A Local Particle Filter and Its Gaussian Mixture Extension Implemented with Minor Modifications to the LETKF

Shunji Kotsuki[1,2,3], Takemasa Miyoshi[1,4,5,6,7], Keiichi Kondo[8,1], and Roland Potthast[9,10]

[1]RIKEN Center for Computational Science, Kobe, Japan
[2]Center for Environmental Remote Sensing, Chiba University, Chiba, Japan
[3]PRESTO, Japan Science and Technology Agency, Chiba, Japan
[4]RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, Kobe, Japan
[5]RIKEN Cluster for Pioneering Research, Kobe, Japan
[6]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[7]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA
[8]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan
[9]Deutscher Wetterdienst, Offenbach, Germany
[10]Applied Mathematics, University of Reading, UK

Correspondence to: Shunji Kotsuki (shunji.kotsuki@chiba-u.jp) and Takemasa Miyoshi (takemasa.miyoshi@riken.jp)

Abstract. A particle filter (PF) is an ensemble data assimilation method that does not assume Gaussian error distributions. Recent studies proposed local PFs (LPFs), which use localization as in the ensemble Kalman filter, to apply the PF for high-dimensional dynamics efficiently. Among others, Penny and Miyoshi developed an LPF in the form of the ensemble transform matrix of the Local Ensemble Transform Kalman Filter (LETKF). The LETKF has been widely accepted for various geophysical systems including numerical weather prediction (NWP) models. Therefore, implementing the LPF consistently with an existing LETKF code is useful.

This study develops a software platform for the LPF and its Gaussian mixture extension (LPFGM) by making slight modifications to the LETKF code with a simplified global climate model known as Simplified Parameterizations, Primitive Equation Dynamics (SPEEDY). A series of idealized twin experiments were accomplished under the ideal model assumption. With large inflation by the relaxation to prior spread, the LPF showed stable filter performance with dense observations but became unstable with sparse observations. The LPFGM showed more accurate and stable performance than the LPF with both dense and sparse observations. In addition to the relaxation parameter, regulating the resampling frequency and the amplitude of Gaussian kernels was important for the LPFGM. With a spatially inhomogeneous observing network, the LPFGM was superior to the LETKF in sparsely observed regions where the background ensemble spread and non-Gaussianity are larger. The SPEEDY-based LETKF, LPF, and LPFGM systems are available as open-source software on Github (https://github.com/skotsuki/speedy-lpf) and can be adapted to various models relatively easily like the LETKF.

Short Summary. Data assimilation takes an important part in numerical weather prediction (NWP) to combine forecasted states and observations. While data assimilation methods in NWP usually assume the Gaussian error distribution, some

variables in the atmosphere are known to have non-Gaussian error statistics such as precipitation. This study extended a widely used ensemble data assimilation algorithm for enabling to assimilate more non-Gaussian observations.

1 Introduction

Ensemble-based data assimilation (DA) has been broadly applied in geoscience fields such as weather and ocean prediction. The ensemble Kalman filter (EnKF) has been intensely investigated for the past two decades, such as for the perturbed observation method (Evensen, 1994; Burgers et al., 1998; Houtekamer and Mitchel, 1998; van Leeuwen 2020) and ensemble square root filters (e.g., Anderson, 2001; Bishop, 2001; Whitaker and Hamill, 2002; Hunt et al., 2007). The EnKF has advantages in flow-dependent error estimates represented by an ensemble and in its relative ease of implementation to nonlinear dynamical systems such as numerical weather prediction (NWP) models. The degrees of freedom of dynamical models (e.g., $> O(10^8)$ for NWP models) are much larger than the typically affordable ensemble size ($< O(10^3)$). On the other hand, atmospheric and oceanic models show local low dimensionality (Patil et al., 2001; Oczkowski et al. 2005), and practical EnKF implementations use a localization technique that limits the impact of distant observations while also reducing the effective degrees of freedom. Among various kinds of EnKFs, the local ensemble transform Kalman filter (LETKF; Hunt et al., 2007) is widely utilized in operational NWP centers in the same manner with the Deutscher Wetterdienst (DWD) and the Japan Meteorological Agency (JMA). Analysis updates of the LETKF are performed by multiplying the ensemble transform matrix to the prior ensemble perturbation matrix, following the ensemble transform Kalman filter (ETKF, Bishop et al. 2001; Wang et al. 2004).

The particle filter (PF) is another ensemble DA method broadly applicable to nonlinear and non-Gaussian problems (cf. van Leeuwen 2009 and van Leeuwen et al. 2019 for reviews on geoscience applications). The PF potentially solves some issues in the basic assumptions of the EnKF by permitting nonlinear observation operators and non-Gaussian likelihood functions (Penny and Miyoshi, 2016). For example, assimilating precipitation observations with a standard EnKF is difficult partly because of their non-Gaussian errors (Lien et al., 2013, 2016; Kotsuki et al., 2017a). The PF can treat such variables with non-Gaussian errors properly. Several PF methods have been explored for low-dimensional problems in early studies (Gordon et al. 1993; van Leeuwen et al. 2009). However, applying the PF to high-dimensional dynamical systems is generally difficult because the number of particles or the ensemble size must to be increased exponentially with the system size to avoid a weight collapse in which very few particles occupy most of the weights (Snyder et al., 2008, 2015). If a weight collapse occurs, the PF loses the diversity of the particles after resampling. Therefore, we need to assure that the weights are similarly distributed among particles. Previous studies developed the equivalent-weights particle filter (EWPF; van Leeuwen et al. 2010; Ades and van Leeuwen 2013, 2015; Zhu et al. 2016) to extend time until a weight collapse by using the proposal density to drive the particles toward the high-probability region of the posterior. Even PFs with proposal densities, however, are unable to prevent the weights from collapsing.

Alternatively, the local particle filter (LPF) uses localization to avoid a weight collapse by limiting the impact of observations within a local domain. Localization is a well-adopted method for the EnKF to treat sampling errors due to a limited ensemble size. Several LPFs have been proposed to apply the PF efficiently to high-dimensional systems (e.g., Bengston et al., 2003; van Leeuwen, 2009; Poterjoy, 2016; Penny and Miyoshi, 2016; Poterjoy and Anderson, 2016; Farchi and Bocquet, 2018; van Leeuwen et al. 2019). Among them, Penny and Miyoshi (2016) developed an LPF by means of the ensemble transform matrix of the LETKF. The LETKF has been commonly used for diverse geophysical systems, and consistent implementation with an available LETKF code would be useful. The DWD implemented the LPF in an operational global model based on the operational LETKF code and described stable performance in the operational setting (Potthast et al., 2019). Walter and Potthast (2022; hereafter WP22) extended the LPF to the Gaussian mixture for further improvements.

