~~Quantifying the minimum ensemble size for asymptotic~~ Noise-scaled accuracy of the ensemble Kalman filter ~~using the degrees of instability~~with an instability-based minimum ensemble size

Kota Takeda[1,2] and Takemasa Miyoshi[2,3]

[1]Department of Applied Physics, Graduate School of Engineering, Nagoya University, Nagoya, Japan
[2]RIKEN Center for Computational Science, Kobe, Japan
[3]RIKEN Center for Interdisciplinary Theoretical and Mathematical Sciences, Wako, Japan

Correspondence: Kota Takeda (takeda@na.nuap.nagoya-u.ac.jp)

Abstract. The ensemble Kalman filter (EnKF) is widely used for state estimation in chaotic dynamical systems, including ~~the atmosphere and ocean. However, the required ensemble size for accurate state estimation remains unclear~~atmospheric and oceanic flows. One of the fundamental questions is how many samples are required for accurate long-term performance of the EnKF. In this study, we ~~define~~ introduce a notion of time-asymptotic filter accuracy based on ~~its time-asymptotic performance relative~~ the scaling of the analysis error with respect to the observation noise ~~. We then~~ level. This formulation provides a qualitative distinction between convergent and divergent filtering behavior, beyond standard criteria based on time-averaged RMSE at a fixed noise level. We investigate the minimum ensemble size ~~, m^* , required to achieve this accuracy, linking it to the degrees of instability in the chaotic dynamics. Since the well-defined characteristic numbers~~ required for this filter accuracy and relate it to intrinsic instability of dynamical systems~~called~~ . Using the Lyapunov exponents (LEs)~~quantify the time-asymptotic exponential growth or decay~~ , which quantify asymptotic exponential growth rates of infinitesimal perturbations, we ~~define the~~ characterize degrees of instability ~~N_+~~ by the number of positive ~~LEs. In the EnKF, capturing such instabilities with~~ exponents N_+. Because spanning the unstable directions by a limited ensemble is ~~crucial for achieving~~ essential for long-term ~~filter accuracy. Therefore~~accuracy, we propose an ensemble spin-up and downsizing ~~method within data assimilation cycles~~strategy. Numerical experiments ~~applying the EnKF~~ with the EnKF applied to the Lorenz 96 model ~~show~~ indicate that the minimum ensemble size required for filter accuracy ~~is estimated by~~ satisfies $m^* = N_+ + 1$. ~~This study provides a practical estimate for the minimum ensemble size~~ These results provide a practical guideline for ensemble-size selection based on a priori ~~information about the target dynamics, along with a methodto achieve long-term accuracy~~dynamical information and bridge idealized theoretical requirements with feasible numerical implementations via the ensemble downsizing method.

1 Introduction

Many geophysical systems, including the motions of the atmosphere and ocean, are modeled as dissipative dynamical systems whose trajectories converge to compact attractors. These dynamics often exhibit chaotic behavior, characterized by sensitivity to initial conditions, which renders long-term forecasts unreliable (Kalnay, 2002). Therefore, quantifying the degree of instability

in chaotic dynamics is essential. One approach to characterizing such instability is through tangent-linear approximations of dynamical systems, known as Lyapunov analysis. The degree of instability is quantified by the Lyapunov exponents (LEs), which ~~are defined as the exponential growth or decay rates of infinitesimal perturbations in the tangent space . For~~ characterize the asymptotic exponential rates of separation of nearby trajectories. These rates are defined through the evolution of infinitesimal perturbations governed by the linearized dynamics along a reference trajectory (Eckmann and Ruelle, 1985; Legras and Vautard, 1996). For autonomous continuous-time dynamical systems, such as autonomous ordinary differential equations, at least one of the LEs is zero, corresponding to perturbations parallel to the vector field (Haken, 1983). At each point on the attractor, the tangent space is decomposed into unstable, neutral, and stable subspaces, spanned by basis vectors with positive, zero, and negative exponential rates in the infinite time limit. We focus on the dimensions of these subspaces. The numbers of positive and non-negative LEs ~~,~~ are denoted by N_+ and N_0, respectively~~, represent the degrees of freedom of unstable and unstable-neutral perturbations in the tangent space, respectively~~. By definition, it follows that $N_0 \geq N_+$.

We consider Bayesian data assimilation for state estimation in chaotic dynamical systems, where noisy observations are obtained at discrete time steps. The ensemble Kalman filter ~~(EnKF)~~ (EnKF, Evensen, 2009) is widely used for this purpose. It estimates uncertainty in the forecast using an ensemble of state vectors and updates the mean and covariance via Bayes' rule. We focus on a deterministic version, the ensemble transform Kalman filter ~~(ETKF)~~ (ETKF, Bishop et al., 2001). The ensemble covariance, ~~C_t~~$\boldsymbol{C}_t$, characterizes forecast uncertainty through its eigenpairs, with the eigenvalues quantifying the magnitude of variability and the eigenvectors specifying the principal directions along which this variability occurs. In the analysis step, corrections are applied more strongly in directions with higher forecast uncertainty. In general, the rank of the ensemble covariance ~~C_t~~ $\boldsymbol{C}_t$ is less than the ensemble size m, i.e., ~~$\mathrm{rank}(C_t) \leq m-1$~~$\mathrm{rank}(\boldsymbol{C}_t) \leq m-1$. Moreover, in geophysical applications, m is limited because each ensemble member incurs a high computational cost. Therefore, it is crucial to estimate uncertain directions with a limited ensemble. If the ETKF underestimates an unstable direction, the state estimation error is not sufficiently corrected and grows to the size of the attractor. This phenomenon is ~~called~~ known as filter divergence and must be avoided. To mitigate this problem, covariance inflation ~~techniques artificially increase~~ artificially increases the ensemble spread to compensate for ~~the underestimation of~~ underestimated uncertainty, thereby helping to prevent filter divergence. Note that inflation cannot remedy an insufficient ensemble size that fails to span the unstable directions. Another numerical technique is localization (Hamill et al., 2001; Hunt et al., 2007), which reduces the influence of distant observations by damping spurious long-range correlations in the ensemble covariance. Although localization can reduce the required ensemble size for practical applications, we do not consider it in order to isolate and analyze the relationship between ensemble size and the degrees of instability.

~~Mathematical studies~~ Mathematical studies of filtering algorithms often focus on the long-term ~~behavior of the analysis error. The key~~ accuracy (Kelly et al., 2014; Kelly and Stuart, 2019; Takeda and Sakajo, 2024; Biswas and Branicki, 2024; Sanz-Alonso and Waniorek, 2025). The central objective is to ~~demonstrate that the mean squared error~~ show that the squared error SE_t remains of order r^2 when the observation noise level r is sufficiently small compared with the attractor size~~. This~~

~~property is referred to as~~ , namely,

$$\limsup_{t\to\infty} \mathbb{E}[\mathrm{SE}_t] = O(r^2) \tag{1}$$

for sufficiently small r, where the expectation is taken with respect to the probability distributions of the observation noise and the initial ensemble. We refer to this property as (r-asymptotic) filter accuracy. ~~Establishing filter accuracy ensures that filter divergence does not occur.~~ When we evaluate filter performance, compared with the standard RMSE-based criterion at a fixed noise level r, the present formulation based on Eq. (1) has the following notable features:

(i) By Jensen's inequality, the expectation of the SE dominates the squared expectation of the RMSE (up to a factor of the state dimension N_x), that is, $\mathbb{E}[\mathrm{SE}_t] \geq N_x(\mathbb{E}[\mathrm{RMSE}_t])^2$, which provides a stronger guarantee for filter performance.

(ii) The observation noise level r is treated as an asymptotic parameter, which facilitates rigorous mathematical analysis and enables a qualitative distinction between convergent and divergent filtering behavior based on the scaling with respect to r.

Takeda and Sakajo (2024) analyzed the ETKF for dissipative dynamical systems and proved filter accuracy under the large-ensemble condition $m \geq N_x + 1$, provided that sufficient inflation is applied. ~~Because the required ensemble size , $N_x + 1$, is impractical for most~~ Since such an ensemble size is impractical in high-dimensional applications, ~~it is important to identify a more relaxed lower bound $m \geq m^*$ depending on the system.~~ a central question is whether filter accuracy can be achieved with a substantially smaller ensemble size determined by the intrinsic instability of the underlying dynamics.

Under more idealized assumptions, González-Tokman and Hunt (2013) investigated ~~the lower bound of~~ a lower bound on m for the ETKF. They proved that for discrete-time dynamical systems, if ~~$m \geq N_+$~~ $m \geq N_+ + 1$, the analysis error is bounded by the order of the observation noise. The proof relies on the following assumptions: the noise is sufficiently small; the initial ensemble is close to the true state and concentrated on the unstable subspace. Their analysis has two limitations: ~~(i)~~ it applies only to discrete-time systems without zero LEs; ~~and (ii)~~ the assumptions on the initial ensemble are not practically verifiable. ~~Nevertheless, their study suggests that the minimum ensemble size is $m^* = N_+ + 1$. Related studies~~ Seminal studies (Trevisan and Uboldi, 2004; Trevisan et al., 2010; Trevisan and Palatella, 2011) investigated the use of the unstable subspace in data assimilation, known as Assimilation in Unstable Subspace (AUS). Related studies (Gurumoorthy et al., 2017; Bocquet et al., 2017; Bocquet and Carrassi, 2017; Grudzien et al., 2018a, b) analyzed the behavior of the (ensemble) Kalman filters and smoothers in relation to the unstable subspaces. These works mainly consider systems ~~with zero LEs~~ that include neutral directions and argue that correcting the state in the N_0-dimensional unstable-neutral subspace is crucial for filter performance. We ~~review~~ now focus on results directly related to the ETKF. Theoretical analyses for linear systems ~~in suggest that the rank of the ETKF covariance is asymptotically bounded by N_+ due to the exponential decay of uncertainty in the stable subspaceand the slower decay in the neutral subspace under some conditions~~have established that the forecast and analysis error covariance matrices of the EnKF asymptotically concentrate in the unstable-neutral subspace. In particular, Bocquet et al. (2017); Bocquet and Carrassi (2017) proved rigorously that, in the linear case, the error covariance matrix becomes asymptotically rank-deficient

with rank at most N_0, and that an ensemble size of at least $N_0 + 1$ is required to properly represent the error covariance matrix. They also present numerical results ~~for the ETKF applied to~~ by the ETKF for the Lorenz 96 model (Lorenz, 1996) with ~~40 variables over a finite-time interval. In this experiment, uncertainty in the neutral direction does not sufficiently decay without assimilating observations. As a result, the rank of the covariance should be larger than N_0 (i.e., $m \geq N_0 + 1$ is required) so that the~~ 40 variables. For instance, in Fig. 8 of Bocquet and Carrassi (2017), the time-averaged ~~analysis error remains small within the finite-time interval~~RMSE of the ETKF is shown as a function of the ensemble size m for the observation noise covariance matrix $R = I$. This figure shows that the RMSE is small relative to the observation noise level (~ 1) when $m \geq N_0 + 1$, supporting the theoretical findings in Bocquet et al. (2017). Similar findings are reported ~~in~~ in Bocquet (2011) for the same model and Carrassi et al. (2022) ~~using~~ for the Quasi-Geostrophic model (Reinhold and Pierrehumbert, 1982) and the Modular Arbitrary-Order Ocean-Atmosphere Model (De Cruz et al., 2016). ~~Based on these results, the minimum ensemble size is estimated as $m^* = N_0 + 1$ when focusing on the time-averaged analysis error within a finite-time interval. However, these results do not address the time-asymptotic accuracy of the ETKF.~~ In these studies, the required ensemble size for accuracy can vary depending on the observation noise level, typically fixed to $r = 1.0$.

