Almost sure error bounds for data assimilation in dissipative systems with unbounded observation noise

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Almost sure error bounds for data assimilation in dissipative systems with unbounded observation noise

Lea Oljača, Jochen Bröcker, Tobias Kuna*

September 7, 2018

Abstract

Data assimilation is uniquely challenging in weather forecasting due to the high dimensionality of the employed models and the nonlinearity of the governing equations. Although current operational schemes are used successfully, our understanding of their long-term error behaviour is still incomplete. In this work, we study the error of some simple data assimilation schemes in the presence of unbounded (e.g. Gaussian) noise on a wide class of dissipative dynamical systems with certain properties, including the Lorenz models and the 2D incompressible Navier-Stokes equations. We exploit the properties of the dynamics to derive analytic bounds on the long-term error for individual realisations of the noise in time. These bounds are proportional to the variance of the noise. Furthermore, we find that the error exhibits a form of stationary behaviour, and in particular an accumulation of error does not occur. This improves on previous results in which either the noise was bounded or the error was considered in expectation only.

1   Introduction

Data assimilation is a term used in the geophysical community to describe efforts to improve our knowledge of a system by combining incomplete observations with imperfect models [1]. Data assimilation is important in many fields of engineering and geophysical applications, and is an essential part of modern numerical weather prediction where it is used to initialise the forecasts based on observations of the atmosphere, combined with short term predictions [13]. In this field, data assimilation is uniquely challenging due to the infinite dimensionality and nonlinearity of the weather problem. Currently employed models

use discretizations with $\mathcal{O}(10^9)$ dimensional state vectors and $\mathcal{O}(10^7)$ partial observations of the atmosphere per day [2]. Furthermore, equations governing the dynamics of the atmosphere are well known to exhibit sensitive dependence on initial conditions [19, 13], meaning that determining them as accurately as possible is a key factor in increasing the length of the forecasting horizon.

Combining noisy data with uncertain models is an inverse problem whose optimal solution is necessarily probabilistic and sits naturally in a Bayesian framework [15] and [13], Sec. 5.5. Due to the nonlinear nature of the underlying equations, deriving an explicit form for the posterior distribution is in general not possible [24]. A sufficiently precise numerical representation (e.g. by MCMC methods or particle filters [18]) of the solution is very computationally expensive and not currently feasible in operational weather forecasting [15], although this is a promising area of research [27]. Therefore, the data assimilation schemes used in practice are approximations based on exact schemes derived for linear systems with Gaussian priors and additive Gaussian noise, known as the Kalman filter [17]. The schemes are applied to the nonlinear dynamics sequentially with various further simplifications, the simplest of which is to assume constant prior covariance. This is known as the 3DVAR method [13], Sec 5.5. A more advance method, the ensemble Kalman filter, involves an evolving prior covariance, estimated through the use of ensembles, that is, several simultaneous runs of the data assimilation cycle using a set of perturbed observations [6]. Although clearly used with great success [2], these are nonetheless ad hoc approximations, and a satisfactory understanding of their fundamental properties is still lacking.

A rigorous study of data assimilation in the context of the full primitive equations (a reasonable model of atmospheric circulation [26]) is currently out of scope. There has been extensive study of a simpler but still infinite dimensional model; the 2D viscous, incompressible Navier-Stokes (N-S) equations. Other models typically studied in the context of data assimilation in geophysical applications (see e.g. [20, 14, 16, 23]) are the Lorenz '63 and Lorenz '96 models, as they exhibit many of the properties of the N-S equations such as being dissipative with a quadratic and energy conserving nonlinearity, while having the advantage of being finite dimensional. Fortunately some remarkable properties of the 2D N-S equations have been known for some time. It was first shown by C. Foias and G. Prodi in 1967 [7] that the solution is completely determined by the temporal evolution of some finite number of spatial Fourier modes, which have since been named the "determining modes". Subsequent work [8, 12] showed that this also holds for a finite set of appropriately chosen nodal values.

More recent work re-frames these results in the context of data assimilation [21, 11], and shows that certain data assimilation schemes have zero asymptotic error even with only finitely rank observations. Hayden, Olson and Titi [11] consider the Lorenz '63 and N-S equations with a data assimilation scheme where noiseless observations are directly replaced into the approximating solution at

discrete times. Their result shows that for a sufficiently large number of observed low modes, the higher modes synchronise, that is, the error goes to zero with the number of assimilation cycles.

In [5], Brett et al build on the results in [11] by allowing for observational errors and using the 3DVAR algorithm. They show that for bounded observational errors, the asymptotic ($t \to \infty$) error between the approximating solution and the true state of the atmosphere is bounded, and of the same order of magnitude as the bound on the noise. The same result is obtained in [10] for another type of data assimilation scheme, which is related to the once widely used 'nudging' schemes. Therefore, in both papers, the overall error is driven by the error in the observations, regardless of initial error. Furthermore, this result is obtained pointwise, that is, it is true for any realization of the noise. The stochastic properties of the observational errors however, do not enter into the derivation of the bound, except the boundedness, which is essential.

In [14, 16], results are obtained in expectation for unbounded noise for the Lorenz '96 and '63 models, respectively. They show that for the 3DVAR scheme, the mean square of the error is of the same order of magnitude as the variance of the noise. In [23], Sanz-Alonso and Stuart extend this result, in expectation, to a wide class of dissipative PDEs, including infinite dimensional systems, that satisfy certain properties; the "absorbing ball" property and the "squeezing property". As is noted in [5], in a remark after Assumption 3.1, there is essentially a trade-off to be made between having bounded noise, with pointwise bounds, and unbounded noise, where similar techniques lead to results in expectation.

The main objective of the present paper is to investigate whether data assimilation into certain dissipative systems of PDEs is well behaved. Our approach is based on the works of [11], [5] and [23]. In those publications, results regarding data assimilation accuracy with unbounded noise are given in expectation, while in the present paper we derive (almost surely) pointwise bounds, even for unbounded noise. More specifically, we prove that for large time, the error is bounded by a finite and stationary process, and give an explicit description of this process in terms of the observation noise. Technically, there are realisations of the noise for which this bound fails, but these have zero probability, and hence are statistically irrelevant.

We use the simple replacement data assimilation scheme as studied by Titi et al in [11] although we expect our results to be extendible to 3DVAR type algorithms. Indeed our results extend to the framework put forward in [5]; this will be presented in a forthcoming publication. We require assumptions similar to the absorbing and squeezing properties of [23] but with some crucial differences. We allow the squeezing function to be random, and require only that its expectation is less than one. We are then able to apply Birkhoff's Ergodic Theorem to show that the squeezing function is sufficiently often less than one to give us a bound which is pointwise finite (Theorem 1 and 2). The result holds for any strength of the noise, given by the variance σ^2, and furthermore, the

bound decreases as the variance of the noise is decreased. Therefore the data assimilation error (for large time) is at least proportional to the strength of the noise. As in [23], we test our assumptions on two finite dimensional systems; Lorenz '63 and '96, before turning to the infinite dimensional N-S system.

The paper is organised as follows. In Section 2 we describe the dynamical system framework, the data assimilation scheme, and the assumptions we require on the observation error. Observations at each data assimilation time are assumed to contain a random error, the nature of which we keep as general as possible. In particular, we do not require i.i.d. or bounded noise, just that the noise is stationary and ergodic. In Section 3, we set out general assumptions on the dynamical systems needed for our main result, Theorem 1, the theorem itself and the proof. In Section 4, we investigate the properties of an apriori bound we derive for the dissipative systems considered in this paper. In Section 5 we show that our assumptions are satisfied by a large class of finite dimensional dissipative systems provided they satisfy certain properties. We discuss the Lorenz '63 and '96 models as examples of such systems. In Section 6, we prove that the N-S equations satisfy the Assumptions of 1 as well. In Section 7 we run numerical simulations for the Lorenz 63' model, showing that the bound hold asymptotically, while Section 8 has a a brief discussion of possible extensions to our results and future work.

2 The data assimilation problem

2.1 Dissipative dynamical system

Informally, we think of an equation as being "dissipative" if all solutions are eventually bounded and this bound is uniform for any initial condition. Formally, a semigroup is dissipative if it possesses a compact absorbing set [22].

Let $\mathbf{H}$ be a Hilbert space with $| \, . \, |$ the induced norm. Let U be the solution of a dissipative system with initial conditions U_0 at t_0 and let ψ be the continuous semi-flow defined by

$$U(t) = \psi(t, t_0, U_0), \tag{1}$$

where

$$\psi(t + s, t_0, U_0) = \psi(t, s, \psi(s, t_0, U_0)), \quad \text{(the semigroup property)},$$

and

$$\psi(0, t, U(t)) = U(t)$$

for all real $t \geq 0$, such that ψ is continuous in t and with respect to initial condition U_0.

We assume that this dynamical system is a perfect representation of the real world system we are interested in; for instance the atmosphere, and we refer to U as the "reference" solution.

4

2.2 Data assimilation

As mentioned in the introduction, we will be using a simple data assimilation method as defined by Titi et al in [11] but with unbounded noise added at each discrete data assimilation time.

Let $\mathbf{O}_P$, the observation *space*, be a finite dimensional subspace of $\mathbf{H}$ and P the orthogonal projection onto $\mathbf{O}_P$.

An observation at time t_n is given by $PU(t_n) + \sigma R_n$, where σR_n is the noise, or random error, in the observation. We will define R_n more precisely in Section 2.3. We assume that R_n is a random variable with values in $\mathbf{O}_P$ so that $PR_n = R_n$.

The finite dimensionality of the observation space $\mathbf{O}_P$ is not a limiting assumption as from an operational perspective it's natural to assume that we have a finite number of observations at each time step. However, the observations are restricted to a linear transformation of the model state. This is a restrictive assumption and is often not the case in weather prediction; the observation operator can be highly non-linear, as for example, in the case of satellite observations. Furthermore assuming additive noise is clearly an idealisation of representing observational uncertainty.

The approximating solution of the discrete data assimilation scheme that we use is obtained as follows. Initially at $t_0 = 0$ we have,

$$\bar{u}_0 = \eta + PU_0 + \sigma R_0,$$

where η is the initial guess of the unobserved part of the solution. Then at discrete times $0 < t_1 < t_2 < ...$ we set

$$\bar{u}_n = Q\psi(t_n, t_{n-1}, \bar{u}_{n-1}) + PU(t_n) + \sigma R_n, \tag{2}$$

where $Q = I - P$ is the projection onto unobserved space.

At intermediate times $t_n \leq t < t_{n+1}$, the approximating solution $u(t)$ is a continuous in time function defined by

$$u(t) = \psi(t, t_n, \bar{u}_n) \text{ for } t \in [t_n, t_{n+1}). \tag{3}$$

We note that u is continuous on each interval $[t_n, t_{n+1})$ but has discontinuities at $t_n, n \in \mathbb{N}$, with u continuous from the right and with limits to the left, since

$$u(t_n^+) = \lim_{t \to t_n^+} \psi(t, t_n, \bar{u}_n) = \bar{u}_n = u(t_n),$$

while

$$u(t_n^-) = \lim_{t \to t_n^-} \psi(t, t_{n-1}, \bar{u}_{n-1}) = \psi(t_n, t_{n-1}, \bar{u}_{n-1}) \neq \bar{u}_n.$$

We are interested in the data assimilation error $\delta(t)$, which is the difference between the reference and approximating solutions described above. In particular,

we are interested in the asymptotic behaviour as $t \to \infty$. Like the approximating solution, $\delta(t)$ is piece-wise continuous in time and defined by

$$\delta(t) = U(t) - u(t) = \psi(t, t_0, U_0) - \psi(t, t_n, \bar{u}_n) \tag{4}$$

in the interval $[t_n, t_{n+1})$. At t_n we have

$$\delta_n := \delta(t_n) = U(t_n) - \bar{u}_n = Q\psi(t_n, t_0, U_0) - Q\psi(t_n, t_{n-1}, \bar{u}_{n-1}) - \sigma R_n.$$

For simplicity, we assume that the time between observational updates (the data assimilation interval),

$$h = t_{n+1} - t_n > 0$$

is constant.

