteorol. Soc. Jpn. Ser. II* **75**(1B), 181–189 (1997).
- ⁵²G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Hist. l'Acad. R. Sci.* 666–704 (1781).
- ⁵³M. Ghil, "A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability?," in *Climate Change: Multidecadal and Beyond*, edited by C.-P. Chang, M. Ghil, M. Latif, and J. Wallace (World Scientific Publishing Co., Singapore, 2015), Vol. 6, pp. 31–52.
- ⁵⁴V. M. Panaretos and Y. Zemel, *An Invitation to Statistics in Wasserstein Space* (Springer Nature, 2020).
- ⁵⁵L. Arnold, "Stabilization by noise revisited," *Z. Angew. Math. Mech.* **70**(7), 235–246 (1990).
- ⁵⁶P. J. Van Leeuwen, Y. Cheng, and S. Reich, *Nonlinear Data Assimilation* (Springer, 2015).
- ⁵⁷A. Budhiraja, "Feller and stability properties of the nonlinear filter," in *The Oxford Handbook of Nonlinear Filtering* (Oxford University Press, 2011), pp. 352–373.
- ⁵⁸P. Del Moral, "Nonlinear filtering using random particles," *Theory Probab. Appl.* **40**(4), 690–701 (1996).
- ⁵⁹P. Del Moral, "Mean field simulation for Monte Carlo integration," in *Monographs on Statistics and Applied Probability* (CRC Press, Boca Raton, FL, 2013), Vol. 126.
- ⁶⁰P. Del Moral, A. Doucet, and S. S. Singh, "Uniform stability of a particle approximation of the optimal filter derivative," *SIAM J. Control Optim.* **53**(3), 1278–1304 (2015).
- ⁶¹P. Del Moral and L. Miclo, "On the stability of nonlinear Feynman-Kac semigroups," *Ann. Fac. Sci. Toulouse Math.* **(6)** **11**(2), 135–175 (2002).
- ⁶²H. Kunita, "Asymptotic behavior of the nonlinear filtering errors of Markov processes," *J. Multivariate Anal.* **1**, 365–393 (1971).
- ⁶³D. Ocone and E. Pardoux, "Asymptotic stability of the optimal filter with respect to its initial condition," *SIAM J. Control Optim.* **34**(1), 226–243 (1996).
- ⁶⁴J. Picard, "Efficiency of the extended Kalman filter for nonlinear systems with small noise," *SIAM J. Appl. Math.* **51**(3), 843–885 (1991).
- ⁶⁵R. van Handel, "Uniform observability of hidden Markov models and filter stability for unstable signals," *Ann. Appl. Probab.* **19**(3), 1172–1199 (2009).
- ⁶⁶R. Atar, "Exponential decay rate of the filter's dependence on the initial distribution," in *The Oxford Handbook of Nonlinear Filtering* (Oxford University Press, Oxford, 2011), pp. 299–318.
- ⁶⁷R. Atar and O. Zeitouni, "Exponential stability for nonlinear filtering," *Ann. Inst. H. Poincaré Probab. Stat.* **33**(6), 697–725 (1997).
- ⁶⁸P. Chigansky and R. Liptser, "Stability of nonlinear filters in nonmixing case," *Ann. Appl. Probab.* **14**(4), 2038–2056 (2004).
- ⁶⁹P. Chigansky, R. Liptser, and R. Van Handel, "Intrinsic methods in filter stability," in *The Oxford Handbook of Nonlinear Filtering* (Oxford University Press, Oxford, 2011), pp. 319–351.
- ⁷⁰M. D. Checkroun, E. Simmonet, and M. Ghil, "Stochastic climate dynamics: Random attractors and time-dependent invariant measures," *Physica D* **240**, 1685–1700 (2011).
- ⁷¹D. Bakry, P. Cattiaux, and A. Guillin, "Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré," *J. Funct. Anal.* **254**(3), 727–759 (2008).