The goal of this study is to develop and provide a software platform to accelerate research on LPF and its Gaussian mixture extension (hereafter, LPFGM). Although many studies on LPF have used simple idealized chaotic models, such as the 40-variable Lorenz model (hereafter, "L96"; Lorenz, 1996; Lorenz and Emanuel, 1998), the present study uses a simplified atmospheric general circulation model defined as the Simplified Parameterizations, Primitive Equation Dynamics (SPEEDY) model (Molteni, 2003). Miyoshi (2005) coupled the LETKF with the SPEEDY model for the first time. This study extends the SPEEDY-LETKF code and implements the LPF and LPFGM. As demonstrated below, the LPF and LPFGM can be implemented easily with simple modifications to the existing LETKF code.

The remainder of this paper is organized as follows. Section 2 explains the mathematical formulation of the LETKF, LPF, and LPFGM, followed by the description of specific modifications made to the existing LETKF code. Section 3 presents the experimental settings. Section 4 outlines the results and discussions. Lastly, section 5 provides a summary.

2 Methodology

2.1 Local ensemble transform Kalman filter

Hunt et al. (2007) introduced the LETKF as a computationally efficient EnKF by combining the local ensemble Kalman filter (Ott et al., 2004) and ETKF (Bishop et al., 2001). Let $\mathbf{X}_t$ be a matrix composed of m ensemble state vectors. The ensemble mean vector and perturbation matrix of $\mathbf{X}_t$ are given by $\bar{\mathbf{x}}_t$ $(\in \mathbb{R}^n)$ and $\mathbf{Z}_t \equiv \left\{ \mathbf{x}_t^{(1)} - \bar{\mathbf{x}}_t, \ldots, \mathbf{x}_t^{(m)} - \bar{\mathbf{x}}_t \right\} (\in \mathbb{R}^{n \times m})$, respectively, where n is the system size. The subscript t indicates the time, and the superscript (i) denotes the ith ensemble member. The analysis equations of the LETKF are specified by:

$$\mathbf{X}_{t,LETKF}^a = \bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b \mathbf{T}_{t,LETKF}, \tag{1}$$

$$\mathbf{T}_{t,LETKF} = \widetilde{\mathbf{P}}_t^a (\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} \left(\mathbf{y}_t^o - \overline{H_t(\mathbf{X}_t^b)} \right) \cdot \mathbf{1} + \left[(m-1)\widetilde{\mathbf{P}}_t^a \right]^{1/2}, \tag{2}$$

where $\mathbf{1}$ is a row vector whose all elements are 1 $(\in \mathbb{R}^m)$, $\mathbf{T}$ is the ensemble transform matrix $(\in \mathbb{R}^{m \times m})$, $\widetilde{\mathbf{P}}$ is the error covariance matrix in the ensemble space $(\in \mathbb{R}^{m \times m})$, $\mathbf{Y} \equiv \mathbf{HZ}$ is the ensemble perturbation matrix in the observation space $(\in$

$\mathbb{R}^{p \times m}$), $\mathbf{R}$ is the observation error covariance matrix ($\in \mathbb{R}^{p \times p}$), $\mathbf{y}$ is the observation vector ($\in \mathbb{R}^{p}$), $\mathbf{H}$ is the linear observation operator matrix ($\in \mathbb{R}^{p \times n}$), and H is the observation operator that may be nonlinear. Here, p is the number of observations. The superscripts o, b, and a denote the observation, background (prior), and analysis (posterior), respectively. The matrix $\mathbf{T}_{LETKF}$ denotes the ensemble transform matrix of the LETKF. The first and second terms of the right-hand side of Eq. (2) correspond to the updates of ensemble mean and perturbation, respectively. The ETKF computes the analysis error covariance matrix in the ensemble space $\widetilde{\mathbf{P}}_t^a$ by:

$$\widetilde{\mathbf{P}}_t^a = \left[\frac{(m-1)}{\beta} \mathbf{I} + (\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} \mathbf{Y}_t^b \right]^{-1}, \tag{3}$$

where β is a multiplicative inflation factor. Equations (1) and (2) are derived from the Kalman filter equations given by:

$$\bar{\mathbf{x}}_t^a = \bar{\mathbf{x}}_t^b + \mathbf{K}_t \left(\mathbf{y}_t^o - \overline{H_t(\mathbf{X}_t^b)} \right), \tag{4}$$

$$\mathbf{K}_t = \mathbf{P}_t^b \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t)^{-1} = \mathbf{P}_t^a \mathbf{H}_t^T \mathbf{R}_t^{-1}, \tag{5}$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^b \Longleftrightarrow (\mathbf{P}_t^a)^{-1} = (\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t, \tag{6}$$

where $\mathbf{K}$ is the Kalman gain ($\in \mathbb{R}^{n \times p}$) and $\mathbf{P}$ is the error covariance matrix in the model space ($\in \mathbb{R}^{n \times n}$). Derivations of Eqs. (5) and (6) are detailed in Appendix A. The EnKF approximates the error covariance matrix by $\mathbf{P} \approx \frac{1}{m-1} \mathbf{Z}(\mathbf{Z})^T$. Hunt et al. (2007) provides more details on deriving Eqs. (1) and (2) from the Kalman filter equations. For nonlinear observation operators, the following approximation is used:

$$\mathbf{Y} \equiv \mathbf{HZ} \approx H(\mathbf{X}) - \overline{H(\mathbf{X})} \cdot \mathbf{1}. \tag{7}$$

The LETKF provides $\widetilde{\mathbf{P}}_t^a$ and $\left(\widetilde{\mathbf{P}}_t^a \right)^{1/2}$ of Eq. (2) by solving the eigenvalue decomposition of $\left(\widetilde{\mathbf{P}}_t^a \right)^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is a square matrix ($\in \mathbb{R}^{m \times m}$) composed of eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix ($\in \mathbb{R}^{m \times m}$) composed of the corresponding eigenvalues. The eigenvalue decomposition leads to $\widetilde{\mathbf{P}}_t^a = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$ and $\left(\widetilde{\mathbf{P}}_t^a \right)^{1/2} = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{U}^T$.

