



# Quantifying the minimum ensemble size for asymptotic accuracy of the ensemble Kalman filter using the degrees of instability

Kota Takeda<sup>1,2</sup> and Takemasa Miyoshi<sup>2,3</sup>

<sup>1</sup>Department of Applied Physics, Graduate School of Engineering, Nagoya University, Nagoya, Japan

<sup>2</sup>RIKEN Center for Computational Science, Kobe, Japan

<sup>3</sup>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. In this study, we define filter accuracy based on its time-asymptotic performance relative to the observation noise. We then investigate the minimum ensemble size,  $m^*$ , required to achieve this accuracy, linking it to the degrees of instability in the chaotic dynamics.

5 Since the well-defined characteristic numbers of dynamical systems called the Lyapunov exponents (LEs) quantify the time-asymptotic exponential growth or decay rates of infinitesimal perturbations, we define the degrees of instability  $N_+$  by the number of positive LEs. In the EnKF, capturing such instabilities with limited ensemble is crucial for achieving long-term filter accuracy. Therefore, we propose an ensemble spin-up and downsizing method within data assimilation cycles. Numerical experiments applying the EnKF to the Lorenz 96 model show that the minimum ensemble size required for filter accuracy  
10 is estimated by  $m^* = N_+ + 1$ . This study provides a practical estimate for the minimum ensemble size based on a priori information about the target dynamics, along with a method to achieve long-term accuracy.

## 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  
15 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 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 (Eckmann and Ruelle, 1985). For continuous-time dynamical systems, such as ordinary differential equations, one of the LEs is zero, corresponding  
20 to perturbations parallel to the vector field. 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, 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  $N_0 \geq N_+$ .



25 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) 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) (Bishop et al., 2001). The ensemble covariance,  $C_t$ ,  
characterizes forecast uncertainty through its eigenpairs, with the eigenvalues quantifying the magnitude of variability and the  
30 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  $C_t$  is less than the  
ensemble size  $m$ , i.e.,  $\text{rank}(C_t) \leq m - 1$ . 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.  
35 This phenomenon is called filter divergence and must be avoided. To mitigate this problem, covariance inflation techniques  
artificially increase the ensemble spread to compensate for the underestimation of uncertainty, thereby helping to prevent filter  
divergence.

Mathematical studies often focus on the long-term behavior of the analysis error. The key objective is to demonstrate that  
the mean squared error remains of order  $r^2$  when the observation noise level  $r$  is sufficiently small compared with the attractor  
40 size. This property is referred to as filter accuracy. Establishing filter accuracy ensures that filter divergence does not occur.  
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 high-dimensional applications, it is important to identify a more relaxed lower bound  $m \geq m^*$  depending  
on the system. Under more idealized assumptions, González-Tokman and Hunt (2013) investigated the lower bound of  $m$  for  
45 the ETKF. They proved that for discrete-time dynamical systems, if  $m \geq N_+$ , 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 (Trevisan and Ubaldi, 2004; Trevisan et al.,  
50 2010; Trevisan and Palatella, 2011; Bocquet et al., 2017; Bocquet and Carrassi, 2017) investigated the use of the unstable  
subspace in data assimilation, known as Assimilation in Unstable Subspace (AUS). These works mainly consider systems with  
zero LEs and argue that correcting the state in the  $N_0$ -dimensional unstable-neutral subspace is crucial for filter performance.  
We review results directly related to the ETKF. Theoretical analyses for linear systems in (Bocquet et al., 2017; Bocquet and  
Carrassi, 2017) suggest that the rank of the ETKF covariance is asymptotically bounded by  $N_+$  due to the exponential decay  
55 of uncertainty in the stable subspace and the slower decay in the neutral subspace under some conditions. They also present  
numerical results for the ETKF applied to 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 time-averaged analysis error  
remains small within the finite-time interval. Similar findings are reported in Carrassi et al. (2022) using the Quasi-Geostrophic



60 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 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 long-time limit. Hence, we conjecture that the minimum ensemble size for  
65 asymptotic accuracy of the ETKF is

$$m^* = N_+ + 1, \quad (1)$$

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.,  $N_0 = N_+ + 1$ ). In general, it holds that  $N_+ + 1 \leq N_0$   
for continuous-time dynamical systems. For this case, we still conjecture that the minimum ensemble size is  $m^* = N_+ + 1$ .  
70 Therefore, we use  $N_+$  rather than  $N_0$  to represent the minimum ensemble size throughout the paper. To verify our conjecture,  
we conduct numerical experiments with the ETKF applied to the Lorenz 96 model with 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  
75 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 ensemble, with its mean close to the true state and its perturbations aligned with  
the unstable subspace. Although our target model 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 from  
this perspective are important because they define and clarify the ensemble size below which filter divergence of the ETKF  
80 cannot be avoided, even when accurate observations and appropriate inflation are used. Moreover, the error bound by order of  
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.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the basics of Lyapunov analysis. In Sect. 3,  
85 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 Bocquet and Carrassi (2017) in estimating the minimum ensemble  
size, highlighting differences in objectives.