~~In this study, we focus on the asymptotic accuracy of the ETKF relative to the order of the observation noise . In this setting, uncertainty in the neutral direction will decay in the~~ The objective of this study is to clarify the relationship between the minimum ensemble size required for filter accuracy and the degrees of dynamical instability, under idealized conditions on the inflation factor and the observation noise level. Accordingly, we adopt a definition of filter accuracy based on its asymptotic behavior with respect to the observation noise level, as given in Eq. (1). Motivated by the studies reviewed above, and under the working assumption that the influence of neutral directions on filter accuracy becomes negligible in the joint long-time ~~limit. Hence, we~~ and small-noise limits, we revisit the conjecture that the minimum ensemble size for ~~asymptotic accuracy of the ETKF is~~ the r-asymptotic filter accuracy is given by

$$m^* = N_+ + 1, \tag{2}$$

~~where only the unstable directions are tracked by the forecast ensemble covariance.~~ In the numerical experiments presented in this study, we use dynamical systems with a single zero LE~~(i.e., , so that $N_0 = N_+ + 1$). In general~~. More generally, it holds that $N_+ + 1 \leq N_0$ for continuous-time dynamical systems. ~~For this case~~Even in such cases, we still conjecture that the minimum ensemble size is $m^* = N_+ + 1$. Therefore, we use N_+ rather than N_0 to represent the minimum ensemble size throughout the paper. To ~~verify our conjecture~~ examine this conjecture within the formulation of the r-asymptotic filter accuracy, we conduct numerical experiments with the ETKF applied to the Lorenz 96 model with ~~40~~ 40 variables. We estimate the minimum ensemble size m^* so that the asymptotic analysis error is bounded by the order of the observation noise when an appropriate multiplicative inflation factor is chosen. We then compare this value with the dimension of the unstable subspace N_+, computed via Lyapunov analysis. In our experiments, we also introduce an ensemble downsizing method for the ETKF: we begin with a sufficiently large ensemble size, $m = N_x + 1$, and reduce it to a smaller size after a fixed spin-up time. This procedure is ~~designed to generate a small but accurate~~ intended to produce a small yet efficient ensemble, with its mean close to the true state and its perturbations aligned with the unstable subspace~~. Although our target model~~, thus approximately realizing the

idealized initial conditions assumed in González-Tokman and Hunt (2013). Although the target model considered is the same as that in Bocquet and Carrassi (2017), our objective differs in that we focus on ~~asymptotic accuracy and its dependence on the order of the observation noise.Numerical studies~~ the reformulated minimum ensemble size based on the accuracy criterion as Eq. (1). Numerical investigations from this perspective are important because they define and qualitatively clarify the ensemble size below which filter divergence of the ETKF cannot be avoided, even when accurate observations and appropriate inflation are used. Moreover, ~~the~~ establishing an error bound by order of the observation noise enables further mathematical analysis of the ETKF. Our approach is applicable when the LEs of the target dynamical system can be estimated, and it offers practical guidance for selecting ensemble size in high-dimensional ETKF applications.

Our main contributions are summarized as follows:

(i) By adopting the reformulated r-asymptotic filter accuracy, we provide a qualitative criterion that distinguishes between divergent and accurate filtering behavior based on noise scaling, rather than relying on time-averaged RMSE at a fixed noise level.

(ii) We propose an ensemble spin-up and downsizing method that enables a practical realization of ensembles aligned with the unstable subspace, bridging idealized theoretical assumptions and feasible numerical implementations.

(iii) Through numerical experiments with the ETKF applied to the Lorenz 96 model, we provide evidence that the minimum ensemble size required for r-asymptotic filter accuracy scales with the number of unstable directions as $m^* = N_+ + 1$, rather than $N_0 + 1$.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the basics of Lyapunov analysis. In Sect. 3, we define the ETKF with the ensemble downsizing method. In Sect. 4, we present the numerical results with the Lorenz 96 model, combining Lyapunov analysis and the ETKF. In Sect. 5, we summarize our results and outline future directions. ~~We also discuss the consistency between our findings and those of in estimating the minimum ensemble size, highlighting differences in objectives.~~

2 Characterizing the degrees of instability in dynamics

2.1 The Lyapunov exponents and their computation

Let $N_x \in \mathbb{N}$~~, we~~. We consider the dynamics governed by

$$\frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \quad t > 0 \tag{3}$$

with $\boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^{N_x}$, where ~~$f : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$~~ $\boldsymbol{f} : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ is a smooth vector field. To study the instability of the dynamics, we examine the evolution of a perturbation $\delta\boldsymbol{x}(t) \in \mathbb{R}^{N_x}$, defined as the difference between two trajectories separated by $\delta\boldsymbol{x}_0 \in \mathbb{R}^{N_x}$ at $t = 0$. Assuming that $\delta\boldsymbol{x}(t)$ is sufficiently small and smooth, its evolution is approximated by the linearization

of Eq. (3):

$$\frac{d}{dt}\delta\boldsymbol{x}(t) = J\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}(t))\,\delta\boldsymbol{x}(t), \quad t > 0 \tag{4}$$

with $\delta\boldsymbol{x}(0) = \delta\boldsymbol{x}_0$, where ~~$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}(t))$~~ $\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}(t))$ denotes the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x} = \boldsymbol{x}(t)$. ~~Eq.~~Equation (4) is referred to as the tangent linear model. Let ~~$\Phi(t, \boldsymbol{x}_0) \in \mathbb{R}^{N_x \times N_x}$~~ $\boldsymbol{\Phi}(t, \boldsymbol{x}_0) \in \mathbb{R}^{N_x \times N_x}$ denote the fundamental matrix solution to Eq. (4) with ~~$\Phi(0, \boldsymbol{x}_0) = I_{N_x}$~~ $\boldsymbol{\Phi}(0, \boldsymbol{x}_0) = \boldsymbol{I}_{N_x}$. The unique solution is then

$$\delta\boldsymbol{x}(t) = \Phi\boldsymbol{\Phi}(t, \boldsymbol{x}_0)\,\delta\boldsymbol{x}_0, \quad t \geq 0. \tag{5}$$

The matrix ~~$\Phi(t, \boldsymbol{x}_0)$~~ $\boldsymbol{\Phi}(t, \boldsymbol{x}_0)$ encodes the deformation and amplification of infinitesimally small perturbations. For the singular values ~~$\sigma_1(\Phi(t, \boldsymbol{x}_0)) \geq \sigma_2(\Phi(t, \boldsymbol{x}_0)) \geq \cdots \geq \sigma_{N_x}(\Phi(t, \boldsymbol{x}_0)) > 0$~~ ~~of~~ ~~$\Phi(t, \boldsymbol{x}_0)$~~ $\sigma_1(\boldsymbol{\Phi}(t, \boldsymbol{x}_0)) \geq \sigma_2(\boldsymbol{\Phi}(t, \boldsymbol{x}_0)) \geq \cdots \geq \sigma_{N_x}(\boldsymbol{\Phi}(t, \boldsymbol{x}_0)) > 0$ of $\boldsymbol{\Phi}(t, \boldsymbol{x}_0)$, we define

$$\lambda_j(t, \boldsymbol{x}_0) = \frac{1}{t}\log\sigma_j(\Phi\boldsymbol{\Phi}(t, \boldsymbol{x}_0)) \in \mathbb{R}, \quad j = 1, \ldots, N_x. \tag{6}$$

For each j, the singular vector $\boldsymbol{v}_j \in \mathbb{R}^{N_x}$ associated with σ_j exponentially grows or decays at rate λ_j over $[0, t]$ under the tangent linear model, i.e.,

$$\|_j(t)\| = \|\Phi(t\boldsymbol{\Phi}(t, \boldsymbol{x}_0)\,\boldsymbol{v}_j\| = e^{\lambda_j t}\|\boldsymbol{v}_j\|, \tag{7}$$

where $\|\cdot\|$ is the Euclidean norm. Thus, the deformation of the initial perturbation $\delta\boldsymbol{x}_0$ is expressed as exponential growth/decay along the directions $\boldsymbol{v}_j$. Taking the limit $t \to \infty$, we obtain the asymptotic rates

$$\lambda_j(\boldsymbol{x}_0) = \lim_{t\to\infty}\lambda_j(t, \boldsymbol{x}_0) \in \mathbb{R}, \quad j = 1, \ldots, N_x, \tag{8}$$

known as the Lyapunov exponents (LEs). The existence of these limits is guaranteed by Oseledets' Multiplicative Ergodic Theorem (Oseledets, 1968; Barreira and Pesin, 2002). If the dynamics Eq. (3) is ergodic, the LEs are uniquely determined regardless of the choice of $\boldsymbol{x}_0$ ~~in~~ belonging to an invariant subset of $\mathbb{R}^{N_x}$. For autonomous continuous-time ~~dynamics~~ dynamical systems of the form Eq. (3), at least one exponent is ~~always zero~~zero, $\lambda_j = 0$, corresponding to a perturbation parallel to the vector field $\delta\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t))$ (Haken, 1983). If the dynamics admits a positive exponent $\lambda_1 > 0$, there exists at least one unstable direction in which perturbations grow exponentially, i.e., the dynamics is chaotic. According to Sect. 1, we define the following dimension to quantify the degrees of freedom of unstable perturbations:

$$N_+ = \#\{j \in \{1, \ldots, N_x\} \mid \lambda_j > 0\}. \tag{9}$$

See Legras and Vautard (1996) for a more comprehensive introduction to LEs and their associated vectors.