The localization is practically important for mitigating sampling errors in the ensemble-based error covariance with a limited ensemble size (Houtekamer and Zhang, 2016). With localization, the LETKF computes the transform matrix $\mathbf{T}_{LETKF}$ at every model grid point independently by assimilating a subset of observations within the localization cut-off radius. The LETKF employs localization by inflating the observation error variance so that observations distant from the analysis model grid point have fewer impacts (Hunt et al., 2007; Miyoshi and Yamane, 2007).

For local analysis schemes, a spatially smooth transition of the transform matrix $\mathbf{T}_{LETKF}$ is essential to prevent abrupt changes in the analyses of neighboring grid points. The LETKF realizes a smooth transition of the transform matrix by using the symmetric square root of $\widetilde{\mathbf{P}}_t^a$ (Hunt et al., 2007). The symmetric square root matrix minimizes the mean-square distance

between identity matrix $\mathbf{I}$ and $\left[(m-1)\widetilde{\mathbf{P}}_t^a\right]^{1/2}$; therefore, the analysis ensemble perturbations can be closer to the background ensemble perturbation.

The EnKF generally underestimates the error variance mainly because of model errors, nonlinear dynamics, and limited ensemble size. Therefore, a covariance inflation technique is used to inflate the underestimated error variance (Houtekamer and Mitchell, 1998). Among several kinds of covariance inflations methods, the present study considers multiplicative inflation (Anderson and Anderson, 1999) and relaxation to prior scheme (Zhang et al., 2004) as implemented by Whitaker and Hamill (2012). In multiplicative inflation, the ensemble-based covariance is multiplied by a factor β ($\mathbf{P}^b \rightarrow \beta\mathbf{P}^b$). This multiplicative inflation is employed in Eq. (3) in the LETKF. In this study, the LETKF experiments use the approach of Miyoshi (2011) which adaptively estimates a spatially varying inflation factor on the basis of the innovation statistics of Desroziers et al. (2005). In realistic problems, covariance relaxation methods are often used to inflate the posterior perturbation (e.g., Terasaki et al. 2019; Kotsuki et al., 2019b). This study utilizes the relaxation to prior spread (RTPS; Whitaker and Hamill, 2012) given by:

$$\mathbf{Z}_{t(k)}^a \leftarrow \left((1-\alpha) + \alpha\frac{\sigma_{t(k)}^b}{\sigma_{t(k)}^a}\right)\mathbf{Z}_{t(k)}^a, \tag{8}$$

where α is the RTPS parameter, and σ is the ensemble spread. The subscript (k) denotes the kth model variable or the kth component of state vector. Although this study uses only multiplicative inflation for the LETKF experiments, the posterior error perturbation is inflated by RTPS for LPF and LPFGM.

2.2 Local particle filter with ensemble transform matrix

Here, we describe the LPF in the form of the ensemble transform matrix of the LETKF (Reich, 2013; Penny and Miyoshi, 2016; Potthast et al. 2019). The PF is a direct Monte Carlo realization of Bayes' theorem given by:

$$\pi(\mathbf{x}_t|\mathbf{y}_{1:t}^o) = \frac{\pi(\mathbf{y}_t^o|\mathbf{x}_t)\pi(\mathbf{x}_t|\mathbf{y}_{1:t-1}^o)}{\pi(\mathbf{y}_t^o|\mathbf{y}_{1:t-1}^o)}, \tag{9}$$

where $\pi(\mathbf{x}_t|\mathbf{y}_{1:t}^o)$ is the probability of state $\mathbf{x}$ given all observations $\mathbf{y}^o$ up to time t. The PF approximates the prior probability density function (PDF), which appears in the numerator of Eq. (9), using an ensemble forecast:

$$\pi_{LPF}(\mathbf{x}_t|\mathbf{y}_{1:t-1}^o) \approx \sum_{i=1}^m w_t^{b(i)}\delta(\mathbf{x}_t - \mathbf{x}_t^{b(i)}), \tag{10}$$

where δ is the Dirac's delta function and $w^{b(i)}$ is the prior weight of the ith particle (i.e., ensemble member) at the previous analysis time. If particles are resampled at the previous analysis time, all particles have the same weight $w^{b(i)} = 1/m$ ($i = 1, \dots, m$). The sum of the weights $\sum_{i=1}^m w_t^{a,b(i)}$ is always 1. The denominator of the right-hand side of Eq. (9) can be estimated by:

$$\pi(\mathbf{y}_t^o|\mathbf{y}_{1:t-1}^o) = \int \pi(\mathbf{y}_t^o|\mathbf{x}_t)\pi(\mathbf{x}_t|\mathbf{y}_{1:t-1}^o)d\mathbf{x}_t \approx \sum_{i=1}^m w_t^{b(i)}\pi(\mathbf{y}_t^o|\mathbf{x}_t^{b(i)}). \tag{11}$$

150   This study assumes a Gaussian likelihood function given by:

$$\pi(\mathbf{y}_t^o | \mathbf{x}_t) \propto \exp\left[-\frac{1}{2}\left(\mathbf{y}_t^o - H_t(\mathbf{x}_t)\right)^T \mathbf{R}_t^{-1}\left(\mathbf{y}_t^o - H_t(\mathbf{x}_t)\right)\right]. \tag{12}$$

This assumption means the observation error distribution is assumed to be Gaussian. Then, the posterior can be expressed by:

$$\pi_{LPF}(\mathbf{x}_t | \mathbf{y}_{1:t}^o) \approx \sum_{i=1}^m w_t^{a(i)} \delta\left(\mathbf{x}_t - \mathbf{x}_t^{b(i)}\right), \tag{13}$$

$$w_t^{a(i)} = w_t^{b(i)} \pi_{LPF}\left(\mathbf{y}_{1:t}^o | \mathbf{x}_t^{b(i)}\right) / \sum_{j=1}^m w_t^{b(i)} \pi_{LPF}\left(\mathbf{y}_{1:t}^o | \mathbf{x}_t^{b(j)}\right) = w_t^{b(i)} q_t^{(i)} / \sum_{k=1}^m w_t^{b(k)} q_t^{(k)}, \tag{14}$$

155
$$q_t^{(i)} = \exp\left[-\frac{1}{2}\left(\mathbf{y}_t^o - H_t\left(\mathbf{x}_t^{b(i)}\right)\right)^T \mathbf{R}_t^{-1}\left(\mathbf{y}_t^o - H_t\left(\mathbf{x}_t^{b(i)}\right)\right)\right], \tag{15}$$

where q is the likelihood. Equation (14) results in posterior weights that satisfy $\sum_{i=1}^m w_t^{a(i)} = 1$.