2.3   Observations

As we will be considering the asymptotic data assimilation error, we will be looking at a sequence of noise realisation that extends into infinite time and in fact it will be useful to extend it backward in time also.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $T : \Omega \to \Omega$ a measure preserving map such that T and T^{-1} are ergodic with respect to $\mathbb{P}$. Let $R : \Omega \to \mathbf{O}_P$ be a random variable on $(\Omega, \mathscr{F})$ and denote $R_n = R \circ T^n$; a sequence of random variables, with $n \in \mathbb{Z}$. R_n will serve to model the noise in the observations at time t_n. We let

$$\bar{R} : (\Omega, \mathscr{F}) \to (\mathbf{O}_P^\infty, \mathscr{B}_\infty)$$

be given by

$$\omega \to (..R_{-1}(\omega), R_0(\omega), R_1(\omega)...).$$

This is a measurable map and represents a realisation of the noise for all time, extending to infinite past and future. We denote the probability distribution of $\bar{R}$ by $P_{\bar{R}}$.

We note that with T measure preserving, R_n is a strictly stationary sequence (see e.g. [4], Proposition 6.9. for proof). We further assume that $\mathbb{E}(R) = 0$ and $\mathbb{E}(|R|^2) = 1$ and we model the random noise in our observation at time t_n as σR_n, where $\sigma \in \mathbb{R}^+$. Therefore σ^2 is the variance of the observation noise. If R were to have non-zero mean, this would represent a systematic error.

As an example, suppose that the R_n are i.i.d random variables with $T : \mathbf{O}_P^\infty \to \mathbf{O}_P^\infty$ being the shift map defined by $(T^k(\bar{r}))_n = r_{n+k}$ for $\bar{r} \in \mathbf{O}_P^\infty$. Then the distribution $P_{\bar{R}}$ of $\bar{R}$ is the product probability and $(\mathbf{O}_P^\infty, \mathscr{B}_\infty, P_{\bar{R}})$ is the canonical probability model[1]. It can be shown that T is measure preserving and T, T^{-1} are ergodic. The proof is similar to the Kolmogorov zero-one law [4], Theorem 3.12.

---

[1]Since the distribution of the process contains all the information we are interested in, we have discarded the original process on Ω and have represented it in term of the coordinate representation process instead on $\mathbf{O}_P^\infty$.

3 Assumptions and main result

In this section we state the main assumptions that we will need in order to prove our main result, Theorem 1. Assumption 1 requires the existence of an absorbing ball which is natural to dissipative systems. Assumption 2 can often be deduced from the same estimates that give us Assumption 1, as is demonstrated in Lemma 5, and is an a priori bound on the error dynamics. Assumptions 3 and 4 are generally more difficult to prove, particularly Assumption 4 in the presence of unbounded random error. They represent a kind of contraction or squeezing on the unobserved part of the dynamics.

Assumption 1. *(Absorbing ball property) There exists $K > 0$, depending on the dynamical system, such that the ball $\mathscr{B} = \{U; |U|^2 \leq K\}$ is absorbing and forward invariant.*

Assumption 2. *(A priori bound) For all $\sigma, h > 0$, there exists a measurable function $\rho_0 : \mathbb{R}^+ \times \mathbb{R}^+ \times \Omega \to \mathbb{R}^+$ with*

$$|\delta_n|^2 \leq \rho_0(h, \sigma) \circ T^n(\omega) := \rho_n$$

such that ρ_n is a continuous monotone increasing function of σ.

Assumption 3. *There exist continuous functions $M, \gamma : (\mathbb{R}^+, \mathbb{R}^+) \to \mathbb{R}^+$ such that whenever $U \in \mathscr{B}$ and $|U - V| \leq \rho$,*

$$|Q\{\psi(t+\tau, t, U) - \psi(t+\tau, t, V)\}|^2 \leq M(\tau, \rho)|Q(U-V)|^2 + \gamma(\tau, \rho)|P(U-V)|^2.$$

Without loss of generality we can assume that M and γ are not decreasing in ρ because we can always replace M, γ by functions that are larger and not decreasing.

Assumption 4. *With ρ_0 as in Assumption 2 and $M(\tau, \rho)$ and $\gamma(\tau, \rho)$ as in Assumption 3; for every $\sigma > 0$ there exists an $h > 0$, such that*

$$\mathbb{E}M(h, \rho_0(h, \sigma)) < 1,$$

and

$$\mathbb{E}\gamma(h, \rho_0(h, \sigma)) < \infty.$$

We note that for any measurable function $f : \mathbb{R} \to \mathbb{R}$, the process $f \circ \rho_n$ is stationary and ergodic, since T is assumed to be measure preserving and ergodic.

In particular, we can write,

$$M_n(\tau) := M(\tau, \rho_n) = M_0(\tau, \rho_0) \circ T^n(\omega), \tag{5}$$

and

$$\gamma_n(\tau) := \gamma(\tau, \rho_n) = \gamma_0(\tau, \rho_0) \circ T^n(\omega). \tag{6}$$

Remark: We want to point out that Assumptions 1 to 4 are related to 'absorbing ball' and 'squeezing' properties which are used in the literature on dissipative dynamics. These properties can take various similar forms and are at the heart of the analysis in several of the works mentioned in the introduction, e.g. [11], [5] and [23]. In particular, we can compare our assumptions to that in [23], in which the authors also treat unbounded noise and apply results to a class of general dissipative systems as we do in this paper. We note first that the absorbing ball property of [23] implies our Assumptions 1 and 2, as shown in Section 4, Corollary 1, as will be shown.

Our 'squeezing' assumption and that of [23] are not directly comparable. In [23] the squeezing applies only to a neighbourhood of the attractor. Our squeezing property, that is assumptions 3 and 4, may appear to be global (i.e. not restricted to a neighbourhood), however note that the coefficient M_n can be larger than one and hence not always contracting. Assumption 4 ensures that squeezing happens often enough by requiring that the expected value of the coefficient is less than one. Although [23] treats unbounded noise, the approximating solution or filter is in fact truncated so that it is always inside the required neighbourhood of the attractor. In our case, the noise is allowed to take the approximating solution outside of any neighbourhood of the attractor.

We now state the main result of the paper.

Theorem 1. *Suppose Assumptions 1 to 4 hold. Let $\sigma^* > 0$ and take $h > 0$ as in Assumption 4 with σ^* instead of σ. Then there exists a stationary and a.s. finite process C_n, a non-negative constant $\bar{\beta} < 1$ and a random variable D, such that for all $\sigma < \sigma^*$, the error $\delta_n = U(t_n) - u(t_n)$ satisfies*

$$|\delta_n|^2 \leq \sigma^2 C_n + D\bar{\beta}^n |QU_0 - \eta|^2, \tag{7}$$

almost surely. In particular,

$$\limsup_n \left(|\delta_n|^2 - \sigma^2 C_n \right) \leq 0, \tag{8}$$

a.s., where C_n, $\bar{\beta}$ and D are given in the proof by Equations (20), (21) and (22). In particular, $C_n, \bar{\beta}$ and D only depend on σ^*.

Theorem 1 shows that, for almost all realisations of the noise, at any data assimilation update time t_n, the error δ_n is bounded. In addition, asymptotically for large time, the bound is given by $\sigma^2 C_n$ which constitutes a stationary process so that its distribution is time independent. Furthermore as $\sigma \to 0$ the bound decreases to zero like σ^2.

To get a bound for intermediate times $t \in (t_n, t_{n+1})$, we require a further assumption.

Assumption 5. *There exists a constant $\kappa > 0$ such that $|\delta(t)|^2 \leq e^{\kappa(t-t_n)}|\delta_n|^2$ for $t \in [t_n, t_{n+1})$.*

We can easily see that if Assumption 5 holds, then the following modified version of Theorem 1 follows.

Theorem 2. *Suppose Assumptions 1 to 5 hold. Let $\sigma^* > 0$ and take $h > 0$ as in Assumption 4 with σ^* instead of σ. Then there exists a stationary and a.s. finite process C_n, a non-negative constant $\bar{\beta} < 1$ and a random variable D, such that for all $\sigma < \sigma^*$, the error $\delta(t) = U(t) - u(t)$ with $t \in [t_n, t_{n+1}) := I_n$ satisfies*

$$|\delta(t)|^2 \leq (\sigma^2 C_n + D\bar{\beta}^n |QU_0 - \eta|^2)e^{\kappa h}, \tag{9}$$

almost surely. In particular,

$$\limsup_n \left[\sup_{t \in I_n} \left(|\delta(t)|^2 - e^{\kappa h}\sigma^2 C_n \right) \right] \leq 0, \tag{10}$$

a.s., where C_n, $\bar{\beta}$ and D are given in the proof by Equations (20), (21) and (22). In particular, $C_n, \bar{\beta}$ and D only depend on σ^*.

Before turning to the proof of the main result, we require some lemmas.

Lemma 1. *Under Assumptions 1 to 3, $\delta_n = U(t_n) - u(t_n)$ satisfies*

$$|\delta_n|^2 \leq \sigma^2 \sum_{l=1}^n \prod_{k=l}^{n-1} M_k |R_{l-1}|^2 \gamma_{l-1} + \prod_{k=0}^{n-1} M_k |QU_0 - \eta|^2 + \sigma^2 |R_n|^2, \tag{11}$$

where $M_k := M(h, \rho_k(h))$ and $h = t_{n+1} - t_n$ is the update interval.

Proof. By Assumption 1 we have that the solution $U(t) \in \mathscr{B}$ for some $t > 0$. Without loss of generality we can assume that $U(t_0) \in \mathscr{B}$. Then, $U(t_n) \in \mathscr{B}$, by the forward invariance of $\mathscr{B}$. Furthermore, by Assumption 2, for any $h > 0$, we have a stationary process ρ_n such that $|\delta_n|^2 \leq \rho_n$ for all $n \in \mathbb{N}$. Therefore we can apply Assumption 3 at each update time t_n. Let $t \in [t_n, t_{n+1})$, $U = U(t_n)$, $V = u(t_n)$, and $M_n(\tau)$(respectively $\gamma_n(\tau)$) be as in Equation (5) (respectively Eq. (6)) where $\tau = t - t_n \in [0, h]$. We obtain

$$
\begin{aligned}
|Q\delta_{n+1}|^2 &= \lim_{t \to t_{n+1}} |Q\delta(t)|^2 \\
&\leq \lim_{t \to t_{n+1}} M_n(t - t_n)|Q\delta_n|^2 + \sigma^2 \gamma_n(t - t_n)|R_n|^2 \\
&= M_n(h)|Q\delta_n|^2 + \sigma^2 \gamma_n(h)|R_n|^2,
\end{aligned}
$$

where we have used the continuity of $Q\delta(t)$ at t_{n+1}. Write $M_n := M_n(h)$ and $\gamma_n := \gamma_n(h)$ for simplicity. By induction on the above,

$$|Q\delta_n|^2 \leq \sigma^2 \sum_{l=1}^n \prod_{k=l}^{n-1} M_k |R_{l-1}|^2 \gamma_{l-1} + \prod_{k=0}^{n-1} M_k |QU_0 - \eta|^2,$$

since $|Q\delta_0|^2 = |QU_0 - \eta|^2$ and we define $\prod_{k=n}^{n-1} M_k = 1$.

Finally, using that $|P\delta_n|^2 = \sigma^2 |R_n|^2$,

$$|\delta_n|^2 = |Q\delta_n|^2 + |P\delta_n|^2,$$

$$\leq \sigma^2 \sum_{l=1}^{n} \prod_{k=l}^{n-1} M_k |R_{l-1}|^2 \gamma_{l-1} + \prod_{k=0}^{n-1} M_k |QU_0 - \eta|^2 + \sigma^2 |R_n|^2,$$

as required. □

To obtain a meaningful bound as stated in Theorem 1, we need that the RHS of estimate (11) is almost surely finite in the long term. This would clearly be the case if M_k would be less than one, for all k (with some conditions on γ_n). Unfortunately, since the a priori bound is stochastic, the M_k are also stochastic and it is not, in general, possible to guarantee that $M_k < 1$ for all k, whatever the value of h. However, we are able to use the Ergodic Theorem to show that if $\mathbb{E}(M_k) < 1$, it ensures $M_k < 1$ often enough to guarantee that estimate (11) is almost surely finite. That is, for almost all realizations of the sequence $\{M_k\}_k$, the proportion of $M_k < 1$ is sufficient to ensure that the product is less than 1.