We estimate the LEs numerically using the standard algorithm based on QR decomposition, as detailed in Algorithm 1 (Sandri, 1996; von Bremen et al., 1997). To implement this algorithm, we require a vector field $\boldsymbol{f} : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$, its Jacobian ~~$\boldsymbol{J}_{\boldsymbol{f}} : \mathbb{R}^{N_x \times N_x}$~~ $\boldsymbol{J}_{\boldsymbol{f}} \in \mathbb{R}^{N_x \times N_x}$, an initial state $\boldsymbol{x}_0 \in \mathbb{R}^{N_x}$, an ODE integrator IntegrateODE, a time step size $\Delta t > 0$ and a number of iterations $n \in \mathbb{N}$. For the ODE integrator, we use the fourth-order Runge-Kutta method.

Algorithm 1 Computing LEs using QR decomposition (Sandri, 1996; von Bremen et al., 1997)

Require: $\boldsymbol{f}$, ~~$J_{\boldsymbol{f}}$~~ $\boldsymbol{J}_{\boldsymbol{f}}$, $\boldsymbol{x}_0$, IntegrateODE, Δt, n

Ensure: ~~$S = (\boldsymbol{x}, V), F(S) = (\boldsymbol{f}(\boldsymbol{x}), J_{\boldsymbol{f}}(\boldsymbol{x})V)$~~ $\boldsymbol{S} = (\boldsymbol{x}, \boldsymbol{V}), \boldsymbol{F}(\boldsymbol{S}) = (\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x})\boldsymbol{V})$

1: ~~$S \leftarrow (\boldsymbol{x}_0, I_{N_x})$~~ $\boldsymbol{S} \leftarrow (\boldsymbol{x}_0, \boldsymbol{I}_{N_x})$

2: $LE \leftarrow \mathbf{0} \in \mathbb{R}^{N_x}$

3: **for** $i = 1$ **to** n **do**

4: ~~$S \leftarrow \mathrm{IntegrateODE}(F, S, \Delta t)$~~ $\boldsymbol{S} \leftarrow \mathrm{IntegrateODE}(\boldsymbol{F}, \boldsymbol{S}, \Delta t)$

5: ~~$Q, R \leftarrow \mathrm{QR}(S)$~~ $\boldsymbol{Q}, \boldsymbol{R} \leftarrow \mathrm{QR}(\boldsymbol{S}[2])$

6: ~~$S \leftarrow Q$~~ $\boldsymbol{S}[2] \leftarrow \boldsymbol{Q}$

7: ~~$LE \leftarrow LE + \log(diag(|R|))$~~ $LE \leftarrow LE + \log(|\mathrm{diag}(\boldsymbol{R})|)$

8: **end for**

9: **return** $LE/(n\Delta t)$

2.2 The Lorenz 96 model

For ~~a~~ $N_x \in \mathbb{N}$ number of variables~~$N_x \in \mathbb{N}$~~, external forcing $F \in \mathbb{R}$ and the state ~~variable~~ vector $\boldsymbol{x} = (x^1, \ldots, x^{N_x})^\top \in \mathbb{R}^{N_x}$, the Lorenz 96 model (Lorenz, 1996) is given by

$$\frac{dx^i}{dt} = (x^{i+1} - x^{i-2})x^{i-1} - x^i + F, \quad i = 1, \ldots, N_x \tag{10}$$

with $x^{-1} = x^{N_x - 1}$, $x^0 = x^{N_x}$ and $x^{N_x+1} = x^1$. This is a spatio-temporal chaotic model on a one-dimensional periodic domain and often used ~~to~~ in data assimilation algorithms. We use this model to show examples of chaotic dynamics with various degrees of instability by changing the parameter F.

3 The ensemble Kalman filter with the ensemble downsizing method

3.1 The filtering problem and the ensemble transform Kalman filter

We consider a discrete-time filtering problem for the dynamics Eq. (3) with noisy observations. Let $t_n = n\tau$, $n = 0, 1, 2, \ldots$, denote the observation times with a fixed interval $\tau > 0$. We define the flow map ~~$\Psi_\tau : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ such that $\Psi_\tau(\boldsymbol{x}_0) = \boldsymbol{x}(\tau)$~~ $\boldsymbol{\Psi}_\tau : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ such that $\boldsymbol{\Psi}_\tau(\boldsymbol{x}_0) = \boldsymbol{x}(\tau)$, where $\boldsymbol{x}(t)$ is the solution to Eq. (3) with $\boldsymbol{x}(0) = \boldsymbol{x}_0$. This yields the discrete-time dynamical system

$$\boldsymbol{x}_n = \Psi\boldsymbol{\Psi}(\boldsymbol{x}_{n-1}), \quad n = 1, 2, \ldots, \tag{11}$$

where ~~$\Psi = \Psi_\tau$~~ $\boldsymbol{\Psi} = \boldsymbol{\Psi}_\tau$ and $\boldsymbol{x}_n = \boldsymbol{x}(t_n)$. The observations are obtained at each t_n as

$$\boldsymbol{y}_n = H\boldsymbol{H}\,\boldsymbol{x}_n + \boldsymbol{\eta}_n, \quad n = 1, 2, \ldots, \tag{12}$$

where ~~$H \in \mathbb{R}^{N_y \times N_x}$~~ $\boldsymbol{H} \in \mathbb{R}^{N_y \times N_x}$ is the observation matrix, and ~~$\boldsymbol{\eta}_n \sim \mathcal{N}(0, R)$~~ $\boldsymbol{\eta}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$ is a Gaussian observation noise with a symmetric positive definite covariance matrix ~~$R \in \mathbb{R}^{N_y \times N_y}$~~ $\boldsymbol{R} \in \mathbb{R}^{N_y \times N_y}$. To estimate the state $\boldsymbol{x}_n$ from the

observations $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, we employ the ensemble Kalman filter ~~(EnKF)~~ (EnKF, Evensen, 2009), which approximates the mean and covariance of the filtering distribution with an ensemble of state vectors. The EnKF consists of the forecast and analysis steps. In the forecast step, each ensemble member evolves according to the model dynamics as

$$\boldsymbol{x}_n^{f(k)} = \Psi \boldsymbol{\Psi}\left(\boldsymbol{x}_{n-1}^{a(k)}\right), \quad k = 1, \ldots, m, \tag{13}$$

where $m \in \mathbb{N}$ is the ensemble size, and the superscripts f and a denote forecast and analysis, respectively. In the analysis step, the ensemble is updated using Bayes' rule restricted to the Gaussian setting. We employ a particular analysis scheme called the ensemble transform Kalman filter ~~(ETKF)~~ (ETKF, Bishop et al., 2001). The ETKF updates the mean and perturbation part of the ensemble as

$$\overline{\boldsymbol{x}}_n^a = \overline{\boldsymbol{x}}_n^f + K \boldsymbol{K}_n \left(\boldsymbol{y}_n - H \boldsymbol{H} \overline{\boldsymbol{x}}_n^f\right), \tag{14}$$

$$V \boldsymbol{V}_n^a = V_n^f T \boldsymbol{V}_n^f \boldsymbol{T}_n, \tag{15}$$

where $\overline{\boldsymbol{x}}_n^f = \frac{1}{m}\sum_{k=1}^m \boldsymbol{x}_n^{f(k)}$, ~~$V_n^f = [\boldsymbol{x}_n^{f(1)} - \overline{\boldsymbol{x}}_n^f, \ldots, \boldsymbol{x}_n^{f(m)} - \overline{\boldsymbol{x}}_n^f] \in \mathbb{R}^{N_x \times m}$, $K_n = C_n^f H^\top (H C_n^f H^\top + R)^{-1}$~~ $\boldsymbol{V}_n^f = [\boldsymbol{x}_n^{f(1)} - \overline{\boldsymbol{x}}_n^f, \ldots, \boldsymbol{x}_n^{f(m)} - \overline{\boldsymbol{x}}_n^f]$, $\boldsymbol{K}_n = \boldsymbol{C}_n^f \boldsymbol{H}^\top (\boldsymbol{H} \boldsymbol{C}_n^f \boldsymbol{H}^\top + \boldsymbol{R})^{-1}$ is the Kalman gain, ~~$C_n^f = V_n^f (V_n^f)^\top / (m-1)$~~ $\boldsymbol{C}_n^f = \boldsymbol{V}_n^f (\boldsymbol{V}_n^f)^\top / (m-1)$ is the forecast covariance, and ~~$T_n \in \mathbb{R}^{m \times m}$~~ $\boldsymbol{T}_n \in \mathbb{R}^{m \times m}$ is a transform matrix defined as

$$T \boldsymbol{T}_n = \left(I \boldsymbol{I}_m + \frac{1}{m-1} (V \boldsymbol{V}_n^f)^\top H^\top R^{-1} H V \boldsymbol{H}^\top \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{V}_n^f\right)^{-1/2}, \tag{16}$$

where the matrix square root is chosen to be symmetric positive definite. Finally, the analysis ensemble members are reconstructed as

$$\boldsymbol{x}_n^{a(k)} = \overline{\boldsymbol{x}}_n^a + \boldsymbol{v}_n^{a(k)}, \quad k = 1, \ldots, m, \tag{17}$$

where $\boldsymbol{v}_n^{a(k)}$ denotes the k-th column of ~~V_n^a~~ $\boldsymbol{V}_n^a$.

As mentioned in Sect. 1, the forecast ensemble is corrected more strongly in directions with higher uncertainty, as represented by the forecast covariance ~~C_n^f~~ $\boldsymbol{C}_n^f$. However, the rank of ~~C_n^f~~ $\boldsymbol{C}_n^f$ is at most $m-1$ with m ensemble members. As well as this rank deficiency, the ensemble covariance suffers from the underestimation of variance due to the limited ensemble size. To mitigate these issues, we employ multiplicative covariance inflation:

$$V \boldsymbol{V}_n^f \leftarrow \alpha V \boldsymbol{V}_n^f, \tag{18}$$

where ~~$\alpha > 1$~~ $\alpha \geq 1$ is the inflation factor.

To evaluate the long-term performance of the filter along the context of rigorous error analysis (Kelly et al., 2014; Kelly and Stuart, 2019; Takeda and Sakajo, 2024; Biswas and Branicki, 2024; Sanz-Alonso and Waniorek, 2025), We define filter accuracy as follows. Assume ~~$R = r^2 I_{N_y}$ with small~~ $\boldsymbol{R} = r^2 \boldsymbol{I}_{N_y}$ with $r > 0$. The EnKF achieves filter accuracy if there exists a constant $c > 0$, independent of r, such that

$$\limsup_{n \to \infty} \mathbb{E}[\mathrm{SE}_n] = \limsup_{n \to \infty} \mathbb{E}[\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|^2] \leq c r^2 \tag{19}$$

for sufficiently small r, where the expectation is taken ~~over the~~ with respect to the probability distributions of the observation noise and the initial ensemble. ~~This property guarantees that the squared analysis error remains of order r^2.~~ We reformulate the minimum ensemble size m^* for the ETKF based on this definition of filter accuracy as the smallest $m \in \mathbb{N}$ such that the ETKF achieves filter accuracy Eq. (19) with an appropriate choice of the inflation factor $\alpha > 1$. As explained Sect. 1, this formulation differs from the standard criterion using the RMSE with a fixed r in two aspects. First, the filter accuracy Eq. (19) implies that the time-average of the expectation of RMSE is the order of r. But, the converse is not necessarily true due to Jensen's convex inequality. Thus, Eq. (19) provides a stronger guarantee for filter performance than the standard criterion (see Appendix A for more details). Second, the observation noise level r is treated as an asymptotic parameter, which makes rigorous mathematical analysis easier and qualitatively distinguishes between divergent and accurate filtering behavior using the dependency on r. Owing to this formulation, we can clearly define the minimum ensemble size m^* and simplify the analysis of its relationship with the geometric properties of the instability of dynamics.