     To mitigate weight collapse, the *local* PF (LPF) solves the PF equations by assimilating local observations surrounding the analysis grid point. This study uses Penny and Miyoshi (2016)'s approach that computes the analysis at every model grid point independently with the observation error covariance **R** being inflated by the inverse of a localization function
160   as in the LETKF (Penny and Miyoshi, 2016).

     Sampling importance resampling (SIR) is a technique for applying the PF for high-dimensional dynamics with a limited amount of particles. The resampling process rearranges the particles to effectively represent the densest areas of the posterior PDF. After resampling, each particle has equal weights and the posterior PDF is given by:

$$\pi_{LPF}(\mathbf{x}_t | \mathbf{y}_{1:t}^o) \approx \frac{1}{m} \sum_{i=1}^m \delta\left(\mathbf{x}_t - \mathbf{x}_t^{a(i)}\right). \tag{16}$$

165   The resampling process can be expressed equally valid to the ensemble transform matrix of the ETKF (Reich, 2013; Penny and Miyoshi, 2016) given by:

$$\mathbf{X}_{t,LPF}^a = \mathbf{X}_t^b \mathbf{T}_{t,LPF} = (\bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b)\mathbf{T}_{t,LPF} = \bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b \mathbf{T}_{t,LPF}. \tag{17}$$

where $\mathbf{T}_{LPF}$ denotes the ensemble transform matrix of the LPF. Here, we applied $(\bar{\mathbf{x}}_t^b \cdot \mathbf{1})\mathbf{T}_{t,LPF} = \bar{\mathbf{x}}_t^b \cdot \mathbf{1}$ based on the following necessary condition of the ensemble transform matrix for the LPF:

170
$$\sum_{i=1}^m \mathbf{T}_{t,LPF}^{(i,j)} = 1, \qquad j = 1, \dots, m \tag{18}$$

In addition to Eq. (18), the ideal resampling matrix satisfies the following two conditions:

- $\sum_{j=1}^m \mathbf{T}_{t,LPF}^{(i,j)} = m \cdot w_t^{a(i)}, \qquad for\ i = 1, \dots, m$

  $\Leftrightarrow \bar{\mathbf{x}}_t^b = \sum_{i=1}^m \mathbf{x}_t^{b(i)} \cdot w_t^{a(i)}$,   (19)

- spatially smooth transition of $\mathbf{T}_{t,LPF}$.

175   where m is the ensemble size, and $\mathbf{T}^{(i,j)}$ indicates the *i*th-row and *j*th-column element of the matrix **T**.

The resampling matrix significantly affects the filter performance (Farchi and Bocquet, 2018). The present study constructs resampling matrices based on the Monte-Carlo approach. This resampling method uses m random numbers $\mathbf{r}^{(i)}$ ($i = 1, \ldots, m$) drawn from uniform distribution U ([0,1]) and accumulated weights $w_{t,acc}$:

$$w_{t,acc}^{(0)} = 0, \qquad w_{t,acc}^{(i)} = w_{t,acc}^{(i-1)} + w_t^{a(i)}, \qquad i = 1, \ldots, m. \tag{20}$$

180 By definition, $w_{t,acc}^{(m)}$ is 1. After sorting $\mathbf{r}$ to be the ascending order $\mathbf{r}_{sorted}$, we generate the resampling matrix using **Algorithm 1**. This procedure is similar to the resampling of Potthast et al. (2019), but it employs an additional treatment so that the resampling matrix is close to the identity matrix (line 10 of **Algorithm 1**; Fig. 1 b and c).

Classical resampling can produce the same particles resulting in the loss of diversity of posterior particles. Therefore, additional treatments are required to generate slightly different posterior particles. Potthast et al. (2019) proposed adding 185 Gaussian random noise to $\mathbf{T}_{t,LPF}$ to prevent the generation of the same particles. As an alternative, the present study uses the Monte-Carlo approach that repeats **Algorithm 1** many times with different random numbers and takes the average of the generated resampling matrices (Fig. 1). Using different random numbers $\mathbf{r}$, the resampling matrices differ (Fig. 1 b, c) even if the same weight is used (Fig. 1 a). By averaging the resampling matrices generated by using different random numbers, we get matrices that have higher weights in the diagonal components and lower weights in the off-diagonal components (Fig. 1 d, 190 e). The Monte-Carlo approach makes the resampling matrix closer to the identity matrix, which is beneficial for a smooth transition in space. In addition, this stochastic approach approximates Eq. (19) using the Monte-Carlo approach. The generated transform matrix with 200 samples (Fig. 1 c) is close to that with 10,000 samples (Fig. 1 d) in the case of 40 particles. Therefore, the transform matrices are generated by averaging 200 sampled matrices in subsequent LPF experiments. Hereafter, we call this resampling method "MC resampling" which is used in the following experiments. The number of required samples for 195 MC resampling is briefly investigated in Appendix B.

The effective particle (or ensemble) size N_{eff} (Kong et al., 1994) is useful to measure the diversity of particles in LPF:

$$N_{t,eff} = 1 / \sum_{i=1}^{m} \left(w_t^{a(i)} \right)^2 \in [1, m]. \tag{21}$$

If N_{eff} is sufficiently large ($N_{eff} \approx m$), no resampling is needed. This study considers a tunable parameter N_0 as a criterion 200 for resampling:

$$\mathbf{T}_{t,LPF} \leftarrow \begin{cases} \mathbf{I} & if \quad N_{t,eff} > N_0 \\ \mathbf{T}_{t,LPF} & else \end{cases}. \tag{22}$$

Without resampling ($N_{t,eff} > N_0$), the posterior weight is succeeded for the subsequent forecast as follows:

$$\mathbf{w}_{t+1}^b \leftarrow (1 - \tau) \cdot \mathbf{w}_t^a + \tau/m, \tag{23}$$

where τ is the tunable forgetting factor $\tau \in [0:1]$. Here, $\tau = 1$ means the subsequent prior weight $\mathbf{w}_{t+1}^b$ has the same weights $1/m$ whereas the posterior weight $\mathbf{w}_t^a$ is completely succeeded to $\mathbf{w}_{t+1}^b$ when $\tau = 0$. This weight succession (Eq. 23) can be interpreted as a temporal localization that reduces the impact of observations temporally distant from the assimilation time. The weight succession in *local* PF is not trivial because the weight (or likelihood of particles) would move with the dynamical flow. Similar discussions can be found for the advection of localization functions (e.g., Ota et al., 2013; Kotsuki et al. 2019a). This study assumes no flow motion of the weights and simply use Eq. (23) at each grid point independently.