## 2 Characterizing the degrees of instability in dynamics

### 90 2.1 The Lyapunov exponents and their computation

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

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)), \quad t > 0 \quad (2)$$

with  $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^{N_x}$ , where  $\mathbf{f} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$  is a smooth vector field. To study the instability of the dynamics, we examine the evolution of a perturbation  $\delta\mathbf{x}(t) \in \mathbb{R}^{N_x}$ , defined as the difference between two trajectories separated by  $\delta\mathbf{x}_0 \in \mathbb{R}^{N_x}$  at

95  $t = 0$ . Assuming  $\delta\mathbf{x}(t)$  is sufficiently small and smooth, its evolution is approximated by the linearization of Eq. (2):

$$\frac{d}{dt}\delta\mathbf{x}(t) = J_{\mathbf{f}}(\mathbf{x}(t))\delta\mathbf{x}(t), \quad t > 0 \quad (3)$$

with  $\delta\mathbf{x}(0) = \delta\mathbf{x}_0$ , where  $J_{\mathbf{f}}(\mathbf{x}(t))$  denotes the Jacobian matrix of  $\mathbf{f}$  at  $\mathbf{x} = \mathbf{x}(t)$ . Eq. (3) is referred to as the tangent linear model. Let  $\Phi(t, \mathbf{x}_0) \in \mathbb{R}^{N_x \times N_x}$  denote the fundamental matrix solution to Eq. (3) with  $\Phi(0, \mathbf{x}_0) = I_{N_x}$ . The unique solution is then

$$100 \quad \delta\mathbf{x}(t) = \Phi(t, \mathbf{x}_0)\delta\mathbf{x}_0, \quad t \geq 0. \quad (4)$$

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

$$\lambda_j(t, \mathbf{x}_0) = \frac{1}{t} \log \sigma_j(\Phi(t, \mathbf{x}_0)) \in \mathbb{R}, \quad j = 1, \dots, N_x. \quad (5)$$

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

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

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

$$\lambda_j(\mathbf{x}_0) = \lim_{t \rightarrow \infty} \lambda_j(t, \mathbf{x}_0) \in \mathbb{R}, \quad j = 1, \dots, N_x, \quad (6)$$

110 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. (2) is ergodic, the LEs are uniquely determined regardless of  $\mathbf{x}_0$  in an invariant subset of  $\mathbb{R}^{N_x}$ . For continuous-time dynamics of the form Eq. (2), one exponent is always zero,  $\lambda_j = 0$ , corresponding to a perturbation parallel to the vector field  $\delta\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t))$ . 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  
 115 chaotic. According to Sect. 1, we define the following dimension to quantify the degrees of freedom of unstable perturbations:

$$N_+ = \#\{j \in \{1, \dots, N_x\} \mid \lambda_j > 0\}. \quad (7)$$



See (Kuptsov and Parlitz, 2012; Carrassi et al., 2022) 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  $\mathbf{f} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$ , its Jacobian  $J_{\mathbf{f}} : \mathbb{R}^{N_x \times N_x}$ , an initial state  $\mathbf{x}_0 \in \mathbb{R}^{N_x}$ , 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:**  $\mathbf{f}, J_{\mathbf{f}}, \mathbf{x}_0, \text{IntegrateODE}, \Delta t, n$   
**Ensure:**  $S = (\mathbf{x}, V), F(S) = (\mathbf{f}(\mathbf{x}), J_{\mathbf{f}}(\mathbf{x})V)$

- 1:  $S \leftarrow (\mathbf{x}_0, I_{N_x})$
- 2:  $LE \leftarrow \mathbf{0} \in \mathbb{R}^{N_x}$
- 3: **for**  $i = 1$  **to**  $n$  **do**
- 4:    $S \leftarrow \text{IntegrateODE}(F, S, \Delta t)$
- 5:    $Q, R \leftarrow \text{QR}(S)$
- 6:    $S \leftarrow Q$
- 7:    $LE \leftarrow LE + \log(\text{diag}(|R|))$
- 8: **end for**
- 9: **return**  $LE/(n\Delta t)$

---

## 2.2 The Lorenz 96 model

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

$$\frac{dx^i}{dt} = (x^{i+1} - x^{i-2})x^{i-1} - x^i + F, \quad i = 1, \dots, N_x \quad (8)$$

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 data assimilation algorithms. We use this model to show examples of chaotic dynamics with various degrees of instability by changing the parameter  $F$ .

## 130 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. (2) with noisy observations. Let  $t_n = n\tau$ ,  $n = 0, 1, 2, \dots$ , denote the observation times with a fixed interval  $\tau > 0$ . We define the flow map  $\Psi_\tau : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$  such that  $\Psi_\tau(\mathbf{x}_0) = \mathbf{x}(\tau)$ ,

where  $\mathbf{x}(t)$  is the solution to Eq. (2) with  $\mathbf{x}(0) = \mathbf{x}_0$ . This yields the discrete-time dynamical system

$$135 \quad \mathbf{x}_n = \Psi(\mathbf{x}_{n-1}), \quad n = 1, 2, \dots, \quad (9)$$