Lemma 2. *For any real $\xi > 0$, there exist almost surely finite random variables $C_{\omega,\xi}$ and $C'_{\omega,\xi}$, such that for all $N > 0$*

$$\prod_{k=0}^{N-1} M_{-k} \leq C_{\omega,\xi} (\beta + \xi)^N, \tag{12}$$

$$\prod_{k=0}^{N-1} M_k \leq C'_{\omega,\xi} (\beta + \xi)^N, \tag{13}$$

where

$$C_{\omega,\xi} := \max_N \frac{\prod_{k=0}^{N-1} M_{-k}}{(\beta + \xi)^N}, \tag{14}$$

$$C'_{\omega,\xi} := \max_N \frac{\prod_{k=0}^{N-1} M_k}{(\beta + \xi)^N}, \tag{15}$$

where $\{M_k\}$ is as in Lemma 1 and $\beta = \mathbb{E}(M_k)$.

Proof. Assuming $\log M_0(\omega)$ is measurable we can apply the Ergodic Theorem [28, 4] to T^{-1} to obtain

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \log M_{-k}(\omega) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \log M_0(\omega) \circ T^{-k}(\omega)$$

$$= \mathbb{E}(\log M_0(\omega))$$

$$\leq \log \mathbb{E}(M_0(\omega)), \tag{16}$$

where the last inequality follows from Jensen's Inequality.

10

We note that we did not require that $\log M_0(\omega)$ is integrable as we can apply the Ergodic Theorem to random variables that are either bounded below or above. In the present case, $\log M_0(h, \omega)$ could be unbounded below but we may replace it with $\bar{M}_0(h, \omega) = \max(\epsilon, M_0(h, \omega))$ for some small $\epsilon > 0$ and apply the Ergodic Theorem to $\log \bar{M}_0(h, \omega)$.

Let $\beta = \mathbb{E}(M_k)$. From Inequality (16) we have that for a.e. ω, for all $\xi > 0$, there exists $N_{\omega,\xi}$ such that for all $n \geq N_{\omega,\xi}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \log M_{-k} \leq \ln(\beta + \xi),$$

and hence

$$\prod_{k=0}^{n-1} M_{-k} \leq (\beta + \xi)^n.$$

This implies

$$\frac{\prod_{k=0}^{n-1} M_{-k}}{(\beta + \xi)^n} \leq 1. \tag{17}$$

Next, we note that for all $N > 0$ it holds that

$$\prod_{k=0}^{N-1} M_{-k} = \frac{\prod_{k=0}^{N-1} M_{-k}}{(\beta + \xi)^N}(\beta + \xi)^N$$
$$\leq C_{\omega,\xi}(\beta + \xi)^N,$$

where

$$C_{\omega,\xi} := \max_N \frac{\prod_{k=0}^{N-1} M_{-k}}{(\beta + \xi)^N}.$$

$C_{\omega,\xi}$ is finite for a.e. ω since by Inequality (17) it is less than 1 for large enough N. To get estimate (13), we repeat the proof above with $k = -k$ but using the ergodicity and $\mathbb{P}$-invariance of T. $\qquad\square$

Lemma 3. *Let $\chi_n(\omega) = \chi_0 \circ T^n(\omega)$ be a sequence of random variables and let*

$$E_{n,m} := \sum_{l=m}^{n} \Big(\prod_{k=l}^{n} M_k \Big) \chi_l$$

for $n > m$. Then

$$E_{n,0} = E_{0,-n} \circ T^n.$$

11

Proof.

$$E_{n,0}(\omega) = \sum_{l=0}^{n} \Big(\prod_{k=l}^{n} M_k\Big)\chi_l(\omega)$$

$$= \sum_{l=-n}^{0} \Big(\prod_{k=l}^{0} M_{k+n}\Big)\chi_{l+n}(\omega)$$

$$= \sum_{l=-n}^{0} \Big(\prod_{k=l}^{0} M_k \circ T^n(\omega)\Big)\Big(\chi_n \circ T^n(\omega)\Big)_l$$

$$= E_{0,-n} \circ T^n(\omega).$$

$\square$

Lemma 4. *Let $\chi_n(\omega) = \chi_0 \circ T^n(\omega)$ be non-negative random variables with finite expectation, and suppose Assumption 4 holds. Then*

$$\sum_{l=0}^{n}\prod_{k=l}^{n-1} M_k \chi_l \leq B_\xi \circ T^{n-1},$$

where

$$B_\xi = C_{\omega,\xi} \sum_{l=0}^{\infty} (\beta + \xi)^l \chi_{-l} + \chi_1 \qquad (18)$$

is an almost surely finite random variable and $C_{\omega,\xi}$ is as defined by (14).

Proof. By definition and by Lemma 3, we have that

$$\sum_{l=0}^{n-1}\prod_{k=l}^{n-1} M_k \chi_l = E_{n-1,0} = E_{0,-(n-1)} \circ T^{n-1}(\omega),$$

where

$$E_{0,-n}(\omega) = \sum_{l=-n}^{0} \Big(\prod_{k=l}^{0} M_k\Big)\chi_l = \sum_{l=-n}^{0} \Big(\prod_{k=0}^{-l} M_{-k}\Big)\chi_l.$$

Therefore,

$$\sum_{l=0}^{n}\prod_{k=l}^{n-1} M_k \chi_l = E_{n-1,0} + \chi_n = \Big(E_{0,-(n-1)} + \chi_1\Big) \circ T^{n-1}(\omega).$$

Then using estimate (12) from Lemma 2 we have

$$E_{0,-(n-1)} \leq C_{\omega,\xi} \sum_{l=-(n-1)}^{0} (\beta + \xi)^{|l|} \chi_l$$

$$\leq C_{\omega,\xi} \sum_{l=-\infty}^{0} (\beta + \xi)^{|l|} \chi_l,$$

$$= C_{\omega,\xi} \sum_{l=0}^{\infty} (\beta + \xi)^{l} \chi_{-l}.$$

Let

$$B_\xi := C_{\omega,\xi} \sum_{l=0}^{\infty} (\beta + \xi)^{l} \chi_{-l} + \chi_1,$$

then

$$\sum_{l=0}^{n} \prod_{k=l}^{n-1} M_k \chi_l \leq B_\xi \circ T^{n-1},$$

as required.

It is clear that B_ξ is measurable since χ_n are non-negative. We need to show that B_ξ is finite for a.e. ω.

Since $\beta < 1$ by Assumption 4, we can choose $\xi > 0$ such that $\beta + \xi < 1$. We know that $C_{\omega,\xi}$ is a.s. finite by Lemma 2 and χ_1 is non-negative with finite expectation. Hence, by Monotone Convergence Theorem,

$$\mathbb{E}\Big(\sum_{l=0}^{\infty} (\beta + \xi)^{l} \chi_{-l} \Big) = \sum_{l=0}^{\infty} (\beta + \xi)^{l} \mathbb{E}(\chi_{-l}) < \infty.$$

Hence

$$\sum_{l=0}^{\infty} (\beta + \xi)^{l} \chi_{-l} < \infty$$

almost surely. Therefore, B_ξ is a.s. finite as required. $\square$

For clarity, where necessary, we will use σ as a parameter in the notation for the remainder of this section.

Proof of Theorem 1. By our choice of σ^* and h, we have $\mathbb{E}M_k^* < 1$, where $M_k^* := M_k(h, \rho_k(h, \sigma^*))$.

We consider Inequality (11). By monotonicity of M_k and γ_k we can replace σ^* inside the functions so that the Inequality (11) still holds. We have

$$|\delta_n|^2 \leq \sigma^2 \sum_{l=1}^{n} \prod_{k=l}^{n-1} M_k^* |R_{l-1}|^2 \gamma_{l-1}^* + \prod_{k=0}^{n-1} M_k^* |QU_0 - \eta|^2 + \sigma^2 |R_n|^2, \qquad (19)$$

13

where $M_k^* := M(h, \rho_k(h, \sigma^*))$ and $\gamma_k^* := \gamma(h, \rho_k(h, \sigma^*))$.

We note first that the second term of Inequality (19) is bounded a.s. by (13);

$$\prod_{k=0}^{n-1} M_k^* |QU_0 - \eta|^2 \leq C_{\omega,\xi}^{'*} (\beta^* + \xi)^n |QU_0 - \eta|^2,$$

where $\beta^* = \beta(\sigma^*)$ and $C_{\omega,\xi}^{'*} = C_{\omega,\xi}'(\sigma^*)$. Fix $\xi > 0$ so that $\beta^* + \xi < 1$. Then,

$$\lim_{n \to \infty} \prod_{k=0}^{n-1} M_k^* |QU_0 - \eta|^2 \leq \lim_{n \to \infty} D\bar{\beta}^n |QU_0 - \eta|^2 = 0, \tag{20}$$

with $D = C_{\omega,\xi}^{'*}$ and $\bar{\beta} = \beta^* + \xi$.

Next, we use Lemma 4. Let

$$C_n := B_\xi^* \circ T^{n-1} + |R \circ T^n|^2, \tag{21}$$

where B_ξ^* is as defined by Equation (18) with σ replaced by σ^* and $\chi_l = |R_{l-1}|^2 \gamma_{l-1}^*$. Hence explicitly,

$$B_\xi^* = C_{\omega,\xi}^* \sum_{l=0}^{\infty} \bar{\beta}^l |R_{-l+1}|^2 \gamma_{-l+1}^* + |R_0|^2 \gamma_0^*, \tag{22}$$

with $\bar{\beta} = \beta^* + \xi$. The remaining terms of Inequality (19) are bounded by $\sigma^2 C_n$ which is a.s. finite and stationary by Lemma 4 and by our assumptions on R_n.

Therefore,

$$|\delta_n|^2 \leq \sigma^2 C_n + D\bar{\beta}^n |QU_0 - \eta|^2$$

and

$$\limsup_n \left(|\delta_n|^2 - \sigma^2 C_n \right) \leq 0,$$

by Equation (20) a.s. as required. Furthermore it holds that $\sigma^2 C_n \to 0$ as $\sigma \to 0$ since C_n does not depend on σ. $\qquad \square$

4 A priori bound for strongly dissipative systems

The next lemmas show that we can usually have a more explicit candidate for the a priori bound ρ_n, if one has an estimate of the rate of contraction to the attractor. This rate is closely related to the absorbing ball property and to our requirement that the system is dissipative. This contraction can be shown to hold for many important dynamical systems, such as Lorenz '63, '96 and the 2D, incompressible, Navier-Stokes. In fact, it is how we are able to show that

these systems have the absorbing ball property and are dissipative. We will study this in more detail in the subsequent sections.

The next lemma derives a bound on the approximating solution based on a specific rate of contraction. The bound depends on the observation noise up to time t_n, the initial guess η, initial condition $U(t_0)$ and the length of the data assimilation interval h.