Although the LETKF applies inflation to the prior error covariance (i.e., $\mathbf{P}^b \rightarrow \beta \mathbf{P}^b$), this inflation is suboptimal to the LPF partly because the weight collapse occurs more easily if multiplicative inflation is applied to the prior perturbation (cf. Eq. 15). Therefore, the LPF usually applies inflation to the posterior particles (e.g., Penny and Miyoshi 2016; Farchi and Bocquet 2018). The present study uses the RTPS to inflate posterior perturbations for the LPF. We do not use the RTPS for the LETKF experiments since the Miyoshi (2011)'s adaptive multiplicative inflation is known to outperform the manually-tuned RTPS for idealized twin experiments with SPEEDY based on the authors' preliminary experiments and personal communication (Ota et al. 2021).

Preliminary experiments showed that the RTPS outperformed the relaxation to prior perturbation (RTPP; Zhang et al., 2004) for the LPF. Having some noise in the transform matrix is important, so that the LPF can maintain the diversity of posterior particles. The RTPP makes the transform matrix closer to the identity matrix, resulting in less diverged posterior particles. Therefore, the RTPS would be a more suitable relaxation method than the RTPP for LPF.

2.3 LPF with Gaussian mixture extension

Assimilating too many independent observations in the local area is not desirable for the LPF to avoid weight collapse (van Leuuwen et al., 2019). To solve this problem, hybrid algorithms of EnKF and PF have been explored to efficiently assimilate massive observation data. The Gaussian mixture extension of the LPF is one such hybrid algorithms (Hoteit et al. 2008; Stordal et al., 2011; WP22). In the Gaussian mixture extension, the previous PDF is approximated by a combination of Gaussian distributions centered at the values of the particles, given by:

$$\pi_{LPFGM}(\mathbf{x}_t|\mathbf{y}_{1:t-1}^o) \approx \sum_{i=1}^m w_t^{b(i)} N(\mathbf{x}_t^{b(i)}, \widehat{\mathbf{P}}_t^b), \tag{24}$$

where $N(\mathbf{x}_t^{b(i)}, \widehat{\mathbf{P}}_t^b)$ is the Gaussian kernel with mean $\mathbf{x}_t^{b(i)}$ and covariance $\widehat{\mathbf{P}}_t^b$. Here, hat indicates matrices for the Gaussian kernels (e.g., prior error covariance $\widehat{\mathbf{P}}^b$, observation error covariance $\widehat{\mathbf{R}}$ and Kalman gain $\widehat{\mathbf{K}}$). The covariance of the Gaussian kernel uses the sampled covariance matrix $\mathbf{P}_t^b$ such that:

$$\widehat{\mathbf{P}}_t^b = \gamma \mathbf{P}_t^b \approx \frac{\gamma}{m-1} \mathbf{Z}_t^b (\mathbf{Z}_t^b)^T, \tag{25}$$

where γ (>0) is a tunable parameter that regulates the amplitude and width of the Gaussian kernel (i.e., uncertainty of particles). For example, larger γ reduces amplitude and increases width of the Gaussian kernel. Because the kernel is supposed to have a

Gaussian distribution, increasing the Gaussian kernel's width results in a decrease in amplitude. The LPFGM results in the same analysis of the LPF when γ is 0 ($\gamma \to 0$).

The Gaussian mixture performs a two-step update to obtain the posterior particles. The first update moves the center of the Gaussian kernel with observations. Since kernels are Gaussian, we can use the Kalman filter for the first update:

$$\mathbf{x}_t^{a(i)} = \mathbf{x}_t^{b(i)} + \widehat{\mathbf{K}}_t \left(\mathbf{y}_t^o - H(\mathbf{x}_t^{b(i)}) \right), \qquad (i = 1, \dots, m)$$

$$\Leftrightarrow \mathbf{X}_{t,GM}^a = \mathbf{X}_t^b + \widehat{\mathbf{K}}_t \left(\mathbf{y}_t^o \cdot \mathbf{1} - H_t(\mathbf{X}_t^b) \right), \tag{26}$$

where $\widehat{\mathbf{K}} = \widehat{\mathbf{P}}^a \mathbf{H}^T \mathbf{R}^{-1}$ as in Eq. (5). In the LETKF, the Kalman gain $\mathbf{K}$ is computed by $\mathbf{K} = \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1} \approx \mathbf{Z}^b \widetilde{\mathbf{P}}^a (\mathbf{Z}^b)^{\mathsf{T}} \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1}$. Since the Gaussian kernel uses the ensemble-based error covariance (Eq. 25), we can apply the exact algorithms of the LETKF to compute $\widehat{\mathbf{P}}^a$ by replacing β of Eq. (3) with γ (i.e., $\widehat{\mathbf{P}}^a = \left[\frac{(m-1)}{\gamma} \mathbf{I} + (\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} \mathbf{Y}_t^b \right]^{-1}$). Here the same gain matrix $\widehat{\mathbf{K}}$ is applied to update each particle independently while the gain matrix $\widehat{\mathbf{K}}$ is based on the forecast error covariance estimated from the entire ensemble. Consequently, Eq. (26) is equivalent to:

$$\mathbf{X}_{t,GM}^a = \bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b \mathbf{T}_{t,GM}, \tag{27}$$

where

$$\mathbf{S}_t = \widetilde{\mathbf{P}}_t^a (\mathbf{Y}_t^b)^T \mathbf{R}_t^{-1} \left(\mathbf{y}_t^o \cdot \mathbf{1} - H_t(\mathbf{X}_t^b) \right), \text{ and} \tag{28}$$

$$\mathbf{T}_{t,GM} = \mathbf{S}_t + \mathbf{I}. \tag{29}$$

Here, $\mathbf{T}_{t,GM}$ denotes the ensemble transform matrix of the Gaussian mixture.