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

$$\mathbf{y}_n = H \mathbf{x}_n + \boldsymbol{\eta}_n, \quad n = 1, 2, \dots, \quad (10)$$

where  $H \in \mathbb{R}^{N_y \times N_x}$  is the observation matrix, and  $\boldsymbol{\eta}_n \sim \mathcal{N}(0, R)$  is a Gaussian observation noise with a symmetric positive definite covariance matrix  $R \in \mathbb{R}^{N_y \times N_y}$ . To estimate the state  $\mathbf{x}_n$  from the observations  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ , we employ the ensemble  
 140 Kalman filter (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

$$\mathbf{x}_n^{f(k)} = \Psi(\mathbf{x}_{n-1}^{a(k)}), \quad k = 1, \dots, m, \quad (11)$$

where  $m \in \mathbb{N}$  is the ensemble size, and the superscripts  $f$  and  $a$  denote forecast and analysis, respectively. In the analysis step,  
 145 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) (Bishop et al., 2001). The ETKF updates the mean and perturbation part of the ensemble as

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + K_n (\mathbf{y}_n - H \bar{\mathbf{x}}_n^f), \quad (12)$$

$$V_n^a = V_n^f T_n, \quad (13)$$

150 where  $\bar{\mathbf{x}}_n^f = \frac{1}{m} \sum_{k=1}^m \mathbf{x}_n^{f(k)}$ ,  $V_n^f = [\mathbf{x}_n^{f(1)} - \bar{\mathbf{x}}_n^f, \dots, \mathbf{x}_n^{f(m)} - \bar{\mathbf{x}}_n^f] \in \mathbb{R}^{N_x \times m}$ ,  $K_n = C_n^f H^\top (H C_n^f H^\top + R)^{-1}$  is the Kalman gain,  $C_n^f = V_n^f (V_n^f)^\top / (m - 1)$  is the forecast covariance, and  $T_n \in \mathbb{R}^{m \times m}$  is a transform matrix defined as

$$T_n = \left( I_m + \frac{1}{m-1} (V_n^f)^\top H^\top R^{-1} H V_n^f \right)^{-1/2}, \quad (14)$$

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

$$155 \quad \mathbf{x}_n^{a(k)} = \bar{\mathbf{x}}_n^a + \mathbf{v}_n^{a(k)}, \quad k = 1, \dots, m, \quad (15)$$

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

As mentioned in Sect. 1, the forecast ensemble is corrected more strongly in directions with higher uncertainty, as represented by the forecast covariance  $C_n^f$ . However, the rank of  $C_n^f$  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  
 160 these issues, we employ multiplicative covariance inflation:

$$V_n^f \leftarrow \alpha V_n^f, \quad (16)$$



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

We define filter accuracy as follows. Assume  $R = r^2 I_{N_y}$  with small  $r > 0$ . The EnKF achieves filter accuracy if there exists a constant  $c > 0$ , independent of  $r$ , such that

$$\limsup_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{x}_n - \bar{\mathbf{x}}_n^a\|^2] \leq cr^2 \quad (17)$$

for sufficiently small  $r$ , where the expectation is taken over the observation noise and the initial ensemble. This property guarantees that the squared analysis error remains of order  $r^2$ .

### 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,  $n = N_{\text{spinup}}$ . We call the period before  $n = N_{\text{spinup}}$  the ensemble spin-up period. In the ensemble downsizing method, we apply the singular value decomposition (SVD) to the ensemble perturbation,  $V \in \mathbb{R}^{N_x \times m_0}$ , and retain only the leading  $m$  modes. This procedure is detailed in Algorithm 2.

---

**Algorithm 2** The ensemble downsizing method by the singular value decomposition

---

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

**Ensure:**  $X = \bar{\mathbf{x}} + V$

- 1:  $U, \Sigma, _- \leftarrow \text{SVD}(V)$
  - 2:  $V \leftarrow U[:, 1:m] \Sigma[1:m, 1:m]$
  - 3: **return**  $\bar{\mathbf{x}} + V$
- 

The resulting ETKF with the multiplicative covariance inflation Eq. (16) 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 accuracy of the ETKF 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  $R = r^2 I_{N_x}$ , 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  $F = 16$ . 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  $X_0$  is defined as  $X_0 = (\mathbf{x}_0^{(k)})_{k=1}^{m_0}$  with  $\mathbf{x}_0^{(k)} \sim N(\mathbf{x}_0, 25I_{N_x})$ , where  $\mathbf{x}_0 \in \mathbb{R}^{N_x}$  is uniformly sampled from the true trajectory.

To evaluate the filter accuracy of the ETKF, we use the squared error (SE) as in Eq. (17). To approximate the expectation  $\mathbb{E}$ , we compute parallel simulations for  $n_{\text{seeds}}$  random seeds to generate the observation noises. Then, we take the maximum after




---

**Algorithm 3** The ETKF with multiplicative covariance inflation and ensemble downsizing

---

**Require:**  $\Psi, H, R, (\mathbf{y}_n)_{n=1}^N, X_0 \in \mathbb{R}^{N_x \times m_0}, \alpha > 1, N_{\text{spinup}} < N, m < m_0,$