- ⁷²D. Crisan and K. Heine, "Stability of the discrete time filter in terms of the tails of noise distributions," *J. Lond. Math. Soc.* **(2)** **78**(2), 441–458 (2008).
- ⁷³W. Stannat, "Stability of the optimal filter for nonergodic signals—A variational approach," in *The Oxford Handbook of Nonlinear Filtering* (Oxford University Press, Oxford, 2011), pp. 374–399.
- ⁷⁴W. Stannat, "Stability of the filter equation for a time-dependent signal on \mathbb{R}^d ," *Appl. Math. Optim.* **52**(1), 39–71 (2005).
- ⁷⁵D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, "A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case," *Electron. Commun. Probab.* **13**, 60–66 (2008).
- ⁷⁶G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli* **2**(4), 341–363 (1996).
- ⁷⁷A. N. Bishop and P. Del Moral, "On the stability of matrix-valued Riccati diffusions," *Electron. J. Probab.* **24**, 40 (2019).
- ⁷⁸A. N. Bishop and P. Del Moral, "An explicit floquet-type representation of Riccati aperiodic exponential semigroups," *Int. J. Control* **94**(1), 258–266 (2021).
- ⁷⁹M. Bocquet and A. Carrassi, "Four-dimensional ensemble variational data assimilation and the unstable subspace," *Tellus A* **69**(1), 1304504 (2017).
- ⁸⁰M. 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 J. Uncertain. Quantif.* **5**(1), 304–333 (2017).
- ⁸¹A. Carrassi, M. Bocquet, J. Demaeyer, C. Grudzien, P. Raanes, and S. Vannitsem, "Data assimilation for chaotic dynamics," in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)*, edited by S. K. Park and L. Xu (Springer International Publishing, Cham, 2022), pp. 1–42.
- ⁸²C. Grudzien, A. Carrassi, and M. Bocquet, "Asymptotic forecast uncertainty and the unstable subspace in the presence of additive model error," *SIAM/ASA J. Uncertain. Quantif.* **6**(4), 1335–1363 (2018).
- ⁸³K. 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 J. Control Optim.* **55**(2), 741–759 (2017).
- ⁸⁴S. K. Park and L. Xu, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (SIAM, Philadelphia, PA, 2012), Vol. IV.
- ⁸⁵W. Liu and M. Röckner, *Stochastic Partial Differential Equations: An Introduction* (Universitext. Springer, Cham, 2015).
- ⁸⁶B. L. Rozovsky and S. V. Lototsky, "Stochastic evolution systems," in *Probability Theory and Stochastic Modelling* (Springer, Cham, 2018), Vol. 89. Linear theory and applications to non-linear filtering, 2nd ed. of [MR1135324].
- ⁸⁷N. Ikeda and S. Watanabe, "Stochastic differential equations and diffusion processes," *North-Holland Mathematical Library* (North-Holland Publishing Co., Amsterdam, 1981), Vol. 24.
- ⁸⁸H. Kunita, "Stochastic flows and stochastic partial differential equations," in *Proceedings of the International Congress of Mathematicians, Berkeley, CA, 1986*, (Amer. Math. Soc., Providence, RI, 1987), Vols. 1 and 2, pp. 1021–1031.
- ⁸⁹F. Flandoli, "Some remarks on a statistical theory of turbulent flows," in *Probabilistic Methods in Fluids* (World Sci. Publ., River Edge, NJ, 2003), pp. 144–160.
- ⁹⁰P. Del Moral and J. Tugaut, "On the stability and the uniform propagation of chaos properties of ensemble Kalman-Bucy filters," *Ann. Appl. Probab.* **28**(2), 790–850 (2018).
- ⁹¹A. N. Bishop, P. Del Moral, K. Kamatani, and B. Rémillard, "On one-dimensional Riccati diffusions," *Ann. Appl. Probab.* **29**(2), 1127–1187 (2019).

- ⁹²A. N. Bishop, P. Del Moral, and A. Niclas, "A perturbation analysis of stochastic matrix Riccati diffusions," *Ann. Inst. Henri Poincaré Probab. Stat.* **56**(2), 884–916 (2020).