The second update resamples the particles based on the likelihood of the posterior kernel given by:

$$q_t^{(i)} = \exp\left[-\frac{1}{2} \left(\mathbf{y}_t^o - H_t(\mathbf{x}_t^{b(i)}) \right)^T \widehat{\mathbf{R}}_t^{-1} \left(\mathbf{y}_t^o - H_t(\mathbf{x}_t^{b(i)}) \right) \right], \tag{30}$$

where

$$\widehat{\mathbf{R}}_t = \mathbf{R}_t + \mathbf{H}_t \widehat{\mathbf{P}}_t^b \mathbf{H}_t^T. \tag{31}$$

Hoteit et al. (2008) and Stordal et al. (2011) used Equation (30) for computing the likelihood of posterior kernels. Alternatively, WP22 suggested that using Eq. (15) instead of Eq. (30) is a reasonable approximation in the case of a smaller variance of $\widehat{\mathbf{P}}_t^b$ compared to the observation departure $\mathbf{y}_t^o - H_t(\mathbf{x}_t^{b(i)})$. This study follows the WP22's approximation because computing inverse of Eq. (31) is computationally much more expensive than computing inverse of diagonal $\mathbf{R}$. Using WP22's approximation, the solution of the LPFGM is given by:

$$\mathbf{X}_{t,LPFGM}^a = \mathbf{X}_{t,GM}^a \mathbf{T}_{t,LPF} = \left(\bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b \mathbf{T}_{t,GM} \right) \mathbf{T}_{t,LPF} = \bar{\mathbf{x}}_t^b \cdot \mathbf{1} + \mathbf{Z}_t^b \mathbf{T}_{t,GM} \mathbf{T}_{t,LPF}. \tag{32}$$

260 Here, we used $(\bar{\mathbf{x}}_t^b \cdot \mathbf{1})\mathbf{T}_{t,LPF} = \bar{\mathbf{x}}_t^b \cdot \mathbf{1}$ (cf. Eq. 18). Consequently, the transform matrix of the LPFGM ($\mathbf{T}_{LPFGM}$) is given by:

$$\mathbf{T}_{t,LPFGM} = \mathbf{T}_{t,GM}\mathbf{T}_{t,LPF}. \tag{33}$$

Namely, the LPFGM can be described as the ensemble transform matrix form. Representing the LPFGM with only one transform matrix $\mathbf{T}_{LPFGM}$ is practically beneficial if one aimed to reduce computational costs by the weight interpolation as for the LETKF in which the transform matrices at higher-resolution model grid points are interpolated by transform matrices at

265 coarser model grid points (Yang et al. 2009; Kotsuki et al. 2020). The weight interpolation is also useful for ensuring spatially smooth transition of the transform matrix (Potthast et al. 2019).

    The two-step update of the LPFGM may appear to use the same observations twice, but this is not true. To understand the principles, here we consider a simple scalar example with H=1.0 with illustrations of Fig. 2. Let $\pi(y_t^o)$ be an observation PDF (Fig. 2 top row), and the prior and posterior PDFs of the LPF are given by $\pi_{LPF}(x_t) \approx \frac{1}{m}\sum_{i=1}^{m}\delta(x_t - x_t^{b(i)})$ and

270 $\pi_{LPF}(x_t|y_t^o)$, respectively (Fig. 2a, 2nd and 3rd rows). The LPF employs resampling by approximating the posterior PDF as *a combination of prior particles* such that $\pi_{LPF}(x_t|y_t^o) \approx \sum_{i=1}^{m} w_t^{a(i)}\delta(x_t - x_t^{b(i)})$ (Fig. 2a, bottom row). Next, we focus on the 5th particle of the LPFGM (Fig. 2b). Prior and posterior PDFs of the 5th particle are given by $\pi(x_t^{(5)}) = N(x_t^{b(5)}, \hat{P}_t^b)$ and $\pi(x_t^{(5)}|y_t^o) = N(x_t^{b(5)}, \hat{P}_t^b) \cdot N(y_t^o, R_t)$, respectively (Fig. 2b, 2nd and 3rd rows). Since the Gaussian kernels are assumed for the prior particles, the center of the posterior kernel moves such that $\pi(x_t^{(5)}|y_t^o) \propto N(x_t^{a(5)}, \hat{P}_t^a)$, where $x_t^{a(5)}$ and $\hat{P}_t^a$ can be

275 computed by the Kalman filter (from the blue circle to the red circle in Fig. 2b). Since the LPFGM moves all particles, the posterior PDF of the LPFGM is given by $\pi_{LPFGM}(x_t|y_t^o) \approx \frac{1}{m}\sum_{i=1}^{m}\pi(x_t^{(i)}|y_t^o)$ (Fig. 2c, 3rd row). These movements correspond to the first update of the LPFGM. In contrast to the LPF, the posterior PDF of the LPFGM is approximated by *a combination of the posterior Gaussian kernels* (red circles of Fig. 2c, 3rd row). The LPFGM employs the resampling based on the moved particles such that $\pi_{LPFGM}(x_t|y_t^o) \approx \sum_{i=1}^{m} w_t^{a(i)}N(x_t^{a(i)}, \hat{P}_t^a)$ (Fig. 2c, bottom row). This resampling corresponds

280 to the second update of the LPFGM. As seen in this example, the two-step update of the LPFGM does not use the same observations twice.

2.4 Implementation and computational complexity

    We implemented the LPFGM based on the available LETKF code from Miyoshi (2005) and follow-on studies (Miyoshi and Yamane 2007, Kondo and Miyoshi 2019; Kotsuki et al. 2020; https://github.com/takemasa-miyoshi/letkf).

285 Figure 3 compares the workflows of the LETKF, LPF, and LPFGM. All DA methods execute the same first four steps (steps A–D). After step D, the LETKF involves four additional steps (steps E–H) to compute $\mathbf{T}_{LETKF}$. The LPF computes the weights of particles (step J), followed by the generation of the transform matrix $\mathbf{T}_{LPF}$ (step K). The LPFGM first executes steps E–G, as in the LETKF, and then executes step I to compute $\mathbf{T}_{GM}$, followed by the LPF algorithms to compute $\mathbf{T}_{LPF}$ (steps J and K).

Finally, the LPFGM multiplies $\mathbf{T}_{GM}$ and $\mathbf{T}_{LPF}$ to compute $\mathbf{T}_{LPFGM}$ (step L). At the end of the process, the transform matrix is
290 applied to the prior perturbation matrix $\mathbf{Z}^b$ to produce the analysis ensemble (step M), followed by the RTPS (step N).