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

```

1:  $X \leftarrow X_0$ 
2:  $m' \leftarrow m_0$ 
3: for  $n = 1$  to  $N$  do
4:   # Forecast step
5:   for  $k = 1$  to  $m'$  do
6:      $\mathbf{x}^{f(k)} \leftarrow \Psi(\mathbf{x}^{a(k)})$ 
7:   end for
8:    $\bar{\mathbf{x}}^f \leftarrow \frac{1}{m'} \sum_{k=1}^{m'} \mathbf{x}^{f(k)}$ 
9:    $V^f \leftarrow [\mathbf{x}^{f(1)} - \bar{\mathbf{x}}^f, \dots, \mathbf{x}^{f(m')} - \bar{\mathbf{x}}^f]$ 
10:  # Covariance inflation
11:   $V^f \leftarrow \alpha V^f$ 
12:   $C^f \leftarrow V^f (V^f)^\top / (m' - 1)$ 
13:  # Analysis step
14:   $K \leftarrow C^f H^\top (H C^f H^\top + R)^{-1}$ 
15:   $\bar{\mathbf{x}}^a \leftarrow \bar{\mathbf{x}}^f + K(\mathbf{y}_n - H \bar{\mathbf{x}}^f)$ 
16:   $T \leftarrow (I_{m'} + \frac{1}{m'-1} (V^f)^\top H^\top R^{-1} H V^f)^{-1/2}$ 
17:   $V^a \leftarrow V^f T$ 
18:   $X \leftarrow \bar{\mathbf{x}}^a + V^a$ 
19:   $X_n \leftarrow X$ 
20:  # Ensemble downsizing
21:  if  $n = N_{\text{spinup}}$  then
22:     $X \leftarrow \text{EnsembleDownsizing}(X, m)$ 
23:     $m' \leftarrow m$ 
24:  end if
25: end for
26: return  $(X_n)_{n=1}^N$ 

```

---





**Table 1.** Common parameters for the ETKF experiments.

| Parameter           | Value                                    | Description                                 |
|---------------------|------------------------------------------|---------------------------------------------|
| $\Delta t$          | 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             |
| $\alpha$            | 1.0, 1.1, 1.2, 1.3, 1.4, 1.5             | Inflation factor                            |
| $r$                 | $10^0, 10^{-1}, \dots, 10^{-4}$          | Standard deviation of the observation noise |
| $n_{\text{obs}}$    | 5 (for $F = 8$ ), 2 (for $F = 16$ )      | Observation interval (integration steps)    |
| $N_{\text{spinup}}$ | 720 (for $F = 8$ ), 1800 (for $F = 16$ ) | Spin-up period (assimilation steps)         |

$n = N_\infty$  to approximate  $\limsup_{n \rightarrow \infty}$ . This leads to

$$\limsup_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{x}_n - \bar{\mathbf{x}}_n^a\|^2] \approx \max_{n \geq N_\infty} \frac{1}{n_{\text{seeds}}} \sum_{i=1}^{n_{\text{seeds}}} \|\mathbf{x}_n - \bar{\mathbf{x}}_n^a(\omega_i)\|^2, \quad (18)$$

where  $\bar{\mathbf{x}}_n^a(\omega_i)$  is the analysis mean of a sample path with the  $i$ -th random seed. We use  $N_\infty = N/2$  and  $n_{\text{seeds}} = 10$  to approximate Eq. (18). On the other hand, we use the root mean squared error (RMSE) to visualize the time series of the analysis error for a sample run.

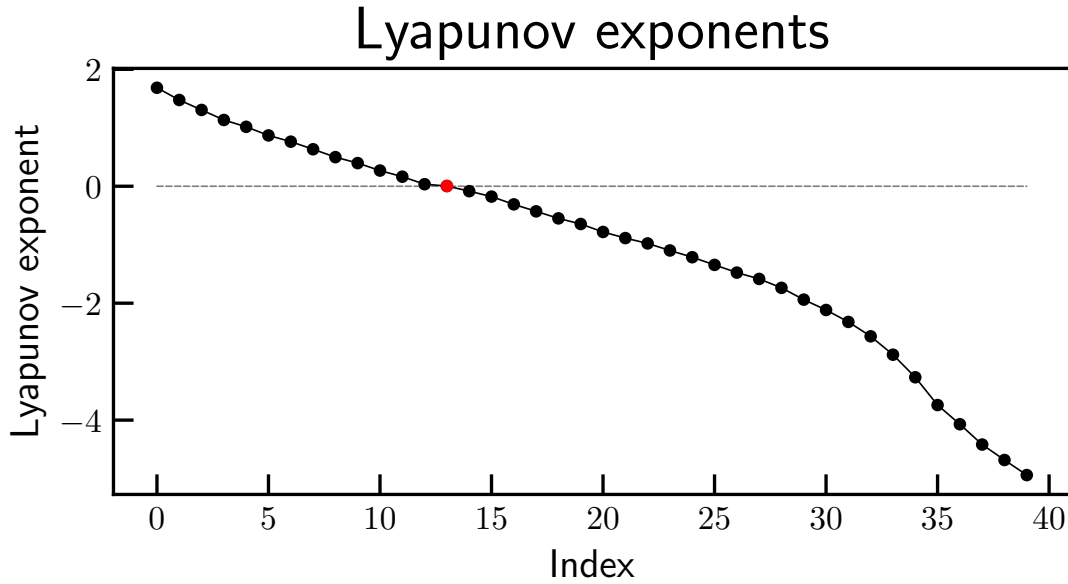
#### 4.1 $F = 8$

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 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  $\lambda_1 \approx 1.67$ .

195 In this section, we assimilate observations every  $n_{\text{obs}} = 5$  integration steps. We reduce the ensemble size after  $N_{\text{spinup}} = 720$  assimilation steps. For each pair  $(r, m)$ , we vary the inflation factor  $\alpha$  to find the optimal value that minimizes the SE defined in Eq. (18). The results are shown in Figure 2 with log-log plots of the SE 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.