Lemma 5. *Let U be as defined in Section 2.1 and suppose that there exist constants $c_1, c_2 > 0$ such that*

$$|U(t)|^2 \leq e^{-c_1(t-s)}|U(s)|^2 + c_2 \tag{23}$$

for all $0 \leq s < t$. Let $u(t)$ be the approximating solution as defined by Equation (3), then

$$|u(t_n)|^2 \leq \phi_n(h, \eta, |U(t_0)|^2) + 2\sigma^2 \sum_{k=0}^{n} e^{-c_1 kh}|R_{n-k}|^2 \tag{24}$$

for all $n \in \mathbb{N}$, where $h = t_n - t_{n-1}$ and

$$\phi_n(h, \eta, x) = |\eta|^2 + \frac{2x}{c_1 h} + 3c_2 \frac{1 - e^{-c_1 nh}}{1 - e^{-c_1 h}}.$$

Proof. By Inequality (23) and because $u_{n-1}(t)$ is a solution in the interval $[t_{n-1}, t_n)$, we have

$$|u(t_n^-)|^2 \leq e^{-c_1 h}|u(t_{n-1})|^2 + c_2. \tag{25}$$

By definition and continuity of $Qu(t)$ at t_n we have

$$|u(t_n)|^2 = |Qu(t_n^-)|^2 + |PU(t_n) + \sigma R_n|^2 \leq |u(t_n^-)|^2 + |PU(t_n) + \sigma R_n|^2. \tag{26}$$

For simplicity, let $O_n = |PU(t_n) + \sigma R_n|^2$ and substitute Inequality (25) into Inequality (26) to get;

$$|u(t_n)|^2 \leq e^{-c_1 h}|u(t_{n-1})|^2 + O_{n-1} + c_2.$$

Therefore by induction

$$|u(t_n)|^2 \leq e^{-c_1 nh}|u(t_0)|^2 + \sum_{k=0}^{n-1} e^{-c_1 kh}\left(O_{n-k} + c_2\right). \tag{27}$$

We note that

$$|PU(t_{n-k}) + \sigma R_{n-k}|^2 \leq 2|U(t_{n-k})|^2 + 2\sigma^2|R_{n-k}|^2$$
$$\leq 2e^{-c_1(n-k)h}|U(t_0)|^2 + 2c_2 + 2\sigma^2|R_{n-k}|^2,$$

15

where we have used Inequality (23) on $U(t_{n-k})$. This implies

$$\sum_{k=0}^{n-1} e^{-c_1 kh} O_{n-k} \leq \sum_{k=0}^{n-1} e^{-c_1 kh} \left(2e^{-c_1(n-k)h}|U(t_0)|^2 + 2c_2 + 2\sigma^2|R_{n-k}|^2 \right)$$

$$= 2ne^{-c_1 nh}|U(t_0)|^2 + 2c_2 \frac{1 - e^{-c_1 nh}}{1 - e^{-c_1 h}} + 2\sigma^2 \sum_{k=0}^{n-1} e^{-c_1 kh}|R_{n-k}|^2.$$

Then Inequality (27) becomes

$$|u(t_n)|^2 \leq |\eta|^2 + \frac{2|U(t_0)|^2}{c_1 h} + 3c_2 \frac{1 - e^{-c_1 nh}}{1 - e^{-c_1 h}} + 2\sigma^2 \sum_{k=0}^{n} e^{-c_1 kh}|R_{n-k}|^2.$$

where we have used that $ne^{-c_1 hn} \leq \frac{1}{c_1 h}$ for all $n \geq 0$ and $h > 0$ and $|u(t_0)| = |\eta|^2 + \sigma^2|R_0|^2$, where η is the initial guess. Thus we have shown Inequality (24). $\square$

We can readily see that Inequality (23) gives us an absorbing ball $B(0, r)$ with $r > c_2^{1/2}$ since any bounded set will eventually be inside the ball. However, we cannot deduce forward invariance. We will see that the actual contractions we encounter in the dynamical systems we study, do guarantee forward invariance and hence imply that Assumption 1 holds.

The following corollary of Lemma 5 gives the a priori bound required for Assumption 2.

Corollary 1. *Let the conditions of Lemma 5 hold and let $\delta_n = U(t_n) - u(t_n)$ be the data assimilation error and $h = t_n - t_{n-1}$ the update interval. Then there exists a stationary, a.s. finite process*

$$\rho_n = \bar{K} + F(h) + 4\sigma^2 \sum_{k=0}^{\infty} e^{-c_1 kh}|R_{n-k}|^2, \tag{28}$$

such that $|\delta_n|^2 \leq \rho_n$, for all $n \in \mathbb{N}$.

Proof. By definition of $|\delta_n|^2$, we have

$$|\delta_n|^2 \leq 2|U(t_n)|^2 + 2|u(t_n)|^2. \tag{29}$$

We insert (23),(24) into (29) to obtain

$$|\delta_n|^2 \leq 2\phi(h, \eta, |U(t_0)|^2) + 4\sigma^2 \sum_{k=0}^{n} e^{-c_1 kh}|R_{n-k}|^2 + 2e^{-c_1 hn}|U(t_0)|^2 + 2c_2.$$

The above simplifies to

$$|\delta_n|^2 \leq \bar{K} + F(h) + 4\sigma^2 \sum_{k=0}^{\infty} e^{-c_1 kh}|R_{n-k}|^2,$$

where $F(h) = \frac{6c_2}{1-e^{-c_1 h}} + \frac{4|U(t_0)|^2}{c_1 h}$ and $\bar{K} = 2\Big(|U(t_0)|^2 + c_2 + |\eta|^2\Big)$, as required.

To see that ρ_n is a measurable process, set

$$\rho_n^N := \bar{K} + F(h) + 4\sigma^2 \sum_{k=0}^{N} e^{-c_1 kh} |R_{n-k}|^2.$$

For each N, ρ_n^N is a finite sum of random variables and therefore measurable and $\{\rho_n^N\}$ is a pointwise non decreasing sequence, since we are adding non-negative terms. Therefore, $\rho_n = \sup_N \rho_n^N$, is measurable. To see that ρ_n is almost surely finite, we note that by the Monotone Convergence Theorem

$$\mathbb{E}(\rho_n) = \sup_N \mathbb{E}(\rho_n^N) = \bar{K} + F(h) + \frac{4\sigma^2}{1 - e^{-c_1 h}} < \infty \tag{30}$$

for all $h > 0$. Furthermore, ρ_n is stationary as R_n is stationary. $\qquad\square$

We can see from Equation (30) that the a priori bound behaves badly at $h = 0$ as its expectation is $\mathcal{O}(\frac{1}{h})$, for small h. In the next lemma we show that for almost all $\omega \in \Omega$, $\lim_{h\to 0} \rho_n h := D_\omega$ exists. Therefore, pointwise, for small h, $\rho_n = \mathcal{O}(\frac{1}{h})$ as well. We note also that ρ_n is decreasing if the noise level σ decreases and converges to a noise-independent constant when $\sigma \to 0$.

Lemma 6. *For ρ_n as defined by Equation (28) we have that*

1. *$\lim_{h\to 0} \mathbb{E}(\rho_n)h = C < \infty$ where $C > 0$ is a constant,*

2. *$\lim_{h\to 0} \rho_n(\omega)h = D_\omega$ for a.e. ω,*

3. *for all $h > 0$, $\rho_n(\omega)$ is monotone in σ and $\lim_{\sigma\to 0} \rho_n(\omega) = \bar{K} + F(h)$ almost surely.*

Proof. To prove item 1, note that

$$\lim_{h\to 0} \mathbb{E}(\rho_n(\omega))h = \lim_{h\to 0} \Big(\bar{K} + F(h) + 4\sigma^2 \sum_{k=0}^{\infty} e^{-c_1 kh}\Big)h$$

$$= \lim_{h\to 0} \frac{6c_2 h}{1 - e^{-c_1 h}} + \frac{4|U(t_0)|^2 h}{c_1 h} + \frac{4\sigma^2 h}{1 - e^{-c_1 h}}$$

$$= \frac{6c_2 + 4|U(t_0)|^2 + 4\sigma^2}{c_1} := C.$$

To prove item 2, it remains to check the pointwise limit of the third term in Equation (28). Using summation by parts, for any $N > 0$,

$$\sum_{k=0}^{N} e^{-c_1 kh} |R_{n-k}|^2 = e^{-Nc_1 h} \sum_{k=0}^{N} |R_{n-k}|^2 + \sum_{k=0}^{N-1} e^{-kc_1 h}(1 - e^{c_1 h}) \sum_{j=0}^{k} |R_{n-j}|^2. \tag{31}$$

17

Considering the first term of RHS of Equation (31), by ergodicity of R_n,

$$\lim_{N\to\infty} Ne^{-Nc_1 h}\frac{\sum_{k=0}^{N}|R_{n-k}|^2}{N} = \lim_{N\to\infty}\left(Ne^{-Nc_1 h}\right)\mathbb{E}(|R_{n-k}|^2) = 0,$$

for a.e. ω.

Next we consider the second term. Again from ergodicity, we have that $\lim_{k\to\infty}\frac{\sum_{j=0}^{k}|R_{n-j}|^2}{k} = 1$, since $\mathbb{E}(|R_n|^2) = 1$. Therefore, for any $\epsilon > 0$, there exists $N_{\omega,\epsilon}$ such that for all $k \geq N_{\omega,\epsilon}$, $\frac{\sum_{j=0}^{k}|R_{n-j}|^2}{k} < 1+\epsilon$. Hence for any $k > 0$,

$$\sum_{j=0}^{k}|R_{n-j}|^2 = \frac{\sum_{j=0}^{k}|R_{n-j}|^2}{k}k \leq \bar{D}_\omega k,$$

where

$$\bar{D}_\omega := \sup_k\left(\frac{\sum_{j=0}^{k}|R_{n-j}|^2}{k}\right),$$

and $\bar{D}_\omega < \infty$ since for large enough k it is smaller than $1 + \epsilon$.

Thus the second term of the RHS of Equation (31) is bounded a.s. by

$$(1 - e^{-c_1 h})\bar{D}_\omega \sum_{k=0}^{N-1} e^{-kc_1 h}k = (1 - e^{-c_1 h})\bar{D}_\omega\frac{e^{-c_1 h}}{(1 - e^{-c_1 h})^2} = \bar{D}_\omega\frac{e^{-c_1 h}}{(1 - e^{-c_1 h})}.$$

In summary, in the limit $h \to 0$,

$$\rho_n h \to \frac{6c_2}{c_1} + \frac{4|U(t_0)|^2}{c_1} + 4\sigma^2 \bar{D}_\omega := D_\omega$$

and $\rho_n = \mathcal{O}(\frac{1}{h})$ a.s. as required.

For item 3, we note that the random term of ρ_n is a.s. finite, therefore for a.e. ω, and $h > 0$, $\lim_{\sigma\to 0}\rho_n = \bar{K} + F(h)$, is a constant that does not depend on the noise. $\qquad\square$

5 Application to finite dimensional systems

In this section we derive more concrete properties, sufficient to imply the general Assumptions 1 to 5 in 3, for dissipative and finite dimensional systems of the form

$$\frac{dU}{dt} + AU + B(U,U) = f, \tag{32}$$

where solutions U and forcing f are functions in a finite dimensional vector space $\mathbf{H} = \mathbb{R}^d$, A is a linear operator and B is a symmetric, bilinear operator; consequently, the results of Theorems 1 and 2 hold. In Subsections 5.1 and 5.2 we apply our results to the Lorenz '63 and Lorenz '96 models respectively.

We assume the following properties,

Property 1.   *1. $B(U,V) = B(V,U)$ for all $U, V \in \mathbf{H}$.*

2. *$(B(U,U),U) = 0$, for all $U \in \mathbf{H}$.*

3. *$B(QU, QU) = 0$, for all $U \in \mathbf{H}$.*

4. *There exists a constant $a_1 > 0$ such that for all $U, V \in \mathbf{H}$,*
   $|(B(U,V)| \leq a_1 |U||V|$.

5. *$(AU, U) \geq |U|^2$, for all $U \in \mathbf{H}$.*

Similar properties are used in [16], [14] and [23]. Indeed, Property 1 is equivalent to *Assumption 5.1* in [23] as will be shown below. In [23], it is shown that *Assumption 5.1* implies the relevant squeezing and absorbing ball property used by those authors (*Assumption 2.1*). As we will show, Property 1 likewise implies Assumptions 1 to 5 in 3 in our paper. As we have discussed in the remark after Assumption 4 in 3, our assumptions are related to those of [23] but are not equivalent.

For the Lorenz '63 model and standard observation operator P, as specified in Subsection 5.1, Properties 1.1 to 1.4 are easily deduced, while Property 1.5 is shown in e.g. [11]. For the Lorenz '96 system and standard P, as specified in Subsection 5.2, all the properties are shown in [16].

Remark 1: Property 1.1 is not a restriction on our dynamical system (32) since only the symmetric part of B enters the dynamics anyway. Property 1.2 implies that the non-linear term does not contribute to the change in energy, analogous with the nonlinear part of the Navier-Stokes Equations. Property 1.3 effectively represents a non trivial condition on the observation operator P, ensuring a form of observability of the system. Property 1.4 is true for any bilinear operator on a finite dimensional space and hence represents no loss of generality. Property 1.5 reflects the fact that Au is considered to be a dissipative term in the dynamics.