If the LETKF code is available, the LPF in the transform matrix form can be developed by coding two steps J and K. The LPFGM can also be developed easily by coding two more steps I and L if the LETKF and LPF codes are available.

Here we compare the computational complexities of the LETKF, LPF and LPFGM algorithms (Table 1). The total cost (C_T) of a DA cycle is identical to the overhead of the assimilation system (C_H), plus n times the average local analysis
295 cost (C_L), and m times the cost of one-member model forecast (C_M):

$$C_T(n, p, m) = C_H(n, p, m) + n \cdot C_L(p_L, m) + m \cdot C_M, \tag{34}$$

where p_L is the number of local observations within the localization cut-off radius. We assume that the overhead and model costs are equivalent among the three methods. In addition, the total computational cost of DA usually depends on the local analysis cost (C_L):

$$300 \quad C_L^{LETKF} = O(m^3 + m^2 p_L), \tag{35}$$

$$C_L^{LPF} = O(m^2 N_{MC}), \tag{36}$$

$$C_L^{LPFGM} = O(m^3 + m^2 p_L + m^2 N_{MC}), \tag{37}$$

where N_{MC} is the number of times the resampling matrices are generated by **Algorithm 1**. The number of local observations p_L is usually much greater than the ensemble size m and N_{MC}. In this case, $O(m^2 p_L)$ is dominant for the LETKF, and LPFGM.

305 The computational cost of LPF here is more expensive than that with a simpler resampling algorithm such as the stochastic uniform resampling ($O(m)$; Penny and Miyoshi 2016, Farchi and Bocquet, 2018) due to the relatively complex **Algorithm 1** ($O(m^2 N_{MC})$). We could not use such simpler approaches in this study since they do not yield ideal resampling matrices that satisfy the two conditions (Eq. 19 and spatially smooth transition of $\mathbf{T}_{t,LPF}$). The computational cost $O(m^2 N_{MC})$ of LPF is still much smaller than the LETKF if $N_{MC} \ll p_L$.

3 Experimental settings

3.1 SPEEDY model

This study used the intermediate global atmospheric general circulation model SPEEDY (Molteni, 2003) to compare the LETKF, LPF and LPFGM. The SPEEDY model is a computationally-inexpensive hydrostatic model with fundamental physical parameterization schemes such as surface flux, radiation, convection, cloud and condensation. The SPEEDY model
315 has 96 × 48 grid points in the horizontal plane (T30 ~ 3.75° × 3.75°) and sigma-coordinate seven vertical layers. The SPEEDY consists of five prognostic variables: temperature (T), specific humidity (Q), zonal wind (U), and meridional wind (V) at seven layers, as well as surface pressure (Ps). The SPEEDY model coupled with the LETKF (SPEEDY−LETKF) has been widely

used in DA studies (Miyoshi 2005 and many follow-on studies, e.g., Miyoshi 2011; Kondo et al., 2013; Kondo and Miyoshi, 2016, 2019; Kotsuki and Bishop 2022). We implemented the LPF and LPFGM based on the existing SPEEDY−LETKF code following the procedures described in section 2.4.

3.2 Experimental design

The experimental settings follow the previous SPEEDY−LETKF experiments (Miyoshi 2005 and follow-on studies, e.g., Kotsuki et al., 2020). A series of idealized and identical twin experiments (also known as observing system simulation experiments) were conducted without model errors. We first performed a spin-up run for one year initialized by the standard atmosphere during rest, followed by the nature run started at 0000 UTC on January 1 of the second year. We assumed diagonal observation error covariance **R** (i.e., uncorrelated observation error). Gaussian noise was added to the nature run to produce observation data at 6 h intervals. The standard deviations of the Gaussian noise are 1.0 K for T, 1.0 m s^{-1} for U and V, 0.1 g kg^{-1} for Q, and 1 hPa for surface pressure. This study considered two observing networks (Figs. 4 a and b): the regularly distributed network (hereafter REG2) and the radiosonde-like inhomogeneous network (hereafter RAOB). We observe T, U and V at all seven layers, whereas Q was observed at the 1st−4th layers. The ensemble size is 40 and their initial conditions were taken from an independent single deterministic SPEEDY forecast with sufficient spin-up simulation.

Table 2 summarizes the settings of the LETKF, LPF, and LPFGM experiments. The observation error variance was inflated for the localization, by using the Gaussian-based function given by:

$$ l = \begin{cases} \exp\left[-\frac{1}{2}\{(d_h/\rho_h)^2 + (d_v/\rho_v)^2\}\right] & if \ \ d_h < 2\sqrt{10/3}\rho_h \ and \ d_v < 2\sqrt{10/3}\rho_v \\ 0 & else \end{cases} , \tag{38} $$

where l is the localization function, and its inverse l^{-1} is multiplied and used to inflate **R**. d_h and d_v are horizontal and vertical distances (km and log(Pa)) between the observation and analysis grid point. ρ_h and ρ_v are tunable horizontal and vertical localization scales (km and log(Pa)). Subscripts h and v represent horizontal and vertical, respectively. This study set the vertical localization scale ρ_v be 0.1 log(Pa) following Greybush et al. (2011) and Kondo et al. (2013). The horizontal localization scales of the LETKF were tuned manually before the experiments to minimize the first-guess root mean square error (RMSE) of the fourth layer (~500 hPa) temperature of the SPEEDY-LETKF since this is among the most important variables for medium-range NWP.

The LETKF experiments used the adaptive multiplicative inflation method of Miyoshi (2011) in which the inflation factor β was estimated adaptively. On the basis of sensitivity experiments for γ, we chose $\gamma = 1.5$. Sensitivity to γ is discussed in section 4.3.3.

We first performed a one-year SPEEDY-LETKF experiment over the second year from January to December following the one-year spin-up. We then performed LETKF, LPF, and LPFGM experiments from January to April in the third year initialized by the first-guess ensemble of the LETKF at 0000 UTC on January 1 of the third year. The results from the last three months, i.e., February to April, were used for verification. We assessed the 6-h forecast or background RMSE for T

at the fourth model level (∼500 hPa). While this study mainly discusses the RMSE and ensemble spread for T, similar results are observed for other variables. Here, this study will concentrate on the fourth-level T as an important variable for NWPs since the major goal of this study is to investigate the stabilities of LPF and LPFGM compared with the LETKF. Humidity verification or accounting for nonlinear observation operators and non-Gaussian observation errors are still crucial studies to investigate advantages of the LPF and LPFGM w.r.t. the LETKF.