200 In the following two experiments, we fix the number of integration steps  $N = 72,000$  and single random seed to generate the observation noise. We then conduct an experiment to investigate the dependence of the RMSE on the spin-up period  $N_{\text{spinup}}$ . We set  $r = 10^{-1}$  and  $m = 15$ , which avoids filter divergence in the experiment for Figure 2. For  $N_{\text{spinup}} = 0$  and 720, we show the time series of the RMSE with various  $\alpha$  in Figure 3. From Figure 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 period when  $N_{\text{spinup}} = 0$ . This suggests that filter stability is achieved even without an ensemble spin-up period

205



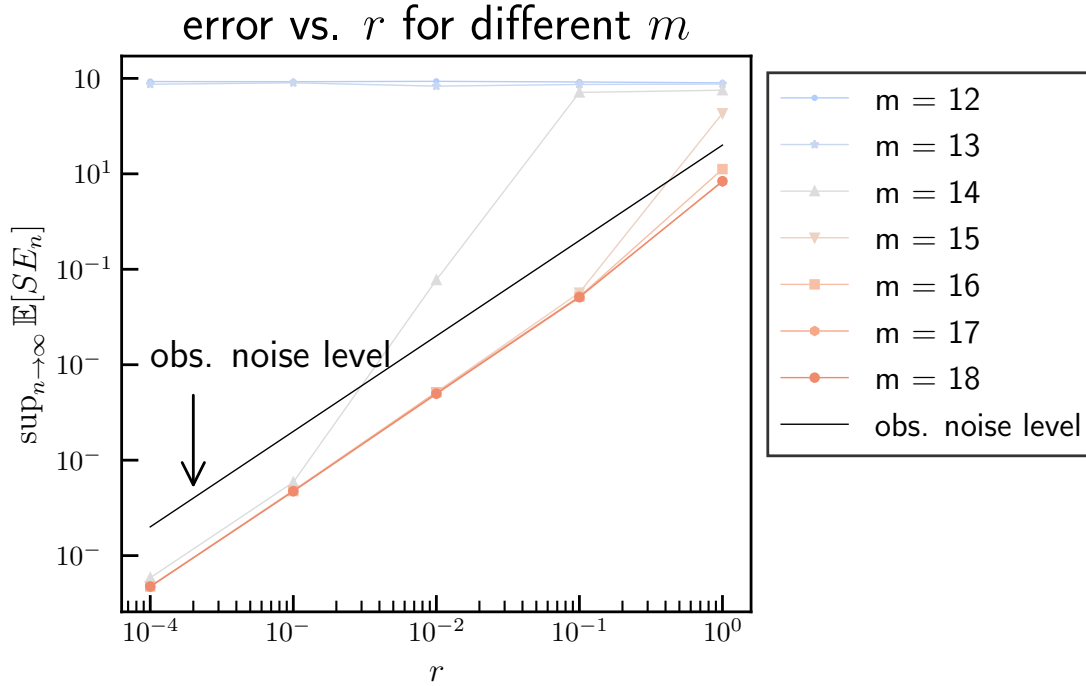
**Figure 1.** The LEs of the Lorenz 96 model with  $(J, F) = (40, 8)$ . The zero exponent  $\lambda_{14} = 0$  is indicated in red.

if a sufficiently large inflation factor is used. In addition, we find that the ensemble spin-up period can reduce the required inflation factor for filter accuracy.

Figure 4 shows the time series of the RMSE with  $r = 10^{-4}$ ,  $N_{\text{spinup}} = 0$  and various  $\alpha$  for  $m = 14$  (a) and  $m = 13$  (b). In Figure 4 (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. 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 4 (b), all RMSE values do not decay and remain large. These results indicate that the minimum ensemble size for filter accuracy is  $m^* = N_+ + 1 = 14$  regardless of the ensemble spin-up period.

## 4.2 $F = 16$

We set  $F = 16$  and compute the LEs as in the previous section, shown in Figure 5. This yields  $N_+ = 15$  and the largest LE  $\lambda_1 \approx 3.82$ . We assimilate observations every  $n_{\text{obs}} = 2$  integration steps in this section. This value of  $n_{\text{obs}}$  is the largest integer with which the approximated error expansion  $n_{\text{obs}}\lambda_1\Delta t$  in the forecast step for  $F = 16$  does not exceed that with  $n_{\text{obs}} = 5$  for



**Figure 2.** 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.

$F = 8$ . Indeed, if we write these quantities for  $F$  as  $n_{\text{obs}}^{(F)}$  and  $\lambda_1^{(F)}$ , the error expansion with each  $F$  is approximated as