- ⁹³T. Karvonen, S. Bonnabel, E. Moulines, and S. Särkkä, "On stability of a class of filters for nonlinear stochastic systems," *SIAM J. Control Optim.* **58**(4), 2023–2049 (2020).
- ⁹⁴A. 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 Process. Geophys.* **15**(4), 503–521 (2008).
- ⁹⁵A. 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).
- ⁹⁶P. Houtekamer and F. Zhang, "Review of the ensemble Kalman filter for atmospheric data assimilation," *Mon. Weather Rev.* **144**(12), 4489–4532 (2016).
- ⁹⁷W. Lawson and J. Hansen, "Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth," *Mon. Weather Rev.* **132**(8), 1966–1981 (2004).
- ⁹⁸I. Hoteit, D.-T. Pham, M. E. Gharamti, and X. Luo, "Mitigating observation perturbation sampling errors in the stochastic EnKF," *Mon. Weather Rev.* **143**(7), 2918–2936 (2015).
- ⁹⁹P. J. Leeuwen, "A consistent interpretation of the stochastic version of the ensemble Kalman filter," *Q. J. R. Meteorol. Soc.* **146**(731), 2815–2825 (2020).
- ¹⁰⁰E. Bach and M. Ghil, "A multi-model ensemble Kalman filter for data assimilation and forecasting," *arXiv:2202.02272* (2022).
- ¹⁰¹H. D. I. Abarbanel, S. Shirman, D. Breen, N. Kadakia, D. Rey, E. Armstrong, and D. Margoliash, "A unifying view of synchronization for data assimilation in complex nonlinear networks," *Chaos* **27**(12), 126802 (2017).
- ¹⁰²G. S. Duane, J. J. Tribbia, and J. B. Weiss, "Synchronicity in predictive modelling: A new view of data assimilation," *Nonlinear Process. Geophys.* **13**(6), 601–612 (2006).
- ¹⁰³X. XX, "Sequential Monte Carlo methods in practice," in *Statistics for Engineering and Information Science*, edited by A. Doucet, N. de Freitas, and N. Gordon (Springer-Verlag, New York, 2001).
- ¹⁰⁴S. Vetra-Carvalho, P. J. van Leeuwen, L. Nerger, A. Barth, M. U. Altaf, P. Brasseur, P. Kirchgeßner, and J. M. Beckers, "State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems," *Tellus A* **70**(1), 1–43 (2018).
- ¹⁰⁵A. Beskos, D. Crisan, and A. Jasra, "On the stability of sequential Monte Carlo methods in high dimensions," *Ann. Appl. Probab.* **24**(4), 1396–1445 (2014).
- ¹⁰⁶C. Cotter, D. Crisan, D. Holm, W. Pan, and I. Shevchenko, "Data assimilation for a quasi-geostrophic model with circulation-preserving stochastic transport noise," *J. Stat. Phys.* **179**(5–6), 1186–1221 (2020).
- ¹⁰⁷C. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko, "A particle filter for stochastic advection by Lie transport: A case study for the damped and forced incompressible two-dimensional Euler equation," *SIAM/ASA J. Uncertain. Quantif.* **8**(4), 1446–1492 (2020).
- ¹⁰⁸N. Kantas, A. Beskos, and A. Jasra, "Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier-Stokes equations," *SIAM/ASA J. Uncertain. Quantif.* **2**(1), 464–489 (2014).
- ¹⁰⁹R. Potthast, A. Walter, and A. Rhodin, "A localized adaptive particle filter within an operational NWP framework," *Mon. Weather Rev.* **147**, 345–362 (2019).
- ¹¹⁰C. Cotter, D. Crisan, D. D. Holm, W. Pan, and I. Shevchenko, "Numerically modeling stochastic Lie transport in fluid dynamics," *Multiscale Model. Simul.* **17**(1), 192–232 (2019).