Remark 2: From the above description of the dynamical system, it is clear there are many parallels with the N-S equations, such as dissipativity, and a nonlinearity which is quadratic and energy conserving. Furthermore, we will see in Section 6 that the N-S equations can be rewritten in a very similar form as Equation (32).

Remark 3: We note that by orthogonality of Q and following from Property 1.5 we always have that

$$(AU, PU) \geq a_2 |PU|^2 - a_3 |U|^2 \tag{33}$$

for some $a_2 > 0$ and $a_3 \geq 0$.[2]

---

[2] $(AU, PU) = (A(P+Q)U, PU) = (APU, PU) + (AQU, PU) \geq |PU|^2 - \|A\|\|U\|^2$, where we have used Property 1.5. Therefore we have $a_2 = 1$ and $a_3 = \|A\|$ but these are not necessarily the sharpest such constants.

Remark 4: We note that if Property 1.3 holds for an orthogonal projection Q then they also hold for any projection whose image is contained in the image of Q.

The next two lemmas follow directly from Property 1. For the case of Lorenz '96, the proofs are given in [14].

Lemma 7. *Properties 1.1 and 1.2 imply that*

$$(B(V,V),U) = -2(B(U,V),V)$$

holds for all $U,V \in \mathbf{H}$.

The proof is simply expanding $(B(U+V,U+V),U+V)$ and $(B(U-V,U-V),U-V)$ using Properties 1.1 and 1.2 and bilinearity of B.

Lemma 8. *Suppose that Properties 1.1, 1.2 and 1.4 are satisfied. Then Property 1.3 is equivalent to the following; there exists a constant $b > 0$ such that*

$$2|(B(U,V),V)| \leq b|PV||U||V|. \tag{34}$$

Proof. By Lemma 7, $2|(B(U,V),V)| = |(B(V,V),U)|$. Note that

$$\begin{aligned}
(B(V,V),U) &= (B(PV+QV,PV+QV),U) \\
&= 2(B(PV,QV),U) + (B(PV,PV),U),
\end{aligned}$$

where we have used Property 1.3. Therefore by Property 1.4,

$$\begin{aligned}
|(B(V,V),U)| &\leq 2a_1|PV||QV||U| + a_1|PV|^2|U| \\
&= a_1|PV||U|(2|QV| + |PV|) \\
&\leq 3a_1|PV||U||V|,
\end{aligned}$$

as required with $b = 3a_1$.

Conversely, suppose that Inequality (34) holds. Then

$$|B(QV,QV),U)| \leq b|PQV||QV||U| = 0$$

since $|PQV| = 0$. As this holds for all $U \in \mathbf{H}$ we get that $B(QV,QV) = 0$ for all $V \in \mathbf{H}$. $\qquad\square$

In the next several lemmas we show that if Property 1 holds, then ODEs of the form (32) satisfy Assumptions 1 to 5 of 3, and consequently Theorems 1 and 2 hold.

We start with showing that Properties 1.2 and 1.5 imply Assumptions 1 and 2 of 3.

Lemma 9. *Let U be the solution of a finite dimensional ODE as defined by (32) and suppose that Properties 1.2 and 1.5 are satisfied. Then Assumption 1 holds for any $K > |f|^2$ and Assumption 2 for ρ_n as given in Corollary 1 with $c_1 = 1$ and $c_2 = |f|^2$.*

Proof. The absorbing ball property is easily verified. Take the inner product of ODE (32) with U and use Property 1.2 and Property 1.5 to get

$$\frac{1}{2}\frac{d|U|^2}{dt} + |U|^2 \leq (f, U).$$

Then, by the Cauchy-Schwarz and Young's inequality we obtain

$$\frac{1}{2}\frac{d|U|^2}{dt} + |U|^2 \leq |(f, U)| \leq |f||U| \leq \frac{1}{2}|f|^2 + \frac{1}{2}|U|^2,$$

and hence,

$$\frac{d|U|^2}{dt} + |U|^2 \leq |f|^2.$$

Assumption 1 follows from using Gronwall's lemma;

$$|U(t)|^2 \leq |U(0)|^2 e^{-t} + |f|^2(1 - e^{-t}). \tag{35}$$

We see that any ball $B(0, K^{1/2})$ with $K > |f|^2$ is absorbing and forward invariant. Furthermore, Inequality (35) implies that the conditions of Corollary 1 are satisfied with $c_1 = 1$ and $c_2 = |f|^2$ and hence Assumption 2 (a priori bound) holds. □

Before proceeding to the next lemmas we derive an equation for the error $\delta = U - u$. Since the approximating solution u satisfies Equation (32) in the interval $[t_n, t_{n+1})$, we have that

$$\frac{d\delta}{dt} + A\delta + 2B(U, \delta) - B(\delta, \delta) = 0, \tag{36}$$

where we have used the bilinearity and symmetry of B to derive the above.

In the next Lemma we show that Assumption 5 holds (Eq. (37)), and we derive a bound on $|P\delta|$ (Eq. (38)) which is used in Lemma 11 to show that Assumption 3 holds. The bound on $|P\delta|$ and its proof are similar to that of the bound obtained in [23], Lemma 5.3, but with an important difference. If we were to simply replace the bound on $|\delta_0|$ (given by r'^2 in that paper) by our a priori bound ρ_n, we would have a term multiplying $|\delta|^2$ that in the limit $h \to 0$ tends to a constant (see Lemma 6). In our bound (38), the a priori bound appears in lower order, $\rho_n^{1/2}$. This means that in the limit, this term goes to zero, which, in turn, enables us to show in Lemma 12, that there is a h for which the squeezing holds in expectation, as required by Assumption 4.

Lemma 10. *Assume that Properties 1.1, 1.2, 1.4 and 1.5 hold. Let U be a solution to ODE (32) contained in the invariant set $\mathscr{B} = B(0, K^{1/2})$. Then $\delta(t) = U(t) - u(t)$ satisfies*

$$|\delta(t)|^2 \leq |\delta_n|^2 e^{\kappa(t-t_n)}, \tag{37}$$

and

$$|P\delta|^2 \leq |\delta_n|^2 (a_4 + a_5 \rho_n^{1/2})(t - t_n) + |P\delta(t_n)|^2, \qquad (38)$$

for $t \in [t_n, t_{n+1})$, $n \in \mathbb{N}_0$, $\kappa = 2(2a_1 K^{1/2} - 1)$, $a_4 = 2e^{\kappa h}(\frac{a_1^2}{a_2} K + a_3)$, $a_5 = 2a_1 e^{3\kappa h/2}$, *and* ρ_n *is as in Lemma 9.*

Outline of Proof: Proof of Eq. (37) is straightforward and similar to the proof given for the Lorenz system in [11], so we omit it for brevity.

Proof of Eq. (38); Taking inner product of the error Equation (36) with $P\delta$ and applying Inequality (33), we get

$$\frac{1}{2}\frac{d|P\delta|^2}{dt} + a_2|P\delta|^2 - a_3|\delta|^2 + 2(B(U,\delta), P\delta) - (B(\delta,\delta), P\delta) \leq 0.$$

Inequality (38) is obtained by applying Cauchy-Schwarz, Property 1.4, Inequality (37), Young's and the a priori bound, which holds by Lemma 9, to the above and then applying Gronwall's lemma.

The next lemma shows that Assumption 3 holds.

Lemma 11. *Let* $U \in \mathscr{B} = B(0, K^{1/2})$ *be a solution to ODE (32), satisfying Properties 1.1 to 1.5 with* $\delta(t)$ *as defined by Equation (4), then there exist continuous functions* $M : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ *and and* $\gamma: \mathbb{R}^+ \to \mathbb{R}^+$ *such that*

$$|\delta(t)|^2 \leq M(t - t_n, \rho_n)|\delta_n|^2 + \gamma(t - t_n)|P\delta_n|^2,$$

for $t \geq t_n$, *where*

$$M(\tau, \rho_n) = e^{-\tau}(1 + a_6 \int_0^\tau (a_4 + a_5 \rho_n^{1/2})e^s s\, ds)$$

and

$$\gamma(\tau) = a_6(1 - e^{-\tau}),$$

with $a_6 = b^2 K$.

Proof. Taking inner product of error Equation (36) with δ and using Properties 1.5 and 1.2 we get

$$\frac{1}{2}\frac{d|\delta|^2}{dt} + |\delta|^2 \leq 2|(B(U,\delta), \delta)|.$$

Note that $|U| \leq K^{1/2}$. Using Lemma 8 and then Young's, we obtain

$$\frac{1}{2}\frac{d|\delta|^2}{dt} + |\delta|^2 \leq |\delta|^2/2 + b^2 K|P\delta|^2/2,$$

and hence

$$\frac{d|\delta|^2}{dt} + |\delta|^2 \leq b^2 K|P\delta|^2. \qquad (39)$$

We use the bound (38) on $|P\delta|^2$ from Lemma 10 and replace in above inequality to obtain

$$\frac{d|\delta|^2}{dt} + |\delta|^2 \le b^2 K \Big(|\delta_n|^2 (a_4 + a_5 \rho_n^{1/2})(t - t_n) + |P\delta(t_n)|^2 \Big).$$

Multiplying by the integrating factor e^{t-t_n} and using Gronwall we get

$$|\delta|^2 \le |\delta_n|^2 M_n(t - t_n, \rho_n) + |P\delta_n|^2 \gamma(t - t_n),$$

where

$$M_n(\tau) := M(\tau, \rho_n) = e^{-\tau} \Big(1 + a_6 \int_0^\tau (a_4 + a_5 \rho_n^{1/2}) e^s s \, ds \Big)$$

and

$$\gamma(\tau) = a_6 (1 - e^{-\tau}),$$

with $a_6 = b^2 K$. Since ρ_n is continuous w.r.t. τ for all $\tau > 0$, so are M_n for a.e. ω. $\qquad \square$

We note that in this case the γ_n are all the same, non-random and finite for all $\tau \ge 0$. Therefore Assumption 4 is satisfied if the following lemma holds.

Lemma 12. *There exists $\tau^* > 0$ such that $\mathbb{E} M_n(\tau) < 1$ and $\mathbb{E} \gamma_n(\tau) < \infty$ for all $\tau \in (0, \tau^*]$.*

Proof. We wish to show that the function

$$m(\tau) = \mathbb{E} M_n(\tau) = e^{-\tau} \Big(1 + a_6 \int_0^\tau (a_4 + a_5 \mathbb{E}(\rho_n^{1/2})) e^s s \, ds \Big)$$

is less than 1 in some neighbourhood around 0. The a priori bound ρ_n, and consequently M_n, is not well defined at zero. However, we will show that $\mathbb{E}(\rho_n(\tau)^{1/2}) \tau^{1/2}$ is finite in a neighbourhood around $\tau = 0$, that is, $\mathbb{E}(\rho_n(\tau)^{1/2}) s^{1/2} < B$ for some constant $B > 0$, for all $s \le \tau$ and τ sufficiently small.

Supposing the above holds, we have that in this neighbourhood

$$m(\tau) \le e^{-\tau} \Big(1 + a_6 \int_0^\tau (a_4 s^{1/2} + a_5 B) s^{1/2} e^s \, ds \Big) := \overline{m}(\tau), \qquad (40)$$

which implies that

$$m(0) = \lim_{\tau \to 0} \mathbb{E} M_n(\tau) \le \lim_{\tau \to 0} \overline{m}(\tau) = 1.$$

Furthermore,

$$\frac{d\overline{m}(\tau)}{d\tau} = -\overline{m}(\tau) + a_6 (a_5 \tau^{1/2} + a_4 B) e^\tau \tau^{1/2},$$

and hence

$$\frac{d\overline{m}(0)}{d\tau} = -1.$$

23

Therefore, there exists a τ^* such that $\overline{m}(\tau) < 1$ for all $0 < \tau \leq \tau^*$. Hence by the bound in (40) the same is true of $m(\tau)$, for sufficiently small τ.