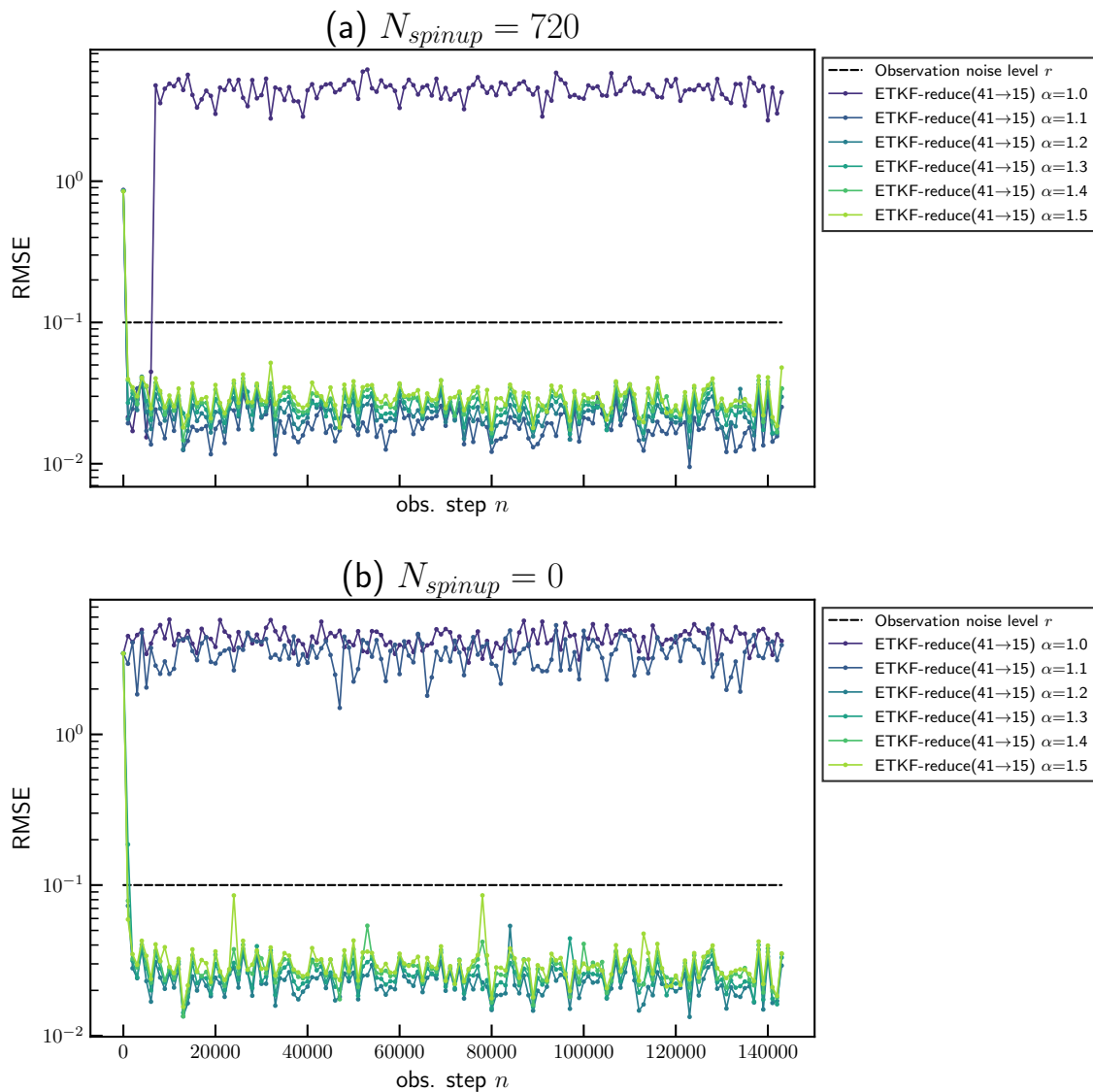
$$n_{\text{obs}}^{(8)} \lambda_1^{(8)} \Delta t \approx 5 \cdot 1.67 \cdot 0.01 = 8.35,$$

$$220 \quad n_{\text{obs}}^{(16)} \lambda_1^{(16)} \Delta t \approx 2 \cdot 3.82 \cdot 0.01 = 7.64.$$