- ¹¹¹P. J. van Leeuwen, "Nonlinear data assimilation in geosciences: An extremely efficient particle filter," *Q. J. R. Meteorol. Soc.* **136**, 1991–1999 (2010).
- ¹¹²D. Crisan and J. Míguez, "Uniform convergence over time of a nested particle filtering scheme for recursive parameter estimation in state-space Markov models," *Adv. Appl. Probab.* **49**(4), 1170–1200 (2017).
- ¹¹³P. Del Moral and J. Tugaut, "Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions," *Stoch. Anal. Appl.* **37**(6), 909–935 (2019).
- ¹¹⁴K. Heine and D. Crisan, "Uniform approximations of discrete-time filters," *Adv. Appl. Probab.* **40**(4), 979–1001 (2008).
- ¹¹⁵F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters," *Ann. Appl. Probab.* **14**(1), 144–187 (2004).
- ¹¹⁶C. E. Leith, "Objective methods for weather prediction," *Annu. Rev. Fluid Mech.* **10**(1), 107–128 (1978).
- ¹¹⁷C. E. Leith, "Theoretical skill of Monte Carlo forecasts," *Mon. Weather Rev.* **102**(6), 409–418 (1974).
- ¹¹⁸R. Stroe and J. F. Royer, "Comparison of different error growth formulas and predictability estimation in numerical extended-range forecasts," *Ann. Geophys.* **11**(4), 296–316 (1993).
- ¹¹⁹A. J. Simmons, R. Mureau, and T. Petroliaigis, "Error growth and estimates of predictability from the ECMWF forecasting system," *Q. J. R. Meteorol. Soc.* **121**(527), 1739–1771 (1995).
- ¹²⁰A. Trevisan, P. Malguzzi, and M. Fantini, "On Lorenz's law for the growth of large and small errors in the atmosphere," *J. Atmos. Sci.* **49**(8), 713–719 (1992).
- ¹²¹P. Malguzzi, A. Trevisan, and A. Speranza, "Statistics and predictability for an intermediate dimensionality model of the baroclinic jet," *Ann. Geophys. Atmos. Hydrospheres Space Sci.* **8**(1), 29–35 (1990).
- ¹²²H. Savijärvi, "Error growth in a large numerical forecast system," *Mon. Weather Rev.* **123**(1), 212–221 (1995).
- ¹²³L. V. Kantorovich, "On the translocation of masses," *J. Math. Sci.* **133**(4), 1381–1382 (2006). originally published in *Doklady Akad. Nauk SSSR*, **37**(7–8), 199–201 (1942).
- ¹²⁴R. L. Dobrushin, "Prescribing a system of random variables by conditional distributions," *Theory Probab. Appl.* **15**(3), 458–486 (1970).
- ¹²⁵L. Wasserstein, "Markov processes with countable state space describing large systems of automata," *Probl. Peredachi Informatsii* **5**, 64–73 (1969). (in Russian).
- ¹²⁶C. Villani, *Optimal Transport: Old and New* (Springer-Verlag, Berlin, 2009).
- ¹²⁷Y. Robin, P. Yiou, and P. Naveau, "Detecting changes in forced climate attractors with Wasserstein distance," *Nonlinear Process. Geophys.* **24**, 393–405 (2017).
- ¹²⁸E. N. Lorenz, "Irregularity: A fundamental property of the atmosphere," *Tellus A* **36**(2), 98–110 (1984).
- ¹²⁹G. Vissio, V. Lembo, V. Lucarini, and M. Ghil, "Evaluating the performance of climate models based on Wasserstein distance," *Geophys. Res. Lett.* **47**(21), e2020GL089385, <https://doi.org/10.1029/2020GL089385> (2020).
- ¹³⁰D. Nualart, "The Malliavin calculus and related topics," in *Probability and its Applications*, 2nd ed. (Springer-Verlag, Berlin, 2006).
- ¹³¹H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems* [Wiley-Interscience (John Wiley & Sons), New York, 1972].