It remains to show that $\mathbb{E}(\rho_n^{1/2})\tau^{1/2} = \mathbb{E}((\rho_n\tau)^{1/2})$ is bounded in a neighbourhood around $\tau = 0$. Recall that

$$\rho_n = \bar{K} + \frac{6|f|^2}{1 - e^{-\tau}} + \frac{4|U(t_0)|}{\tau} + 4\sigma^2 \sum_{k=0}^{\infty} e^{-k\tau}|R_{n-k}|^2.$$

Therefore,

$$\mathbb{E}((\rho_n\tau)^{1/2}) \leq \mathbb{E}(\rho_n\tau)^{1/2}$$
$$= \left(\bar{K}\tau + \frac{6|f|^2}{1 - e^{-\tau}}\tau + 4|U(t_0)|^2 + 4\sigma^2\tau \sum_{k=0}^{\infty} e^{-k\tau}\right)^{1/2}$$
$$= \left(\bar{K}\tau + 4|U(t_0)|^2 + \frac{6|f|^2 + 4\sigma^2}{1 - e^{-\tau}}\tau\right)^{1/2}.$$

This bound is continuous at 0, and the limit is

$$\lim_{\tau \to 0} \mathbb{E}((\rho_n\tau)^{1/2}) \leq (4|U(t_0)|^2 + 6|f|^2 + 4\sigma^2)^{1/2},$$

which is finite. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Before turning to the N-S equations we will analyse two well known finite dimensional systems, known as Lorenz '63 and '96, that are commonly used as model problems for data assimilation.

5.1   Lorenz '63 model

The Lorenz '63 model consists of a system of three coupled ODEs, obtained from the N-S equations by truncation of the Fourier series to the first three modes [19, 25]. It is given by

$$\begin{cases} \dot{U}_1 = -\alpha U_1 + \alpha U_2, \\ \dot{U}_2 = -\alpha U_1 - U_2 - U_1 U_3, \\ \dot{U}_3 = -bU_3 + U_1 U_2 - b(r + \alpha), \end{cases}$$

where the parameters $b, r, \alpha \geq 0$ are real constants with standard values of $b = 10, r = 8/3, \alpha = 28$.

We can write this system in the form of ODE (32), (see e.g. [9]), where

$$A = \begin{pmatrix} \alpha & -\alpha & 0 \\ \alpha & 1 & 0 \\ 0 & 0 & b \end{pmatrix}, B(U, \bar{U}) = \begin{pmatrix} 0 \\ (U_1\bar{U}_3 + U_3\bar{U}_1)/2 \\ -(U_1\bar{U}_2 + U_2\bar{U}_1)/2 \end{pmatrix}, f = \begin{pmatrix} 0 \\ 0 \\ -b(r + \alpha) \end{pmatrix}.$$

The standard observation operator P is the projection onto the U_1 subspace. With this operator P all items of Property 1 are easily verified. Furthermore, we have

$$(B(U,V), PW) = 0$$

for all $U, V, W \in \mathbb{R}^3$, meaning that the nonlinear part of the flow is always perpendicular to the observations.

This last property is specific to Lorenz '63; it does not hold for Lorenz '96 or N - S. It means that we can have a much simplified estimate for $|P\delta|^2$, since taking inner product of the error Equation (36) and $P\delta$ and applying Inequality (33) now yields;

$$\frac{d|P\delta|^2}{dt} + 2a_2|P\delta|^2 \leq 2a_3|\delta_n|^2 e^{\kappa(t-t_n)} \leq 2a_3|\delta_n|^2 e^{\kappa h}.$$

Setting $a_7 = 2a_3 e^{\kappa h}$, the estimate (38) on $|P\delta|^2$ is simplified to

$$|P\delta(t)|^2 \leq e^{-2a_2(t-t_n)}\left(\frac{a_7}{2a_2}|\delta_n|^2(e^{2a_2(t-t_n)} - 1) + |P\delta_n|^2\right).$$

We note that the stochastic ρ_n no longer appears. We follow the proof of Lemma 11 till Equation (39) and then use the simplified bound obtained above. Thus we get,

$$|\delta|^2 \leq |\delta_n|^2 M(t - t_n) + \gamma(t - t_n)|P\delta_n|^2,$$

where

$$M(\tau) = e^{-\tau}\left(1 + a_8 \int_0^\tau e^s - e^{(-2a_2+1)s}ds\right) \tag{41}$$

and

$$\gamma(\tau) = b^2 K e^{-\tau} \int_0^\tau e^{(-2a_2+1)s}ds,$$

where $a_8 = b^2 K \frac{a_7}{2a_2}$.

We can see that in the particular case of Lorenz '63, we get a stronger result because M is deterministic and does not depend on the size of $|\delta(t_n)|$. Consequently we just need to show that the non-random function $M(\tau) < 1$ for Assumption 4 to hold. This can readily be verified as $M(0) = 1$ and $M'(\tau) = -M(\tau) + a_8(e^\tau - e^{-2a_2\tau}) + \kappa e^{-\tau} a_8 \int_0^\tau e^s - e^{(-2a_2+1)s}ds$, so that $M'(0) = -1 < 0$. Therefore, there exists a $\tau^* > 0$ such that $M(\tau) < 1$ for all $\tau < \tau^*$.

In this case the C_n of Theorem 1 have a much simpler form. Choose some $h \in (0, \tau^*)$. Let $\zeta > 0$ be some constant such that $M(h) < \zeta < 1$. We can replace M_k by the constant ζ in Equation (11) and get;

$$|\delta_n|^2 \leq \sigma^2 \sum_{l=0}^{n-1} \zeta^l \gamma |R_{n-l-1}|^2 + \zeta^n |QU_0 - \eta|^2 + \sigma^2 |R_n|^2.$$

Therefore Equation (11) and get;

$$\limsup_n \left(|\delta_n|^2 - \sigma^2 (\sum_{l=0}^{\infty} \zeta^l \gamma |R_{n-l-1}|^2 + |R_n|^2) \right) \leq 0. \qquad (42)$$

Hence, a possible form of C_n is $C_n = \sum_{l=0}^{\infty} \zeta^l \gamma |R_{n-l-1}|^2 + |R_n|^2$, which is a stationary process due to the assumptions on R_n. Furthermore,

$$\mathbb{E}(C_n) = \frac{1 - \zeta + \gamma}{1 - \zeta} < \infty. \qquad (43)$$

Therefore, since $C_n \geq 0$, it is a.s. finite.

We note also that

$$\limsup_n \mathbb{E}|\delta_n|^2 \leq \sigma^2 \mathbb{E}(C_n) = \frac{\sigma^2 (1 - \zeta + \gamma)}{1 - \zeta},$$

so that the long-term mean square of the error is proportional to the strength of the noise, since constants ζ and γ are independent of the noise and only depend on the data assimilation interval h.

The bounding process C_n gives little information in the limit $h \to 0$, because then $\zeta \to 1$. The same problem arises using 3DVAR as shown by [14], however they also give numerical results showing that the accuracy of the filter is fortunately a lot better than the theoretical bound implies. Clearly the bounds we give are not sharp since we make a number of estimates along the way. The main problem with our analysis for small h is that we are always summing the squared magnitude of the observational error. If h is small enough however, the dynamics is close to the identity, which should lead to considerable cancellations between the propagated errors. This is not taken into account in our approach.

We remark also that the above P is not the only observation projection that would allow for Theorem 1 to hold. Any such P would need to satisfy Property 1.3. That is, $B(QU, QU) = 0$, so that the image of Q is contained in the null space of B. The null space of B is given by $U_3 U_1 = 0$ and $U_1 U_2 = 0$ so that it is composed of the plane $U_1 = 0$ and the line $U_3 = U_2 = 0$. This means that Q must project either onto the (U_2, U_3)-plane or the U_1 subspace or the origin. Since $P = I - Q$, P can project either onto the (U_2, U_3)-plane or the U_1 subspace, or the whole space (i.e. P is the identity). We note that observing only the U_2 or only the U_3 subspace would not work.

5.2 Lorenz '96 model

The Lorenz '96 model [20] is given by

$$\frac{dU_i}{dt} = (U_{i+1} - U_{i-2})U_{i-1} - U_i + F,$$

26

for $i = 1...N$, $N = 3M$, for some $M \in \mathbb{N}$ with $U_{-1} = U_{N-1}, U_0 = U_N$, $U_{N+1} = U_1$ and $F = 8$.

As given in [16], in this model A is the $N \times N$ identity matrix, $f = (8, ..., 8)^T$ is an N dimensional vector, and the symmetric bilinear form is given by

$$B(U, V)_i = -\frac{1}{2}((U_{i+1} - U_{i-2})V_{i-1} + (V_{i+1} - V_{i-2})U_{i-1}).$$

The projection operator P is produced by setting every third column of the identity matrix to 0. That is,

$$P = (e_1, e_2, 0, e_4, ..., 0, e_{N-2}, e_{N-1}, 0).$$

With the above observation operator it has been shown, see [16], that Property 1 holds and that $a_2 = 1$ and $a_3 = 0$ since A is the identity matrix. Furthermore, we have that $b = 6$ and $a_1 = 2$.

In some ways the Lorenz '96 model behaves more like the 2D Navier-Stokes, in that the equation for P is not as simple; Lorenz '63 is in this sense exceptional. Thus, in the case of Lorenz '96 we cannot easily deduce an explicit form for the process C_n.

6   Application to Navier Stokes

In this section we show that the 2-D incompressible Navier-Stokes equations, with L-periodic boundary conditions, satisfy Assumptions 1 to 5 of Section 3, and therefore that Theorems 1 and 2 hold also for this model.

As we will see, the strategy for showing Assumptions 3 and 4 of 3 for the N-S equations will differ from the finite dimensional case we saw in Section 5. In the case of N-S, we are able to use only the Q part of the error equation to derive the "squeezing" property of Assumption 3. This is due to the specific form that the observation operator P_λ takes, which means that the Q equation represents the higher modes, which are dissipated quicker, the larger the λ. In the Lorenz models all modes are dissipated at the same rate so we cannot hope to adjust the operator P in order to obtain the same effect.

Following the notation of [11], let $\Omega = [0, L] \times [0, L]$. The equations for the velocity field u and pressure p are given by

$$\frac{\partial u}{\partial t} - \nu \Delta u + (u.\nabla)u + \nabla p = f, \tag{44}$$
$$\nabla \cdot u = 0,$$

where ν is the kinematic viscosity and f the time independent body forcing. Let $\mathbb{V}$ be the space of L-periodic trigonometric polynomials, with zero divergence

and zero constant term. That is,

$$\mathbb{V} = \{u : \mathbb{R}^2 \to \mathbb{R}^2 ; \text{L-periodic trig. polynomial}, \nabla.u = 0, \int_\Omega u = 0\},$$

and let H be the closure of $\mathbb{V}$ in $L^2(\Omega)$ and V the closure of $\mathbb{V}$ in Sobolev space H^1. Let $v \in \mathbb{V}$ and let $u \in V$ be a solution to Equation (44). Take the L^2 inner product of (44) with v to get

$$(\frac{\partial u}{\partial t}, v) - \nu(\Delta u, v) + (u \cdot \nabla u, v) + (\nabla p, v) = (f, v).$$

Since v is divergence-free we obtain for the pressure term

$$(\nabla p, v) = \int_\Omega \nabla p \cdot v = -\int_\Omega p \nabla \cdot v = 0,$$

where we also use that v is periodic. By density of $\mathbb{V} \in H^1$, the weak form

$$\frac{du}{dt} + \nu A u + B(u, u) = f \tag{45}$$

of the N-S equations holds for all $v \in V$. Equation (45) is an ODE in the dual space V^*, so that A and B are operators from V to V^*. If $u \in H^2$ then $(Au, v) = \int_\Omega -\Delta u \cdot v \, dx$ and $(B(u, u), v) = \int_\Omega (u \cdot \nabla u) \cdot v \, dx$.

We can express $u \in H$ by its Fourier series

$$u = \sum_{\bar{k} \in \mathscr{J}} u_{\bar{k}} e^{i\bar{k}.x},$$

where

$$\mathscr{J} = \left\{ \bar{k} = \frac{2\pi}{L}(k_1, k_2) : k_i \in \mathbb{Z}, \bar{k} \neq 0 \right\}.$$

We define norms on H, V and $H^2 \cap H$ respectively as

$$|u|^2 = L^2 \sum_{\bar{k} \in \mathscr{J}} |u_{\bar{k}}|^2,$$

$$\|u\|^2 = L^2 \sum_{\bar{k} \in \mathscr{J}} \bar{k}^2 |u_{\bar{k}}|^2,$$

and

$$|Au|^2 = L^2 \sum_{\bar{k} \in \mathscr{J}} \bar{k}^4 |u_{\bar{k}}|^2,$$

which can be shown to be equivalent to the standard norms on L^2, H^1 and H^2 on these spaces.