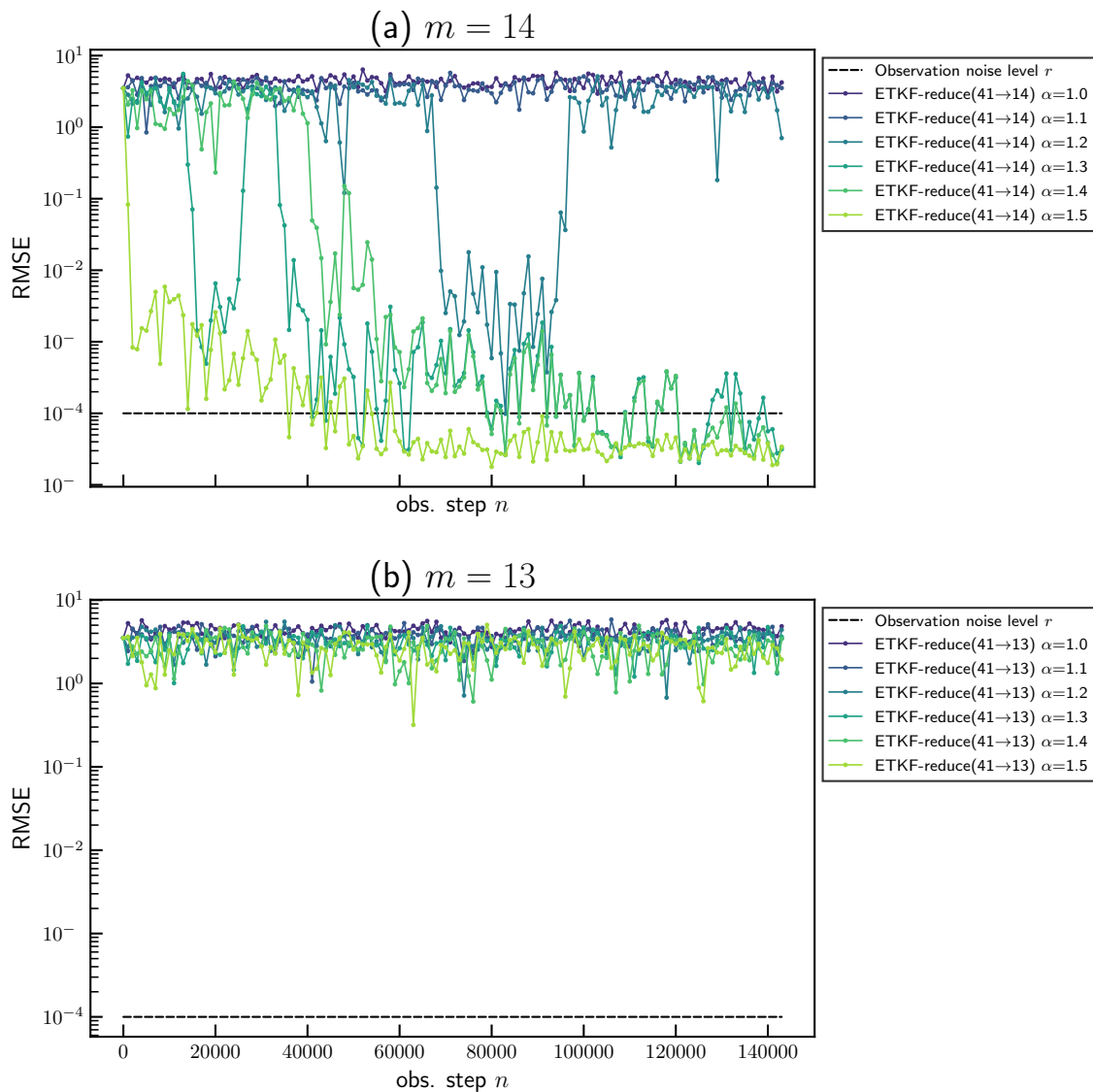
We compute the SE in the same manner as in the previous section with  $N = 72,000$  integration steps and  $N_{\text{spinup}} = 1800$  assimilation steps, which yields the same integration steps before the ensemble downsizing method. The dependence of the SE on  $r$  for different  $m$  is shown in Figure 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  $m^* = N_+ + 1 = 16$ .

## 225 5 Conclusions

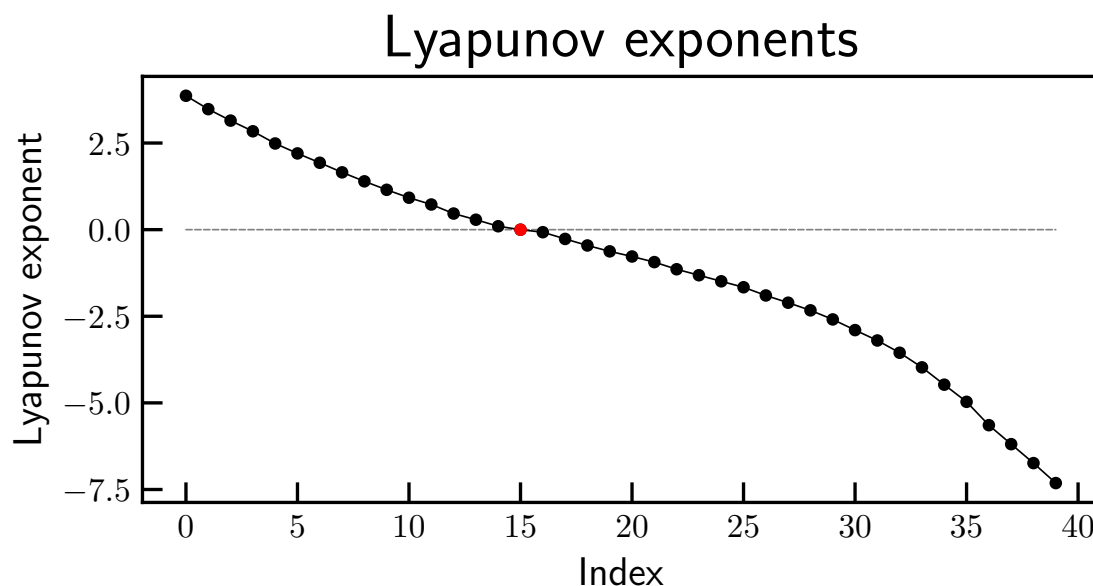
We proposed an ensemble downsizing method for the EnKF to generate an ensemble aligned with the unstable subspace of the dynamics. Through numerical experiments with the ETKF applied to the Lorenz 96 model, we verified our conjecture that



**Figure 3.** The time series of the RMSE with  $r = 10^{-1}$ ,  $m = 15$  and various  $\alpha = 1.0, \dots, 1.5$  for  $N_{spinup} = 720$  (a) and  $N_{spinup} = 0$  (b). The dashed line indicates the level  $r$ .