3.2 The ensemble downsizing method

To generate an ensemble with its mean close to the true state and its perturbations aligned with the unstable subspace, we introduce an ensemble downsizing method. We begin with a sufficiently large ensemble size, m_0, and reduce it to a smaller size, m, at a fixed spin-up time, $n = N_{\mathrm{spinup}}$. We call the period before $n = N_{\mathrm{spinup}}$ the ensemble spin-up period. In the ensemble downsizing method, we apply the singular value decomposition (SVD) to the ensemble perturbation, ~~$V \in \mathbb{R}^{N_x \times m_0}$~~ $\boldsymbol{V} \in \mathbb{R}^{N_x \times m_0}$, and retain only the leading m modes. This procedure is detailed in Algorithm 2. We suppose that the SVD algorithm returns the singular values in decreasing order and the associated singular vectors accordingly. For a matrix A, we use MATLAB-style indexing notation: $A[:, 1:m]$ denotes the submatrix formed by the first m columns of A, and $A[1:m, 1:m]$ denotes the leading $m \times m$ principal submatrix.

Algorithm 2 The ensemble downsizing method by the singular value decomposition

Require: ~~$X \in \mathbb{R}^{N_x \times m_0}$~~ $\boldsymbol{X} \in \mathbb{R}^{N_x \times m_0}$, $m < m_0$

Ensure: ~~$X = \overline{x} + V$~~ $\boldsymbol{X} = \overline{\boldsymbol{x}} + \boldsymbol{V}$

1: ~~$U, \Sigma, \_ \leftarrow \mathrm{SVD}(V)$~~ $\boldsymbol{U}, \boldsymbol{\Sigma}, \_ \leftarrow \mathrm{SVD}(\boldsymbol{V})$

2: ~~$V \leftarrow U[:, 1:m]\, \Sigma[1:m, 1:m]$~~ $\boldsymbol{V} \leftarrow \boldsymbol{U}[:, 1:m]\, \boldsymbol{\Sigma}[1:m, 1:m]$

3: **return** ~~$\overline{x} + V$~~ $\overline{\boldsymbol{x}} + \boldsymbol{V}$

The resulting ETKF with the multiplicative covariance inflation Eq. (18) and the ensemble downsizing method is summarized in Algorithm 3. See Tippett et al. (2003) for an efficient implementation of the analysis step.

4 Numerical results

To verify our conjecture that the minimum ensemble size for ~~asymptotic~~ filter accuracy of the ETKF based on Eq. (19) is $m^* = N_+ + 1$, we perform numerical experiments with the Lorenz 96 model. Throughout this section, we set $N_x = 40$, ~~$H = I_{N_x}$ and~~

Algorithm 3 The ETKF with multiplicative covariance inflation and ensemble downsizing

Require: ~~Ψ, H, R~~ $\boldsymbol{\Psi}, \boldsymbol{H}, \boldsymbol{R}$, $(\boldsymbol{y}_n)_{n=1}^N$, ~~$X_0 \in \mathbb{R}^{N_x \times m_0}$~~ $\boldsymbol{X}_0 \in \mathbb{R}^{N_x \times m_0}$, $\alpha > 1$, $N_{\mathrm{spinup}} < N$, $m < m_0$ ~~,~~

Ensure: ~~$X = (\boldsymbol{x}^{(k)})_{k=1}^{m'}$~~ $\boldsymbol{X} = (\boldsymbol{x}^{(k)})_{k=1}^{m'}$

~~$X \leftarrow X_0$~~ $\boldsymbol{X} \leftarrow \boldsymbol{X}_0$

$m' \leftarrow m_0$

for $n = 1$ **to** N **do**

  # Forecast step

  for $k = 1$ **to** m' **do**

    ~~$\boldsymbol{x}^{f(k)} \leftarrow \Psi(\boldsymbol{x}^{a(k)})$~~ $\boldsymbol{x}^{f(k)} \leftarrow \boldsymbol{\Psi}(\boldsymbol{x}^{a(k)})$

  end for

  $\overline{\boldsymbol{x}}^f \leftarrow \frac{1}{m'} \sum_{k=1}^{m'} \boldsymbol{x}^{f(k)}$

  ~~$V^f \leftarrow [\boldsymbol{x}^{f(1)} - \overline{\boldsymbol{x}}^f, \ldots, \boldsymbol{x}^{f(m')} - \overline{\boldsymbol{x}}^f]$~~ $\boldsymbol{V}^f \leftarrow [\boldsymbol{x}^{f(1)} - \overline{\boldsymbol{x}}^f, \ldots, \boldsymbol{x}^{f(m')} - \overline{\boldsymbol{x}}^f]$

  # Covariance inflation

  ~~$V^f \leftarrow \alpha V^f$~~ $\boldsymbol{V}^f \leftarrow \alpha \boldsymbol{V}^f$

  ~~$C^f \leftarrow V^f (V^f)^\top / (m' - 1)$~~ $\boldsymbol{C}^f \leftarrow \boldsymbol{V}^f (\boldsymbol{V}^f)^\top / (m' - 1)$

  # Analysis step

  ~~$K \leftarrow C^f H^\top (H C^f H^\top + R)^{-1}$~~ $\boldsymbol{K} \leftarrow \boldsymbol{C}^f \boldsymbol{H}^\top (\boldsymbol{H} \boldsymbol{C}^f \boldsymbol{H}^\top + \boldsymbol{R})^{-1}$

  ~~$\overline{\boldsymbol{x}}^a \leftarrow \overline{\boldsymbol{x}}^f + K(\boldsymbol{y}_n - H \overline{\boldsymbol{x}}^f)$~~ $\overline{\boldsymbol{x}}^a \leftarrow \overline{\boldsymbol{x}}^f + K(\boldsymbol{y}_n - \boldsymbol{H} \overline{\boldsymbol{x}}^f)$

  ~~$T \leftarrow (I_{m'} + \frac{1}{m'-1} (V^f)^\top H^\top R^{-1} H V^f)^{-1/2}$~~ $\boldsymbol{T} \leftarrow (\boldsymbol{I}_{m'} + \frac{1}{m'-1} (\boldsymbol{V}^f)^\top \boldsymbol{H}^\top \boldsymbol{R}^{-1} \boldsymbol{H} \boldsymbol{V}^f)^{-1/2}$

  ~~$V^a \leftarrow V^f T$~~ $\boldsymbol{V}^a \leftarrow \boldsymbol{V}^f \boldsymbol{T}$

  ~~$X \leftarrow \overline{\boldsymbol{x}}^a + V^a$~~ $\boldsymbol{X} \leftarrow \overline{\boldsymbol{x}}^a + \boldsymbol{V}^a$

  ~~$X_n \leftarrow X$~~ $\boldsymbol{X}_n \leftarrow \boldsymbol{X}$

  # Ensemble downsizing

  if $n = N_{\mathrm{spinup}}$ **then**

    ~~$X \leftarrow \mathrm{EnsembleDownsizing}(X, m)$~~ $\boldsymbol{X} \leftarrow \mathrm{EnsembleDownsizing}(\boldsymbol{X}, m)$

    $m' \leftarrow m$

  end if

end for

return ~~$(X_n)_{n=1}^N$~~ $(\boldsymbol{X}_n)_{n=1}^N$

~~$R = r^2 I_{N_x}$~~ $\boldsymbol{H} = \boldsymbol{I}_{N_x}$ and $R = r^2 \boldsymbol{I}_{N_x}$, where $r > 0$ is a parameter representing the standard deviation of the observation noise. We consider two settings for the external forcing: $F = 8$~~and $F = 16$~~, 16. For each setting, we compute the LEs and estimate N_+ using Algorithm 1. Then, we apply the ETKF with the ensemble downsizing method (Algorithm 3) to the Lorenz 96 model. We summarize the common parameters for the ETKF experiments in Table 1. The initial ensemble ~~X_0~~ $\boldsymbol{X}_0$ is defined as ~~$X_0 = (\boldsymbol{x}_0^{(k)})_{k=1}^{m_0}$ with $\boldsymbol{x}_0^{(k)} \sim N(\boldsymbol{x}_0, 25 I_{N_x})$~~ $\boldsymbol{X}_0 = (\boldsymbol{x}_0^{(k)})_{k=1}^{m_0}$ with $\boldsymbol{x}_0^{(k)} \sim \mathcal{N}(\boldsymbol{x}_0, 25 \boldsymbol{I}_{N_x})$, where $\boldsymbol{x}_0 \in \mathbb{R}^{N_x}$ is uniformly sampled from the true trajectory.

Table 1. Common parameters for the ETKF experiments.

Parameter	Value	Description
Δt	0.01	Time step size for the model integration
N	$72{,}000$ $(= 10 \times 360 \times 20)$	Total number of integration steps
m	12, 13, 14, 15, 16, 17, 18	Ensemble size after downsizing
m_0	41 $(= N_x + 1)$	Ensemble size before downsizing
α	1.0, 1.1, 1.2, 1.3, 1.4, 1.5	Inflation factor
r	$10^0, 10^{-1}, \ldots, 10^{-4}$	Standard deviation of the observation noise
n_{obs}	5 (for $F = 8$), 2 (for $F = 16$)	Observation interval (integration steps)
N_{spinup}	720 (for $F = 8$), 1800 (for $F = 16$)	Spin-up period (assimilation steps)

To evaluate the filter accuracy of the ETKF, we use the ~~squared error (SE)~~ SE as in Eq. (19). To approximate the expectation $\mathbb{E}$, we compute parallel simulations for ~~n_{seeds}~~ n_{seeds} random seeds to generate the observation noises. Then, we take the maximum after $n = N_\infty$ to approximate $\limsup_{n\to\infty}$. This procedure leads to

$$\limsup_{n\to\infty} \mathbb{E}[\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|^2] \approx \max_{n \geq N_\infty} \frac{1}{n_{seeds}} \sum_{i=1} \frac{1}{n_{\mathrm{seeds}}} \sum_{i=1}^{n_{seeds}} \|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a(\omega_i)\|^2, \tag{20}$$

where $\overline{\boldsymbol{x}}_n^a(\omega_i)$ is the analysis mean of a sample path with the i-th random seed. We use $N_\infty = N/2$ and ~~$n_{seeds} = 10$ to approximate~~ $n_{\mathrm{seeds}} = 10$ to compute Eq. (20). On the other hand, we use the ~~root mean squared error (RMSE)~~ RMSE to visualize the time series of the analysis error for a sample run.