The key idea of the approach taken in [11], and which we follow, is that there is a natural splitting of the phase space V into a finite-dimensional sub-space and its

infinite dimensional orthogonal complement such that the orthogonal projection of the solution onto the finite dimensional subspace dominates.

We define the orthogonal projection P_λ as

$$P_\lambda u = \sum_{|\bar{k}|^2 \leq \lambda} u_{\bar{k}} e^{i\bar{k}.x},$$

where $0 < \lambda \in \mathbb{Z}$. We say that P_λ is a projection onto the low modes.

Let us state some well known properties of the system. In this setting and with initial conditions in V, the existence and uniqueness of strong solutions is shown for example in [22]. Therefore we can define a semi-flow. We will verify Assumption 1 and 2 for Equation (45) by the following Theorem which is proved in [12].

Theorem 3. *Let $u(t)$ solve the N-S Equation (45) and $u_0 \in H$, then the following estimate holds*

$$\|u(t)\|^2 \leq e^{-\nu\lambda_1(t-s)}\|u(s)\|^2 + \frac{1}{\nu}\int_s^t e^{-\nu\lambda_1(t-\tau)}|f|^2 \, d\tau \qquad (46)$$

for every $0 < s \leq t$, where λ_1 is the smallest eigenvector of A. In particular, we have

$$\limsup_{t\to\infty} \|u(t)\|^2 \leq \frac{|f|^2}{\nu^2\lambda_1} := K. \qquad (47)$$

It follows from Corollary 1 that Assumption 2 is satisfied with constant $c_1 = \nu\lambda_1$ and $c_2 = K$.

It follows from Inequality (46) that the ball $B(0, r)$ with $r > K^{1/2}$ is an absorbing set because whatever bounded set we start with there will be a time after which it will be contained in the ball. Furthermore it's straightforward to show from (46) that $B(0, r)$ is forward invariant, as required for Assumption 1.

In the case where no noise is present in the observations, the existence of a function M, as required for Assumption 3, is shown in [11], Theorem 3.9. We follow the same reasoning but with the adjustment that in our setting $P_\lambda\delta(t_n) \neq 0$, so that the induction argument used to ensure a bound on $\|\delta_n\|^2$ is in our case impossible due to the noise term in the observation that can be arbitrarily large. Hence, we replace the $R = \|\delta_0\|^2$ bound from [11], by an a priori bound from Assumption 2. We conclude that whenever there exists a $\rho > 0$ such that $\|\delta_n\|^2 < \rho$ we have for $t \in [t_n, t_{n+1})$,

$$\|Q_\lambda\delta(t)\|^2 \leq M(t - t_n, \rho)\|\delta(t_n)\|^2,$$

where

$$M(h) = e^{-\nu\lambda h}\left(1 + \int_0^h g(s, \rho)e^{\nu\lambda s} \, ds\right)$$

and

$$g(s,\rho) = C_1\lambda^{1/4}e^{\kappa s}(\rho(h,\omega)^{1/2}e^{\kappa s/2}+2K^{1/2})^2+C_2e^{\kappa s}(\rho(h,\omega)^{1/2}e^{\kappa s/2}+2K^{1/2})^{8/3},$$

and where $C_1 = 2^{-1/4}\nu^{-1}\lambda_1^{-1/4}$, $C_2 = 5^{5/3}2^{-22/3}3\nu^{-5/3}\lambda_1^{-1/3}3$. Further, K is the size of the attractor of the N-S dynamical system defined by Equation (47). Finally, $\kappa = 2^{-1/3}(5/8)^{5/3}(3/8)\nu^{-5/3}\lambda_1^{-1/3}K^{4/3}$ is the constant as in [11], Theorem 3.8.

We want to use Theorem 1 to show that this random bound is sufficient to obtain convergence. Indeed, we can show that Assumption 4 holds.

Theorem 4. *Suppose that $\mathbb{E}(|R_0|^{8/3}) < \infty$, then for all $h > 0$, there exists a $\lambda^* < \infty$ such that for all $\lambda > \lambda^*$, Assumption 4 holds. That is, $\mathbb{E}(M(h,\rho_0(h))) < 1$.*

Proof. By the previous discussion, we have that

$$\mathbb{E}(M(h,\rho_0(h))) = e^{-\nu\lambda h}\Big(1 + \int_0^h \bar{g}(s,\rho_0(h))e^{\nu\lambda s}\, ds\Big),$$

where

$$\bar{g}(s,\rho_0(h)) := C_1\lambda^{1/4}e^{\kappa s}\mathbb{E}(l(h,s)^2) + C_2e^{\kappa s}\mathbb{E}(l(h,s)^{8/3}),$$

and where $l(h,s) := \rho_0(h)^{1/2}e^{\kappa s/2} + 2K^{1/2}$.

Note that $\bar{g}(s,\rho_0(h)) \leq \bar{g}(h,\rho_0(h))$ for all $s \leq h$. Then

$$\mathbb{E}(M(h,\rho_0(h))) \leq e^{-\nu\lambda h}\Big(1 + \bar{g}(h,\rho_0(h))\int_0^h e^{\nu\lambda s}\, ds\Big)$$

$$= e^{-\nu\lambda h} + \frac{\bar{g}(h,\rho_0(h))}{\nu\lambda}\Big(1 - e^{-\nu\lambda h}\Big).$$

From the above it follows that $\mathbb{E}(M_n(h)) < 1$ if $-\nu\lambda + \bar{g}(h,\rho_0(h)) < 0$. Using the definition of $\bar{g}$, we get

$$-\nu\lambda + C_1\lambda^{1/4}e^{\kappa h}\mathbb{E}(l^2) + C_2e^{\kappa h}\mathbb{E}(l^{8/3}) < 0, \qquad (48)$$

where $l := l(h,h)$.

It is clear that Inequality (48) will hold for some sufficiently large λ if the second and third terms of (48) are finite. It is sufficient to show that $\mathbb{E}(l^{8/3})$ is finite, since then, any lower moment is finite.

---

[3]We have used the explicit value for the dimensionless constant $c = 2^{-3/2}$ that appears in [11], Theorem 3.4.

Recall that $l^{8/3} = (\rho_0(h)^{1/2}e^{\kappa h/2} + 2K^{1/2})^{8/3}$. It is sufficient to show that $\mathbb{E}(\rho_0(h)^{4/3}) < \infty$ since

$$\mathbb{E}(l^{8/3}) = \int (\rho_0(h)^{1/2}e^{\kappa h/2} + 2K^{1/2})^{8/3}d\mathbb{P}$$
$$= \|(\rho_0(h)^{1/2}e^{\kappa h/2} + 2K^{1/2})\|_{8/3}^{8/3}$$
$$\leq \left(e^{\kappa h/2}\|\rho_0(h)^{1/2}\|_{8/3} + 2K^{1/2}\right)^{8/3},$$

where in the last step we applied the Minkowski inequality.

It's clear that the right hand side of the above inequality is finite if

$$\|\rho_0(h)^{1/2}\|_{8/3} = \|\rho_0(h)\|_{4/3}^{1/2} < \infty.$$

Using the Minkowski inequality on the a priori bound we get

$$\|\rho_0(h)\|_{4/3} \leq \bar{K} + F(h) + 4\sigma^2 \sum_{k=0}^{\infty} e^{-\nu\lambda_1 kh}\|R_{-k}^2\|_{4/3},$$

where $\bar{K}$ and $F(h)$ are both deterministic and the right hand side is finite if $h > 0$ and $\|R_{-k}^2\|_{4/3} < \infty$. $\qquad\square$

The above result does not hold uniformly for small h since the bound diverges at $h = 0$.

In the previous theorem we saw that for any $h > 0$, there exists a finite λ which guarantees that $\mathbb{E}(M(h, \rho_0(h))) < 1$. We can compute an explicit expression for a possible λ from Equation (48), which is given in Lemma 13 in the Appendix.

7 Numerical results

In this section we show some numerical results to illustrate the asymptotic bound obtained in our main result (Theorem 1) applied to Lorenz '63 system as described in Section 5.1.

To obtain the optimal data assimilation step size, we minimize the function $M(h)$ given in equation (41), which is minimal at $h \approx 3 \times 10^{-5}$. At this h, M evaluates to $\approx 1 - 1.4 \times 10^{-5}$.

The forward model is numerically integrated using the function *Odeint* within the *Scipy* package in *Python*. We generate the reference solution from an initial state $[1.0, 1.0, 1.0]$ and integrate forward for $50,000$ time steps n of size h.

The observation data is generated by adding Gaussian random variable σR_n to the U_1 component of the reference solution at each integration step, where

$\sigma = 0.3$ and $\{R_n\}, n \in \mathbb{Z}$ are i.i.d. standard normal. The bound in (42) we write as $\sigma^2 C_n$ where $C_n = \gamma D_n + |R_n|^2$, $D_0 = \sum_{l=0}^{\infty} \zeta^l |R_{-l-1}|^2$ and D_n satisfies the recursion

$D_n = \zeta D_{n-1} + R_{n-1}, n \geq 1$. We note that D_0 is independent of $\{R_n, n \geq 0\}$. We approximate D_0 through

$$D_0 \approx \sum_{l=0}^{L} \zeta^l |R_{-l-1}|^2,$$

with $L = 250,000$. The expected value of the bound is ≈ 8.13 using the equation in (43). Finally, the approximating solution is computed with initial guess of $[0.0, -2.0, 3.0]$.

Figure 1 shows how the bound (blue line) and the data assimilation error (green line) evolve with time. It illustrates that the bound holds asymptotically. In addition, we can see that the DA assimilation error converges quite quickly, after approximately $10,000$ steps to below the bound. It can be seen that the asymptotic error is in fact much smaller than the bound. We believe that this is due to our very conservative estimate of the squeezing constant (ζ in the present case). Small errors seem to decay much faster than ζ^n, which is the 'worst case' decay rate.