**Figure 4.** The time series of the RMSE with  $r = 10^{-4}$ ,  $N_{\text{spinup}} = 0$  and various  $\alpha = 1.0, \dots, 1.5$  for  $m = 14$  (a) and  $m = 13$  (b). The dashed line indicates the level  $r$ .

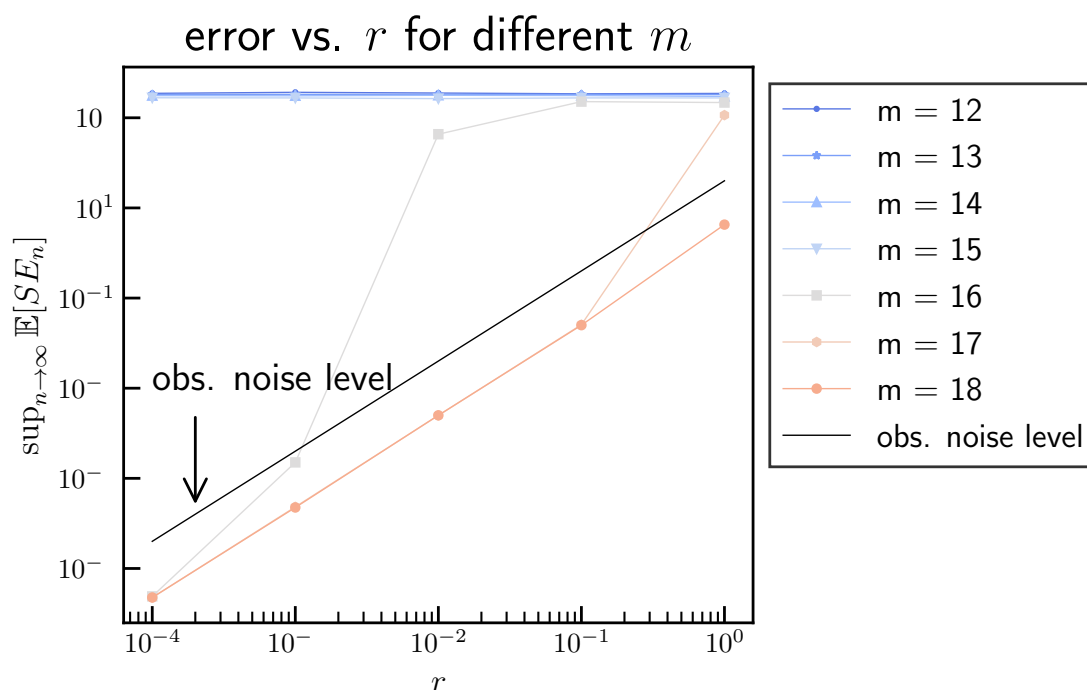


**Figure 5.** The LEs of the Lorenz 96 model with  $(J, F) = (40, 16)$ . The zero exponent  $\lambda_{16} = 0$  is indicated in red.

the minimum ensemble size for asymptotic accuracy is  $m^* = N_+ + 1$ , where  $N_+$  is the number of positive LEs (Figure 1, 5). This estimate of  $m^*$  is valid for the multiple external forcing  $F$  in the Lorenz 96 model (Figures 2 and 6), and the filter  
 230 remains stable over long integration periods (Figures 3 and 4). Moreover, filter accuracy is achieved even without an ensemble spin-up period (Figure 4). 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 factor for filter accuracy. 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  
 235 needed to verify whether the estimate  $m^* = N_+ + 1$  holds in such cases. Suitable dynamical systems for this purpose include 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).

. Code availability

The code is available at [https://github.com/KotaTakeda/enkf\\_ensemble\\_downsizing/releases/tag/v1.0.0](https://github.com/KotaTakeda/enkf_ensemble_downsizing/releases/tag/v1.0.0)  
 240 and archived on Zenodo: <https://doi.org/10.5281/zenodo.17319854>.



**Figure 6.** 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  $F = 16$  ( $N_+ = 15$ ) is used.

. Author Contribution

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

. Competing Interests

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

245 . 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  
250 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  
 255 Aspects, *Mon. Weather Rev.*, 129, 420–436, [https://doi.org/10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2), 2001.
- Bocquet, M. and Carrassi, A.: Four-Dimensional Ensemble Variational Data Assimilation and the Unstable Subspace, *Tellus Dyn. Meteorol. Oceanogr.*, 69, 1304–1304, <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,  
 260 <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](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  
 265 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.
- 270 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.
- Kalnay, E.: *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, New York,  
<https://doi.org/10.1017/CBO9780511802270>, 2002.
- Kuptsov, P. V. and Parlitz, U.: Theory and Computation of Covariant Lyapunov Vectors, *J Nonlinear Sci*, 22, 727–762,  
 275 <https://doi.org/10.1007/s00332-012-9126-5>, 2012.
- 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,  
 280 1105–1145, [https://doi.org/10.1175/1520-0493\(1982\)110<1105:DOWRQS>2.0.CO;2](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.
- 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,  
 285 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 Ubaldi, 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](https://doi.org/10.1175/1520-0469(2004)061<0103:AOSATO>2.0.CO;2), 2004.
- 290 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., Udvardi, 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](https://doi.org/10.1016/S0167-2789(96)00216-3), 1997.