4.1 $F = 8$

We first set $F = 8$, a typical parameter for which the Lorenz 96 model exhibits chaotic behavior. The LEs are computed using Algorithm 1 with $\Delta t = 0.001$ and $n = 10^6$ (~~Figure~~Fig. 1). In the computation, we define the index of the zero exponent as the minimizer of $i \mapsto |\lambda_i|$. This yields $N_+ = 13$ and the largest LE $\lambda_1 \approx 1.67$.

In this section, we ~~assimilate observations every~~ set the observation time interval 0.05 (i.e., $n_{\mathrm{obs}} = 5$~~integration steps~~). We reduce the ensemble size after $N_{\mathrm{spinup}} = 720$ assimilation steps. For each pair (r, m), we vary the inflation factor α to find the optimal value that minimizes the SE defined in Eq. (20). The results are shown in ~~Figure~~Fig. 2 with log-log plots of the SE

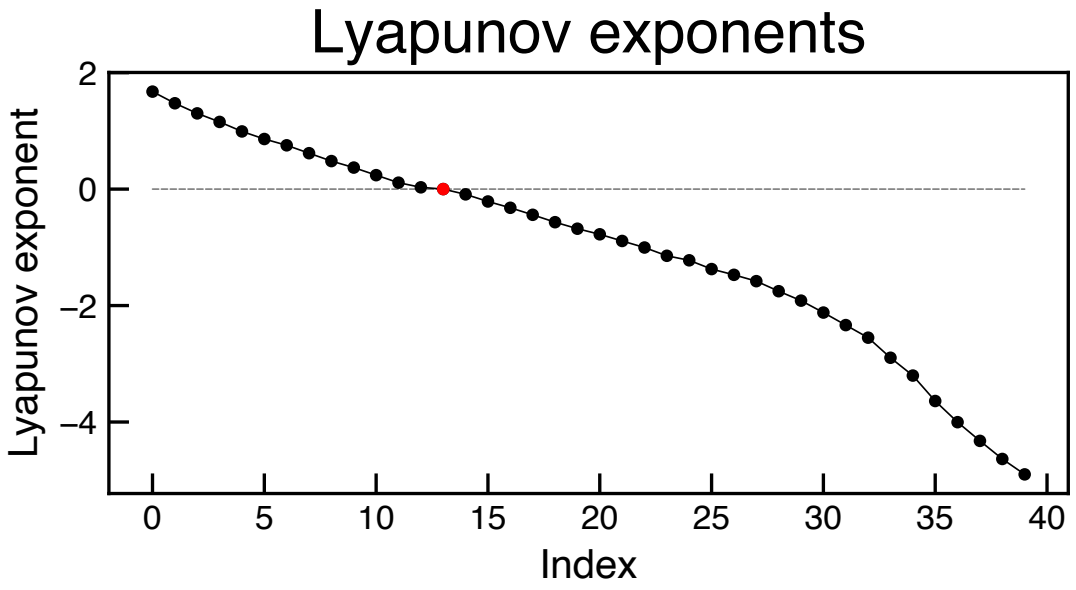
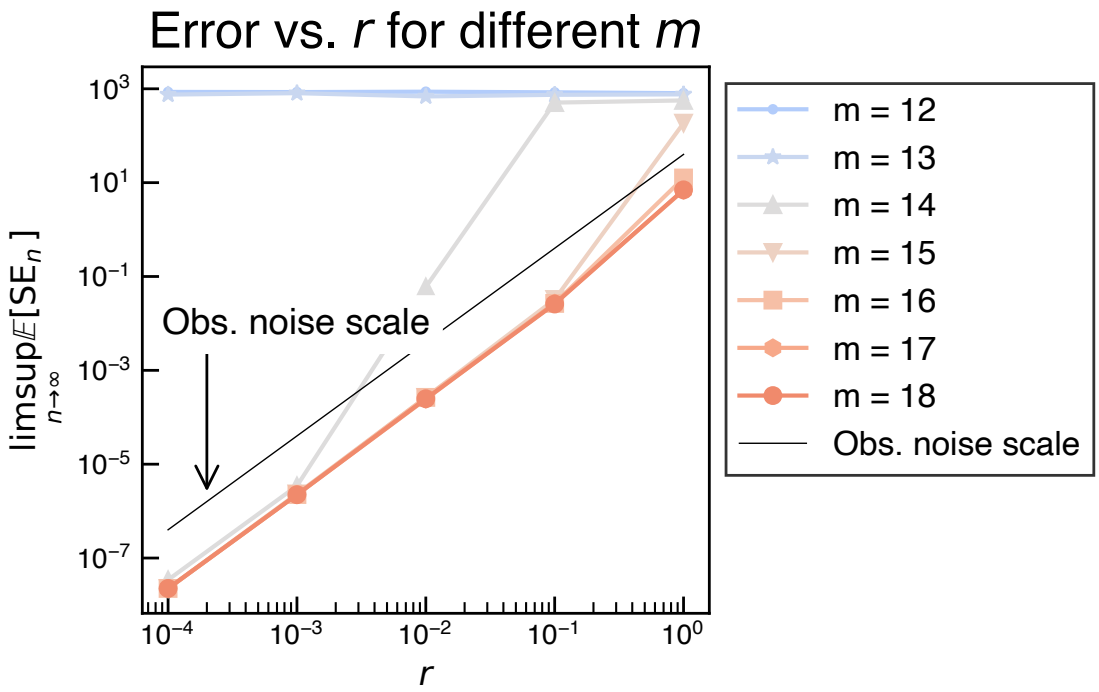


Figure 1. The LEs of the Lorenz 96 model with $(J, F) = (40, 8)$. The zero exponent $\lambda_{14} = 0$ is indicated in red.

against r for different m. If m is larger than or equal to $N_+ + 1 = 14$, the SE is bounded by the order of r^2, indicating that the ETKF achieves filter accuracy. Conversely, if m is smaller than 14, the SE stays around $O(1)$ even for small r, indicating that the ETKF does not achieve filter accuracy. Accordingly, our formulation of the filter accuracy Eq. (19) qualitatively distinguishes between divergent and accurate filtering behavior and evaluates the minimum ensemble size as $m^* = N_+ + 1 = 14$ in this setting. In particular, in the border case $m = 14$, the ensemble spin-up works effectively because the SE is larger than the order of r^2 for relatively large r but becomes comparable to it for small r. Since the requirements of small r in rigorous mathematical analysis are for asymptotic approximation, we examine an alternative condition, a short time interval, for practical settings with fixed r in Appendix B.

~~In the following two experiments, we fix the number of integration steps $N = 72,0000$ and single random seed to generate the observation noise.~~ We then conduct ~~an experiment~~ a longer experiment ($N = 720,000$) to investigate the dependence of the RMSE on the spin-up period N_{spinup} with a single random seed to generate the observation noise. We set ~~$r = 10^{-1}$ and $m = 15$~~$r = 10^{-4}$ and $m = 14$, which avoids filter divergence in the experiment for ~~Figure~~Fig. 2. For $N_{\text{spinup}} = 0$ and 720, we show the time series of the RMSE with various α in ~~Figure~~Fig. 3. ~~From Figure~~In Fig. 3 (a), ~~we observe that the RMSE with $\alpha \geq 1.1$ remains small even for a longer period when $N_{\text{spinup}} = 720$. Similarly, in Figure 3 (b), the RMSE with $\alpha \geq 1.2$ remains small for a longer periodwhen $N_{\text{spinup}} = 0$. This suggests that filter stability is achieved even without an ensemble spin-up period if a sufficiently large inflation factor is used. In addition, we find that the ensemble spin-up period can reduce the~~



~~Log-log plots of the SE vs. r for different ensemble sizes m after the downsizing. The dashed line indicates the order of r^2, corresponding to the observation noise level. The Lorenz 96 model with $F = 8$ ($N_+ = 13$) is used.~~

Figure 2. Log-log plots of the SE vs. r for different ensemble sizes m after the downsizing. The black line indicates the order of r^2, corresponding to the observation noise scale. The Lorenz 96 model with $F = 8$ ($N_+ = 13$) is used.

~~required inflation factor for filter accuracy.The time series of the RMSE with $r = 10^{-1}$, $m = 15$ and various $\alpha = 1.0, \ldots, 1.5$ for $N_{\mathrm{spinup}} = 720$ (a) and $N_{\mathrm{spinup}} = 0$ (b). The dashed line indicates the level r. Figure ?? shows the time series of the RMSE with $r = 10^{-4}$~~all series of the RMSE quickly decay to values with the order of r. Moreover, the filter remains stable over a long assimilation period. In Fig. 3 (b), ~~$N_{\mathrm{spinup}} = 0$ and various α for $m = 14$ (a) and $m = 13$ (b). In Figure ?? (a),~~ the RMSE with $\alpha = 1.5$ decays to a small value. The time required for the RMSE to decay is much longer than that ~~for the results in Figure 3~~in Fig. 3 (a). A potential explanation for this phenomenon is the slow decay of the uncertainty in the neutral direction. Since we focus on the time asymptotic accuracy, this phenomenon is not further investigated in this study. ~~In Figure ?? (b~~From these results, we conclude that the ensemble spin-up and downsizing method effectively saves the computational time to generate a small ensemble sustaining filter accuracy, which is crucial in more high-dimensional applications.