Next, we investigate whether the DA error asymptotically behaves like $\mathcal{O}(\sigma^2)$. In fact, from Figure 1, we hypothesize that $\mathbb{E}(\frac{1}{n}\sum_{i=1}^{n}|\delta_i|^2) = \sigma^2$. Although we have no proof of this we note that the error being equal to σ is the best possible performance for this type of algorithm. The reason is that the noise of order σ is injected at each data assimilation step. Let $E_n^j = \frac{1}{n}\sum_{i=1}^{n}|\delta_i|^2$ for a particular instance j of the noise realisation. Let $X_n^j = \frac{E_n^j - \sigma^2}{\sigma^2}$, then our hypothesis is equivalent to $\mathbb{E}(X_n^j) = 0$. In order to test the DA error asymptotically, we run data assimilation with initial condition of $[1.0, 1.0, 1.0]$ for a "true" solution starting from $[1.0, 1.0, 1.0]$ as well, so that the only deviations is the noise injected by the observations. We ran 100 MC simulations for different values of σ, using the same 100 noise samples for each σ. More specifically, for 10 values of σ; $\sigma_i = 1 - \frac{i}{10}$ for $i = [0, .., 9]$ we perform a $30,000$ step ($n = 30,000$) data assimilation run for 100 instances of noise realisations ($I = 100$). We found no evidence to reject this null hypothesis for any value of σ. Our 95 % confidence interval is $[-0.001813, 0.001813]$ with no appreciable dependence on σ. Computing the expectation with the same 100 realisations for each σ_i gives very similar results with variation of three orders of magnitude lower. Figure 2 shows the results of the same experiment but with re-sampling of the 100 realisations for each σ. It demonstrates that also in this experiment the expectations remains within the confidence interval for each σ_i.
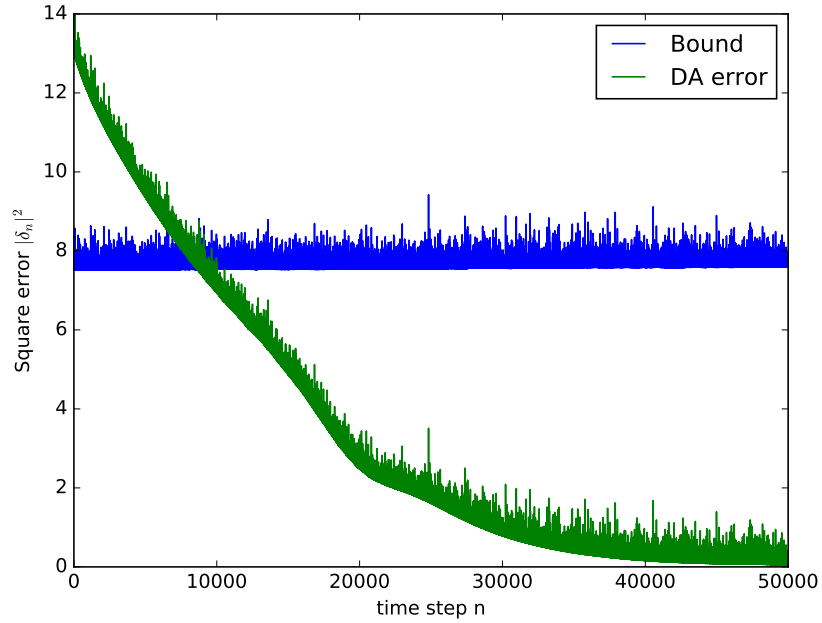
Figure 1: Comparison of the asymptotic bound as given in (42) for Lorenz '63 (blue line), versus a data assimilation error for 50,000 data assimilation steps and noise variance of 0.09 (green line). Note that the bound is asymptotic, that is, it does not take initial conditions into account.
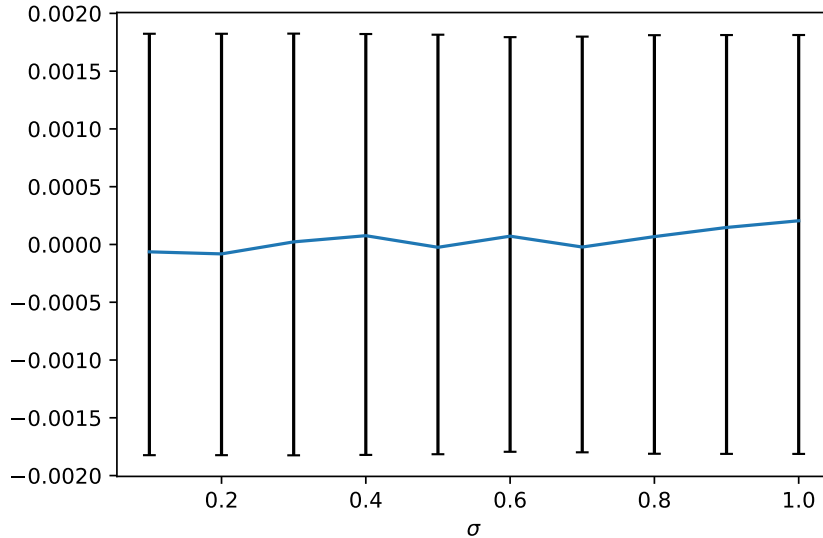
Figure 2: The expectation $\mathbb{E}(X_n^j)$, $n = 30,000$, estimated with 100 MC simulations re-sampled for each of the 10 instances of the standard deviation of the noise σ. We can see that this remains within the 95% confidence interval providing no evidence to reject the hypothesis.

8 Further work

There are several natural continuations to this work. Most immediately, our framework could be extended to other, more complex, data assimilation algorithms. As already mentioned, we expect that extension to 3DVAR algorithms (for example as set up in [5]) is possible.

We can also ask whether it is possible to apply our method to continuous in time data assimilation. Results regarding accuracy of the filter in continuous in time data assimilation already exist, see for example [3] where Bessaih, Olsen and Titi study the case of the 2D N-S model, with observations taken to be finite volume averages or at nodal points with stochastic measurement error. The authors show that this leads to a stochastically forced partial differential equation for the approximating solution and obtain asymptotic bounds on the error in expectation. Furthermore they show that the bound scales with the variance of the noise in the observations. Deriving pointwise bounds in this setting would require a suitable comparison theorem for stochastic differential equations. Comparison theorems do exist but whether they are applicable in

this setting remains to be investigated.

In this work, and in most works that we are aware of concerning the accuracy of filters, the model for the dynamics is assumed to have no error, that is, the model used for the 'true' signal and for the data assimilation is the same. However, given the chaotic nature of many models relevant for real world applications, e.g. weather, it could be natural to model the uncertainty in the signal with a stochastic forward model.

Acknowledgements

References

[1] A. Apte et al. "Data assimilation: mathematical and statistical perspectives". In: *Internat. J. Numer. Methods Fluids* 56.8 (2008), pp. 1033–1046.

[2] P. Bauer, A. Thorpe, and G. Brunet. "The quiet revolution of numerical weather prediction". In: *Nature* 525.7567 (2015), pp. 47–55.

[3] H. Bessaih, E. Olson, and E. S. Titi. "Continuous data assimilation with stochastically noisy data". In: *Nonlinearity* 28.3 (2015), pp. 729–753.

[4] L. Breiman. *Probability.* Society for Industrial and Applied Mathematics, 1992, p. 421.

[5] C. E. A. Brett et al. "Accuracy and stability of filters for dissipative PDEs". In: *Phys. D Nonlinear Phenom.* 245 (2013), pp. 34–45.

[6] G. Evensen. *Data assimilation.* Second. The ensemble Kalman filter. Springer-Verlag, Berlin, 2009.

[7] C. Foias and G. Prodi. "Sur le comportement global des solutions non-stationnaires des équations de Navier- Stokes en dimension 2". In: *Rend. del Semin. Mat. della Univ. di Padova* 39 (1967), pp. 1–34.

[8] C. Foias and R. Temam. "Determination of the solutions of the Navier-Stokes equations by a set of nodal values". In: *Math. Comput.* 43.167 (1984), pp. 117–133.

[9] C. Foias et al. "The Lorenz equation as a metaphor for the Navier-Stokes equations". In: *Discrete Contin. Dynam. Systems* 7.2 (2001), pp. 403–429.

[10] C. Foias, C. F. Mondaini, and E. S. Titi. "A discrete data assimilation scheme for the solutions of the two-dimensional Navier-Stokes equations and their statistics". In: *SIAM J. Appl. Dyn. Syst.* 15.4 (2016), pp. 2109–2142.

[11] K. Hayden, E. Olson, and E. S. Titi. "Discrete data assimilation in the Lorenz and 2D Navier-Stokes equations". In: *Phys. D Nonlinear Phenom.* 240.18 (2011), pp. 1416–1425.

[12] D. A. Jones and E. S. Titi. "On the number of determining nodes for the 2D Navier-Stokes equations". In: *J. Math. Anal. Appl.* 168.1 (1992), pp. 72–88.

[13] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 2003.

[14] K. J. H. Law, A. Shukla, and A. M. Stuart. "Analysis of the 3DVAR filter for the partially observed Lorenz'63 model". In: *Discret. Contin. Dyn. Syst. Ser. A* 34.3 (2014), pp. 1061–1078.

[15] K. J. H. Law and A. M. Stuart. "Evaluating data assimilation algorithms". In: *Internat. J. Numer. Methods Fluids* 140 (2012), pp. 3757–3782.

[16] K. J. H. Law et al. "Filter accuracy for the Lorenz 96 model: fixed versus adaptive observation operators". In: *Phys. D* 325 (2016), pp. 1–13.

[17] K. J. H. Law, A. M. Stuart, and K. Zygalakis. *Data assimilation.* Vol. 62. Texts in Applied Mathematics. A mathematical introduction. Springer, Cham, 2015.

[18] P. J. van Leeuwen. "Particle Filtering in Geophysical Systems". In: *Mon. Weather Rev.* 137.12 (2009), pp. 4089–4114.

[19] E. N. Lorenz. "Deterministic Nonperiodic Flow". In: *J. Atmos. Sci.* 20.2 (1963), pp. 130–141.

[20] E. N. Lorenz. "The Slow Manifold – What Is It?" In: *Proc. Seminar on Predictability* 1 (1996), pp. 1–18.

[21] E. Olson and E. S. Titi. "Determining modes for continuous data assimilation in 2D Turbulence". In: *J. Statist. Phys.* 113.5-6 (2003), pp. 799–840.

[22] J. C. Robinson. *Infinite-dimensional dynamical systems.* Cambridge Texts in Applied Mathematics. An introduction to dissipative parabolic PDEs and the theory of global attractors. Cambridge University Press, Cambridge, 2001.

[23] D. Sanz-Alonso and A. M. Stuart. "Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems". In: *SIAM/ASA J. Uncertain. Quantif.* 3.1 (2015), pp. 1200–1220.

[24] A. M. Stuart. "Inverse problems: a Bayesian perspective". In: *Acta Numer.* 19 (2010), pp. 451–559.

[25] R. Temam. *Infinite-dimensional dynamical systems in mechanics and physics.* Second. Vol. 68. Applied Mathematical Sciences. Springer-Verlag, New York, 1997.

[26] G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation.* Cambridge University Press, 2006.

[27] P. J. Van Leeuwen. "Aspects of particle filtering in high-dimensional spaces". In: *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).* Vol. 8964. 2015, pp. 251–262.

[28] P. Walters. *An introduction to ergodic theory.* Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1982.

9 Appendix

Lemma 13. *Equation* (48) *holds for all*

$$\lambda \geq \max\left(2^{-1}e^{4/3\kappa h}\mathbb{E}(l^2)^{4/3}, 5^{5/3}2^{-19/3}3e^{\kappa h}\mathbb{E}(l^{8/3})\right)\lambda_1^{-1/3}\nu^{-8/3} \tag{49}$$

Proof. We consider two possible cases of the second term of Inequality (48) being greater or smaller than the third term, which correspond to λ being greater or smaller than the expression

$$\left(\frac{\mathbb{E}(l^{8/3})}{\mathbb{E}(l^2)}\right)^4 (5^{5/3}2^{-19/12}3)^4\lambda_1^{-1/3}\nu^{-8/3} := M_1. \tag{50}$$

Replacing in Inequality (48), we have that for λ greater than or equal to (50), if λ holds for below equation then it holds for (48) as well;

$$-\nu\lambda + 2^{-3/4}\nu^{-1}\lambda_1^{-1/4}\lambda^{1/4}e^{\kappa h}\mathbb{E}(l^2) < 0,$$

so that

$$\lambda > 2^{-1}\nu^{-8/3}\lambda_1^{-1/3}e^{4/3\kappa h}\mathbb{E}(l^2)^{4/3} := M_2,$$

and hence

$$\lambda > \max\left(M_1, M_2\right). \tag{51}$$

On the other hand, if λ is less than expression (50), we can replace Inequality (48) with

$$-\nu\lambda + 5^{5/3}2^{-19/3}3\nu^{-5/3}\lambda_1^{-1/3}e^{\kappa h}\mathbb{E}(l^{8/3}) < 0,$$

so that

$$\lambda > 5^{5/3}2^{-19/3}3\lambda_1^{-1/3}\nu^{-8/3}e^{\kappa h}\mathbb{E}(l^{8/3}) := M_3,$$

and hence

$$M_3 < \lambda < M_1. \tag{52}$$

There are solutions for λ in Inequality (52) if and only if

$$e^{\kappa h} < \frac{\mathbb{E}(l^{8/3})^3}{\mathbb{E}(l^2)^4}5^5 2^{-16}3^3,$$

so that

$$e^{4/3\kappa h} < \frac{\mathbb{E}(l^{8/3})^4}{\mathbb{E}(l^2)^{16/3}}(5^5 2^{-16}3^3)^{4/3}.$$

Multiplying both sides by $2^{-1}\nu^{-8/3}\lambda_1^{-1/3}\mathbb{E}(l^2)^{4/3}$ we get precisely that

$$M_2 < M_1.$$

Conversely, when $M3 > M_1$, we have that $M_2 > M_1$, which means that Inequality (51) becomes

$$\lambda > M_2. \tag{53}$$

Putting Inequalities (52) and (53) together, we see that we require that

$$\lambda > \max\left(M_2, M_3\right).$$

$\square$