We then investigate the effect of the ensemble spin-up on the ensemble alignment with the unstable subspace. The integration time is $N = 360,000$ and a single random seed is used to generate the observation noise. To clarify the effect of the ensemble

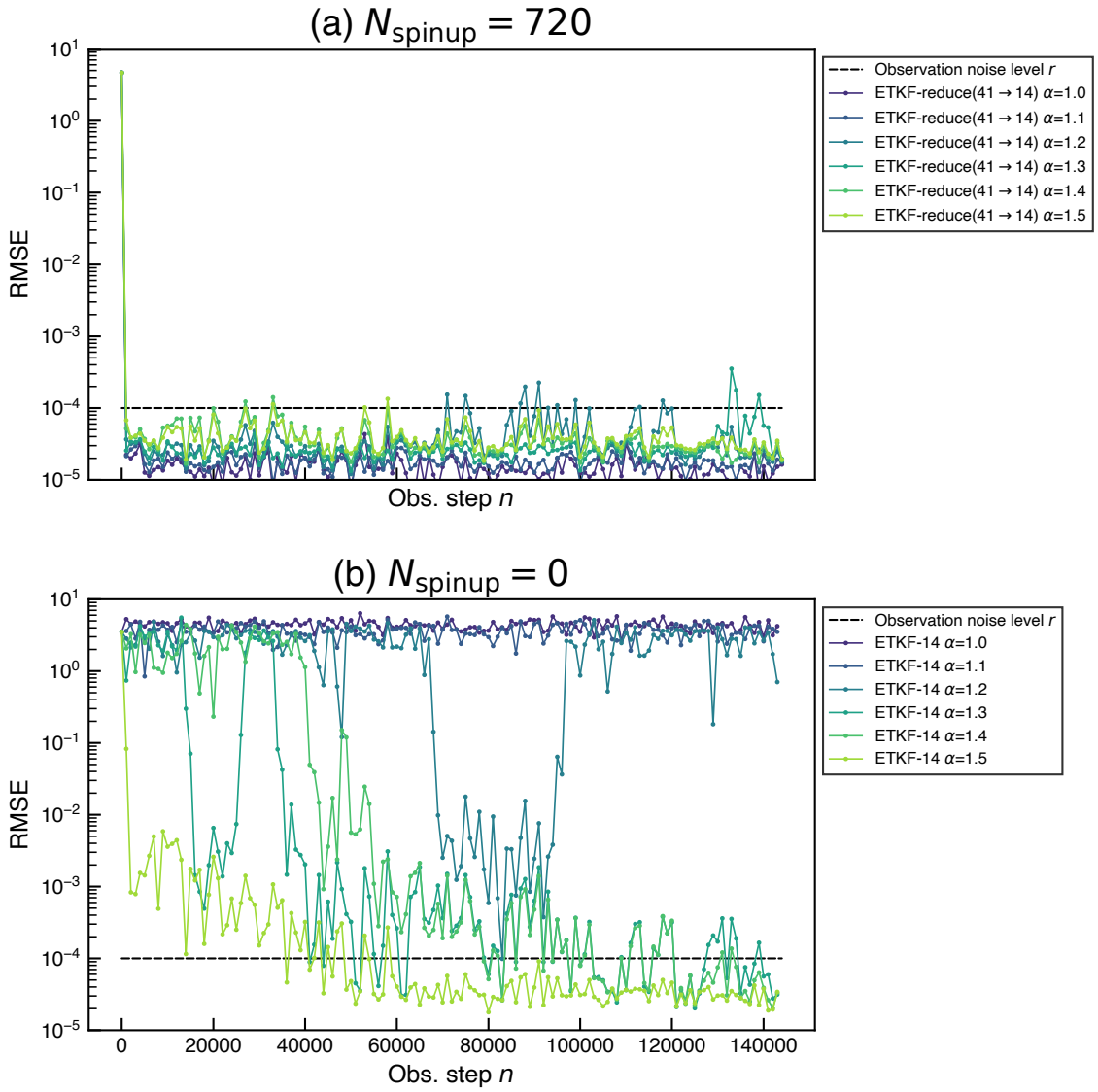


Figure 3. The time series of the RMSE with $r = 10^{-4}$, $m = 14$ and various $\alpha = 1.0, \ldots, 1.5$ for (a) $N_{\mathrm{spinup}} = 720$ and (b) $N_{\mathrm{spinup}} = 0$. The dashed line indicates the level r.

downsizing method related to the unstable subspace, we perform experiments with initial ensemble which has accurate mean close to the true state and inaccurate perturbations not aligned in the unstable subspace. Specifically, the initial ensemble $\boldsymbol{X}_0 \in \mathbb{R}^{N_x \times m}$ is generated as

$$\boldsymbol{X}_0 = [\overline{\boldsymbol{x}}_0 + \boldsymbol{v}_0^{(k)}]_{k=1}^m, \quad \overline{\boldsymbol{x}}_0 \sim \mathcal{N}(\boldsymbol{x}_0, 0.01\boldsymbol{R}), \quad \boldsymbol{v}_0^{(k)} \sim \mathcal{N}(\boldsymbol{0}, 0.01\boldsymbol{R}), \mathit{allRMSEvaluesdonotdecayandremainlarge} \tag{21}$$

for some $m \in \mathbb{N}$. This construction yields that the expected initial RMSE is the order of $0.1r$, but the alignment of the ensemble perturbations with the unstable subspace is not guaranteed. Figure 4 shows the time series of the RMSE with $r = 10^{-4}$, $m = 14$, various $\alpha = 1.0, \ldots, 1.5$ for (a) $N_{\mathrm{spinup}} = 720$ and (b) $N_{\mathrm{spinup}} = 0$. While the RMSE with the ensemble spin-up and downsizing method in Fig. 4 (a) sustains the order of r for almost all α, the RMSE without the ensemble spin-up immediately diverges for all α and then decays very slowly for only some α. These results ~~indicate that the minimum ensemble size for filter accuracy is $m^* = N_+ + 1 = 14$ regardless of the~~ distinguish the ETKF with the ensemble spin-up and downsizing method from the one without it in terms of sustaining filter accuracy with inaccurate initial perturbations not aligned with the unstable subspace. To capture the unstable subspace of nonlinear dynamical systems, accurate state estimation is necessary but not sufficient since the unstable subspace depends on the state. The result in Fig. 4 implies that the ensemble spin-up ~~period.~~ method can assist not only the convergence of the ensemble mean to the true state but also the alignment of the ensemble perturbations with the unstable subspace.

4.2 $F = 16$

~~We~~ To verify that $m^* = N_+ + 1$ also holds with different value of N_+, we set $F = 16$ and compute the LEs as in the previous section, shown in ~~Figure~~Fig. 5. This yields $N_+ = 15$ and the largest LE $\lambda_1 \approx 3.82$. We ~~assimilate observations every~~ set the observation time interval 0.02 (i.e., $n_{\mathrm{obs}} = 2$~~integration steps~~) in this section. ~~This value of n_{obs} is~~ Since we control the interval using n_{obs} in the experiments, the value is determined by the largest integer n_{obs} with which the approximated error expansion $n_{\mathrm{obs}}\lambda_1\Delta t$ in the forecast step for $F = 16$ does not exceed that with $n_{\mathrm{obs}} = 5$ for $F = 8$. Indeed, if we write these quantities for F as $n_{\mathrm{obs}}^{(F)}$ and $\lambda_1^{(F)}$, the error expansion with each F is approximated as

$$n_{\mathrm{obs}}^{(8)}\lambda_1^{(8)}\Delta t \approx 5 \cdot 1.67 \cdot 0.01 = 8.35,$$
$$n_{\mathrm{obs}}^{(16)}\lambda_1^{(16)}\Delta t \approx 2 \cdot 3.82 \cdot 0.01 = 7.64.$$

We compute the SE in the same manner as in the previous section with $N = 72{,}000$ integration steps and $N_{\mathrm{spinup}} = 1800$ assimilation steps, which yields the same integration steps before the ensemble downsizing method. The dependence of the SE on r for different m is shown in ~~Figure~~Fig. 6. As in the previous section, $m \geq N_+ + 1 = 16$ gives filter accuracy, while $m < 16$ does not. Therefore, the minimum ensemble size for filter accuracy is $m^* = N_+ + 1 = 16$.

5 Conclusions

We ~~proposed an ensemble~~ reformulated the minimum ensemble size for filter accuracy of the EnKF based on the r-asymptotic filter accuracy Eq. (19), and investigated its relationship with the instability of dynamics characterized by the LEs. To obtain the

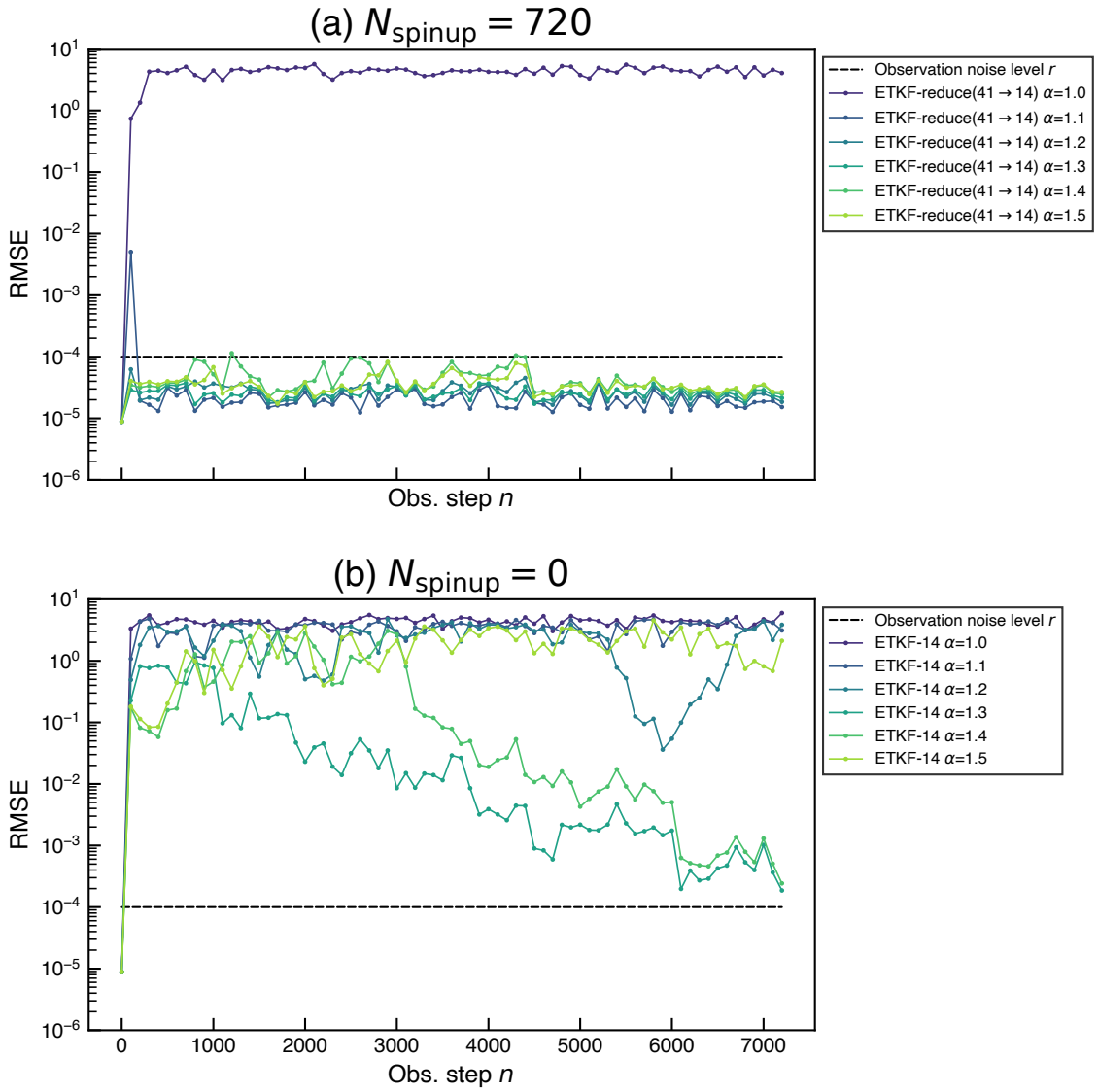


Figure 4. The time series of the RMSE with $r = 10^{-4}$, ~~$N_{\mathrm{spinup}} = 0$ and~~ $m = 14$, various $\alpha = 1.0, \ldots, 1.5$ for ~~$m = 14$~~ (a) $N_{\mathrm{spinup}} = 720$ and ~~$m = 13$~~ (b) $N_{\mathrm{spinup}} = 0$. The accurate initial ensemble is given by Eq. (21). The dashed line indicates the level r.

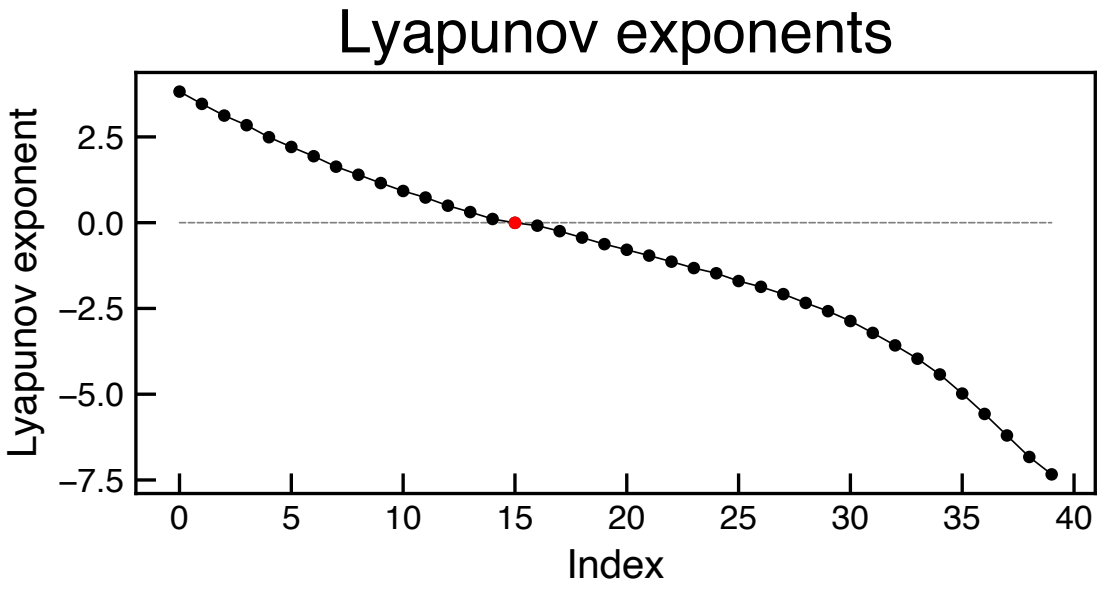


Figure 5. The LEs of the Lorenz 96 model with $(J, F) = (40, 16)$. The zero exponent $\lambda_{16} = 0$ is indicated in red.

effective small ensemble sustaining the filter accuracy, we proposed the ensemble spin-up and downsizing method for the EnKF to ~~generate~~ generating an ensemble aligned with the unstable subspace of the dynamics. Through numerical experiments with the ETKF applied to the Lorenz 96 model, we verified our conjecture that the minimum ensemble size for ~~asymptotic~~ the filter accuracy is $m^* = N_+ + 1$, where N_+ is the number of positive LEs (~~Figure~~Fig. 1, 5). In particular, our formulation qualitatively distinguishes between divergent and accurate filtering behavior through the observation noise scaling. This estimate of m^* is valid for the multiple external forcing F in the Lorenz 96 model (~~Figures~~Figs. 2 and 6), and the filter remains stable over long ~~integration periods (Figures 3and ??). Moreover, filter accuracy is achieved even~~ assimilation periods (Fig. 3). Even without an ensemble spin-up period~~(Figure ??~~, the filter accuracy is achieved (Fig. 3). The ensemble downsizing method ~~offers practical advantages: it mitigates the slow convergence of uncertainty in the neutral direction when $m = m^*$ and can reduce the required inflation factorfor filter accuracy.~~ has an advantage in the border case $m = m^*$ as the error decays quickly with it while the error decays very slowly without it, which is a practical advantage of this method. Since our results depend on the choice of the inflation factor, we recommend employing an adaptive inflation scheme such as the EnKF-N (Bocquet, 2011) to avoid manual tuning.

In this study, the estimate of the minimum ensemble size $m^* = N_+ + 1$ has been verified only for systems with a single zero LE. In general, there may exist multiple zero LEs, which can lead to a larger difference between N_+ and N_0. Further studies are needed to verify whether the estimate $m^* = N_+ + 1$ holds in such cases. Suitable dynamical systems for this purpose include

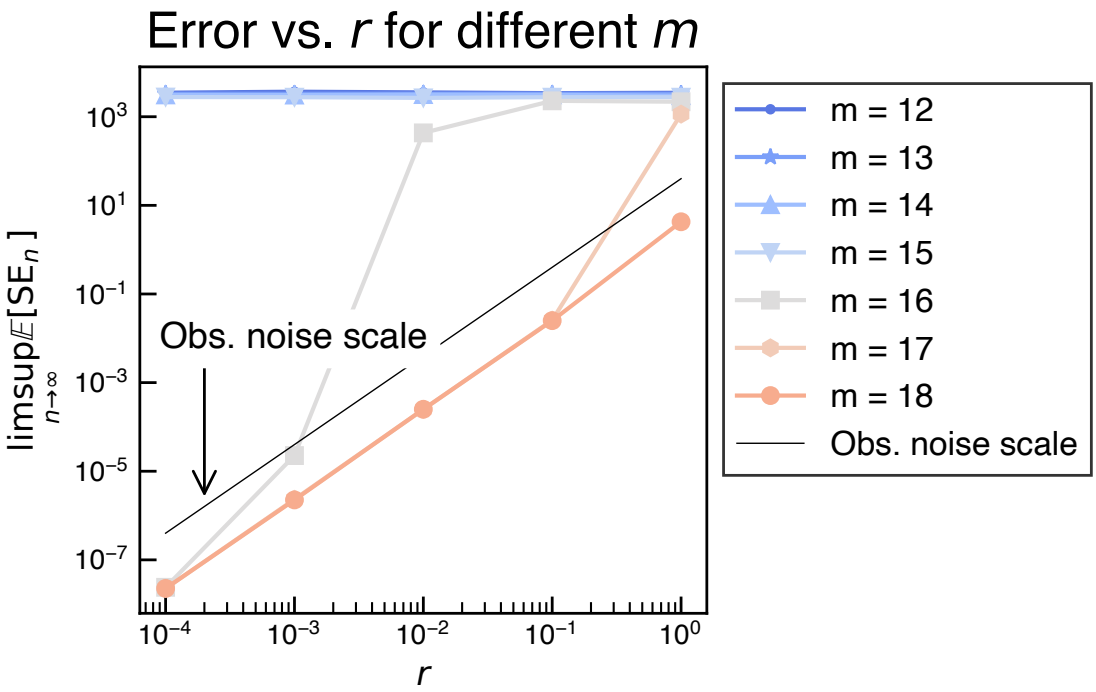


Figure 6. Log-log plots of the SE vs. r for different ensemble sizes m after the downsizing. The ~~dashed~~ black line indicates the order of ~~r^2~~r^2, corresponding to the observation noise ~~level~~scale. The Lorenz 96 model with $F = 16$ ($N_+ = 15$) is used.

Hamiltonian systems with multiple zero LEs or the Modular Arbitrary-Order Ocean-Atmosphere Model (De Cruz et al., 2016) which exhibits many negative LEs close to zero as discussed in (Carrassi et al., 2022). Additionally, an explicit quantitative analysis of the alignment between ensemble covariance eigenvectors and Lyapunov vectors is an important topic, left for future work. This analysis will clarify the mechanism of the ensemble alignment within the spin-up period and enhance the validity of the ensemble downsizing method. The other future direction is to evaluate the minimum ensemble size with localization, in which we need to define 'local degrees of instability' associated with a localization radius.

. Code availability

The code is available at https://github.com/KotaTakeda/enkf_ensemble_downsizing/releases/tag/v1.0.0 and archived on Zenodo: https://doi.org/10.5281/zenodo.17319854.

Appendix A: Comparison of error evaluation criteria

Mathematical studies for filters often focus on the long-term behavior of the analysis error (Kelly et al., 2014; Kelly and Stuart, 2019; Takeda and Sakajo, 2024; Biswas and Branicki, 2024; Sanz-Alonso and Waniorek, 2025) and aim to establish the bound known as time-asymptotic filter accuracy, which is defined as Eq. (19) in the manuscript. It aims to bound the expectation of the squared error

$$\mathrm{SE}_n = \mathbb{E}[\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|^2].$$

Here, we compare this criterion with the commonly used RMSE. The expectation of the RMSE at time t_n is defined as

$$\mathrm{RMSE}_n = \mathbb{E}\left[\frac{\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|}{\sqrt{N_x}}\right].$$

From Jensen's convex inequality, we have

$$N_x \mathrm{RMSE}_n^2 = \mathbb{E}\left[\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|\right]^2 \leq \mathbb{E}[\|\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n^a\|^2] = \mathrm{SE}_n.$$

This implies that small values of the expectation of the squared error lead to small RMSE values in expectation, but not vice versa. In addition, supremum over time in the asymptotic limit is larger than or equal to the time-averaged value in general. Hence, the criterion Eq. (19) is stronger than the time-averaged RMSE criterion.

Appendix B: RMSE with a small observation interval

In Fig. 2, the SE with $m = 14$ (i.e., $m = N_+ + 1$) is of order 10^3 and exceeds the $O(r^2)$ scaling for large r (around 10^0). Although this does not contradict filter accuracy, such a large observation noise level can be found in practical applications. We consider an alternative condition of a short observation time interval to achieve filter accuracy. Hence, in the following experiments, we apply the ETKF with $m = 14$ to the Lorenz 96 model with $(N_x, F) = (40, 8.0)$ for large observation noise $r = 1.0$ and a short observation interval 0.001 (i.e., $n_{\mathrm{obs}} = 1$ and $\Delta t = 0.001$). We set $N = 720,000$ integration time steps and $n_{\mathrm{seeds}} = 10$ for approximating the expectation. We plot the time series of $\mathbb{E}[\mathrm{SE}_n]$ in Fig. B1. Compared to the SE with $m = 14$ and $r = 1$ shown in Fig. 2, the supremum of the SE in Fig. B1 is substantially reduced, from values of order 10^3 to values only slightly larger than $N_x r^2 = 40$. This result implies that a small observation interval improves accuracy to the order of the observation noise r, even when r is large.

. Author Contribution

KT is responsible for all plotting, analysis, and writing. TM provided significant discussions and inputs for this study.

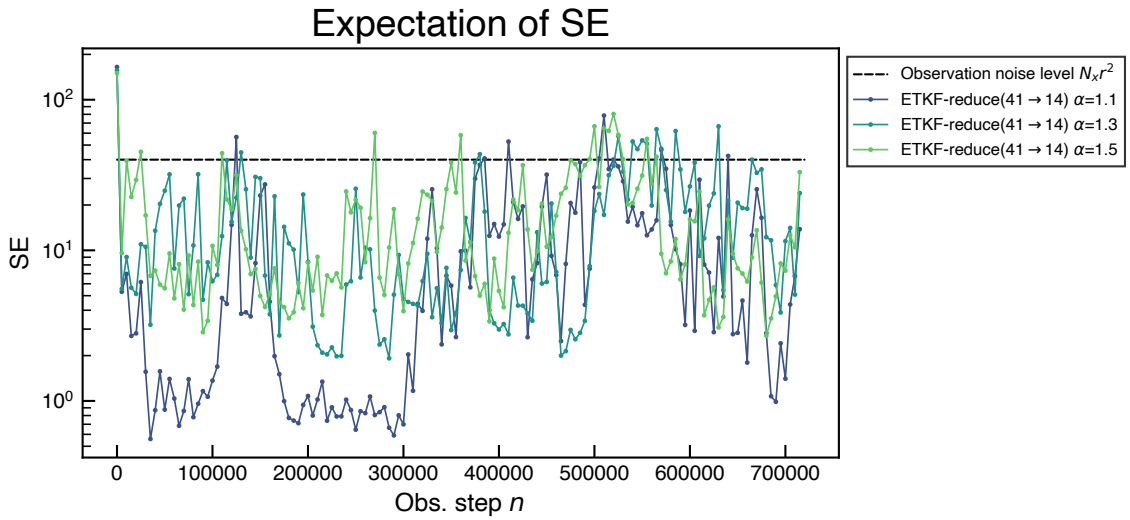


Figure B1. The time series of the SE with $r = 1.0$, $m = 14$, $n_{\mathrm{obs}} = 1$, $\Delta t = 0.001$, and various α. The dashed line indicates the level $N_x r^2$.

. Competing Interests

Some authors are members of the editorial board of journal NPG.

. Acknowledgements We used an AI tool to edit or polish the authors' written text for spelling, grammar, or general style.

. Financial Support

The first author was partially supported by RIKEN Junior Research Associate Program and JST SPRING JPMJSP2110. The second author was supported by JST SATREPS (JPMJSA2109), JST CREST (JPMJCR24Q3), JSPS KAKENHI (JP24H00021), Japan Aerospace Exploration Agency (EORA4), the COE research grant in computational science from Hyogo Prefecture and Kobe City through Foundation for Computational Science and the RIKEN TRIP initiative (RIKEN Prediction Science).

References

Barreira, L. and Pesin, Y.: Lyapunov Exponents and Smooth Ergodic Theory, vol. 23 of *University Lecture Series*, American Mathematical Society, Providence, Rhode Island, ISBN 978-0-8218-2921-9 978-1-4704-2170-0, https://doi.org/10.1090/ulect/023, 2002.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Weather Rev., 129, 420–436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Biswas, A. and Branicki, M.: A Unified Framework for the Analysis of Accuracy and Stability of a Class of Approximate Gaussian Filters for the Navier–Stokes Equations, Nonlinearity, 37, 125 013, https://doi.org/10.1088/1361-6544/ad805b, 2024.

Bocquet, M.: Ensemble Kalman Filtering without the Intrinsic Need for Inflation, Nonlinear Process. Geophys., 18, 735–750, https://doi.org/10.5194/npg-18-735-2011, 2011.

Bocquet, M. and Carrassi, A.: Four-Dimensional Ensemble Variational Data Assimilation and the Unstable Subspace, Tellus Dyn. Meteorol. Oceanogr., 69, 1304 504, https://doi.org/10.1080/16000870.2017.1304504, 2017.

Bocquet, M., Gurumoorthy, K. S., Apte, A., Carrassi, A., Grudzien, C., and Jones, C. K. R. T.: Degenerate Kalman Filter Error Covariances and Their Convergence onto the Unstable Subspace, SIAM/ASA J. Uncertainty Quantification, 5, 304–333, https://doi.org/10.1137/16M1068712, 2017.

Carrassi, A., Bocquet, M., Demaeyer, J., Grudzien, C., Raanes, P., and Vannitsem, S.: Data Assimilation for Chaotic Dynamics, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV), edited by Park, S. K. and Xu, L., pp. 1–42, Springer International Publishing, Cham, ISBN 978-3-030-77722-7, https://doi.org/10.1007/978-3-030-77722-7_1, 2022.

De Cruz, L., Demaeyer, J., and Vannitsem, S.: The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0, Geosci. Model Dev., 9, 2793–2808, https://doi.org/10.5194/gmd-9-2793-2016, 2016.

Eckmann, J. P. and Ruelle, D.: Ergodic Theory of Chaos and Strange Attractors, Rev. Mod. Phys., 57, 617–656, https://doi.org/10.1103/RevModPhys.57.617, 1985.

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, Springer, Berlin, Heidelberg, ISBN 978-3-642-03710-8 978-3-642-03711-5, https://doi.org/10.1007/978-3-642-03711-5, 2009.

González-Tokman, C. and Hunt, B. R.: Ensemble Data Assimilation for Hyperbolic Systems, Physica D: Nonlinear Phenomena, 243, 128–142, https://doi.org/10.1016/j.physd.2012.10.005, 2013.

Grudzien, C., Carrassi, A., and Bocquet, M.: Asymptotic Forecast Uncertainty and the Unstable Subspace in the Presence of Additive Model Error, SIAMASA J. Uncertain. Quantif., https://doi.org/10.1137/17M114073X, 2018a.

Grudzien, C., Carrassi, A., and Bocquet, M.: Chaotic Dynamics and the Role of Covariance Inflation for Reduced Rank Kalman Filters with Model Error, Nonlinear Process. Geophys., 25, 633–648, https://doi.org/10.5194/npg-25-633-2018, 2018b.

Gurumoorthy, K. S., Grudzien, C., Apte, A., Carrassi, A., and Jones, C. K. R. T.: Rank Deficiency of Kalman Error Covariance Matrices in Linear Time-Varying System With Deterministic Evolution, SIAM J. Control Optim., 55, 741–759, https://doi.org/10.1137/15M1025839, 2017.

Haken, H.: At Least One Lyapunov Exponent Vanishes If the Trajectory of an Attractor Does Not Contain a Fixed Point, Physics Letters A, 94, 71–72, https://doi.org/10.1016/0375-9601(83)90209-8, 1983.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, Mon. Weather Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter, Physica D: Nonlinear Phenomena, 230, 112–126, https://doi.org/10.1016/j.physd.2006.11.008, 2007.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, New York, https://doi.org/10.1017/CBO9780511802270, 2002.

Kelly, D. and Stuart, A. M.: Ergodicity and Accuracy of Optimal Particle Filters for Bayesian Data Assimilation, Chin. Ann. Math. Ser. B, 40, 811–842, https://doi.org/10.1007/s11401-019-0161-5, 2019.

Kelly, D. T. B., Law, K. J. H., and Stuart, A. M.: Well-Posedness and Accuracy of the Ensemble Kalman Filter in Discrete and Continuous Time, Nonlinearity, 27, 2579–2603, https://doi.org/10.1088/0951-7715/27/10/2579, 2014.

Legras, B. and Vautard, R.: A Guide to Liapunov Vectors, in: ECMWF Workshop Predict., pp. 135–146, Reading, United-Kingdom, 1996.

Lorenz, E. N.: Predictability: A Problem Partly Solved, in: Proc. Semin. Predict., vol. 1, pp. 1–18, ECMWF, Shinfield Park, Reading, 1996.

Oseledets, V. I.: A Multiplicative Ergodic Theorem. Lyapunov Characteristic Numbers for Dynamical Systems, Trans. Mosc. Math. Soc., 19, 197–231, 1968.

Reinhold, B. B. and Pierrehumbert, R. T.: Dynamics of Weather Regimes: Quasi-Stationary Waves and Blocking, Mon. Weather Rev., 110, 1105–1145, https://doi.org/10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2, 1982.

Sandri, M.: Numerical Calculation of Lyapunov Exponents, Math. J., 6, 78–84, 1996.

Sanz-Alonso, D. and Waniorek, N.: Long-Time Accuracy of Ensemble Kalman Filters for Chaotic Dynamical Systems and Machine-Learned Dynamical Systems, SIAM J. Appl. Dyn. Syst., pp. 2246–2286, https://doi.org/10.1137/24M1719232, 2025.

Takeda, K. and Sakajo, T.: Uniform Error Bounds of the Ensemble Transform Kalman Filter for Chaotic Dynamics with Multiplicative Covariance Inflation, SIAM/ASA J. Uncertainty Quantification, 12, 1315–1335, https://doi.org/10.1137/24M1637192, 2024.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble Square Root Filters, Mon. Weather Rev., 131, 1485–1490, 2003.

Trevisan, A. and Palatella, L.: On the Kalman Filter Error Covariance Collapse into the Unstable Subspace, Nonlinear Process. Geophys., 18, 243–250, https://doi.org/10.5194/npg-18-243-2011, 2011.

Trevisan, A. and Uboldi, F.: Assimilation of Standard and Targeted Observations within the Unstable Subspace of the Observation–Analysis–Forecast Cycle System, J. Atmospheric Sci., 61, 103–113, https://doi.org/10.1175/1520-0469(2004)061<0103:AOSATO>2.0.CO;2, 2004.

Trevisan, A., D'Isidoro, M., and Talagrand, O.: Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension, Q. J. R. Meteorol. Soc., 136, 487–496, https://doi.org/10.1002/qj.571, 2010.

von Bremen, H. F., Udwadia, F. E., and Proskurowski, W.: An Efficient QR Based Method for the Computation of Lyapunov Exponents, Physica D: Nonlinear Phenomena, 101, 1–16, https://doi.org/10.1016/S0167-2789(96)00216-3, 1997.