4 Results and discussion

4.1 Experiments with a regularly distributed observing network

We first compare the LETKF, LPF, and LPFGM with REG2 in which the manually-tuned horizontal localization scale for the LETKF is 600 km. First, sensitivities to the horizontal localization scale (ρ_h) and RTPS parameter (α) are investigated for the LPF. Figure 5 indicates the time-mean background RMSEs and ensemble spreads for LETKF and LPF. The LPF requires large inflation ($\alpha \geq 1.0$) to avoid filter divergence. With large inflation ($\alpha \geq 1.0$), the LPF experiment with $\rho_h = 600$ km resulted in the smallest RMSEs among the three LPF experiments. The best performing localization scale of LPF was found similar to that of the LETKF. The four LPF experiments exhibited larger RMSEs than the LETKF, as demonstrated by the previous studies with a low-dimensional L96 model (Poterjoy, 2016; Penny and Miyoshi, 2016; Farichi and Bouquet, 2018). The LPF experiment with $\rho_h = 700$ km shows filter divergence when $\alpha = 1.0$. The ensemble spreads of LPF are smaller than the RMSEs when $\alpha = 1.0$. This under dispersive ensemble is a typical behavior of LPF (e.g., Poterjoy and Anderson 2016).

Second, we compare the time series of the background RMSEs, ensemble spreads, and effective particle sizes N_{eff} (Fig. 6). The RMSE and ensemble spread are consistent in the LETKF. However, the LPF shows generally smaller ensemble spreads than the RMSEs. Since the beginning of the experiments on January 1, the three LPF experiments showed rapid increases in the ensemble spread (Fig. 6 a), and a rapid decrease in the effective particle size (Fig. 6 b) within two weeks. After the rapid changes, the ensemble spreads and effective particle sizes were stabilized. The three LPF experiments increased RMSEs until the beginning of March. The LPF with $\rho_h = 500$ km showed filter divergence in April. The LPF with $\rho_h = 400$ km and $\rho_h = 600$ km also showed increasing trends in RMSE until the end of April while their ensemble spreads and effective particle sizes were stable.

Next, sensitivities to the RTPS parameter (α), resampling frequency (N_0), and forgetting factor (τ) are investigated for the LPFGM. The best performing localization scale of LPF was identical to that of the LETKF (Fig. 5). Therefore, we choose the same 600-km localization scale for the LPFGM. Figure 7 compares the time-mean background RMSEs and ensemble spreads for LETKF and LPFGM. In the LPFGM experiments, we investigate six experimental settings for $\tau = 0.0$, 0.1, 0.2, 0.3, 0.5, and 1.0. With $\tau = 0.0$, (Fig. 7 a), the LPFGM shows RMSEs similar to those of LETKF with best performing parameters for α and N_0. Regulating the resampling frequency (N_0) would be needed for LPFGM because the LPFGM shows filter divergence when resampling is employed at all assimilation steps excluding the case of $\alpha = 1.0$ (magenta line in Fig. 7

a). The LPFGM employs two ensemble updates: the Gaussian mixture and resampling. Owing to the first update by the Gaussian mixture, the LPFGM would not require resampling at all assimilation steps. With $\tau = 0.0$, the best performing relaxation parameter α is approximately 0.6–0.9, excluding the case employing resampling at all assimilation steps.

Increasing τ leads to the LPFGM being less sensitive to the resampling parameter N_0 and RTPS parameter α (Fig. 7 b-f), implying that the LPFGM is more stable when the LPFGM forgets weights ($\tau \geq 0.1$). Additionally, the LPFGM requires less inflation (i.e., smaller α) when the forgetting factor τ becomes larger. Owing to the first update by the Gaussian mixture, succeeding weights completely may be unnecessary for the LPFGM. In addition, the method of succeeding weights in the *local* PF is not trivial because the weights would move with the dynamical flow. Because the results of the LPFGM experiments without weight succession ($\tau = 1.0$) were superior to those with weight succession ($\tau = 0.0$), the remainder of this paper focuses on the LPFGM experiments without weight succession ($\tau = 0.0$).

The time series of background RMSEs, and effective particle sizes for LPFGM experiments show that the LPFGM with every-time resampling ($N_0 = 40$) exhibits large RMSEs and the smallest effective particle sizes (Fig. 8, magenta). In contrast, the LPFGM experiments with infrequent resampling ($N_0 = 10$, 5, and 2) show small RMSEs similar to that of the LETKF and maintain larger effective particle sizes than the LPF (Fig. 8, red, blue and green lines).

Finally, we compare the spatial patterns of the RMSEs at a fourth model level for the LETKF, LPF, and LPFGM with best performing parameter settings (Fig. 9). The LETKF and LPFGM show larger RMSEs in the tropics and polar regions (Figs. 9 a and c), possibly because of uncertainties from convective dynamics in the tropics and sparser observations in the polar regions. The LPF shows a larger RMSE than the LETKF globally (Fig. 9 b, d). By contrast, slight improvements were observed globally in the LPFGM relative to the LETKF as indicated by generally warm colors in Fig. 9 (e), especially around the North Pole.

4.2 Experiments with a realistic observing network

Here we compare the LETKF, LPF, and LPFGM with the realistic observing network RAOB. With this spatially inhomogeneous observing network, all LPF experiments showed filter divergence, even with a broad range of the localization and RTPS parameters. Since there are fewer assimilated observations in RAOB than in REG2, the weight collapse would not be the primary cause of the LPF filter divergence. Because the LPF creates posterior particles by linearly combining prior particles, it is preferable for the LPF to have observations within the range of the prior particles. Filter divergence must be prevented by maintaining synchronization between the LPF and the observations. The LPF would require more observations for the synchronization than the LETKF, which was also shown in the authors' initial tests with L96. In addition, the LPFGM was unstable with weight succession ($\tau = 0.0$). Therefore, this section focuses on the comparison of the LETKF and LPFGM without weight succession ($\tau = 1.0$) only.

First, sensitivities to the RTPS parameter (α) and resampling frequency (N_0) are investigated for the LPFGM. Figure 10 compares the time-mean background RMSEs and ensemble spreads for the LETKF and LPFGM. The best performing horizontal localization scale for the LETKF is 1100 km, and the same localization scale is used for the LPFGM. The LPFGM

with every-time resampling ($N_0 = 40$) shows the largest RMSEs (Fig. 10 a, magenta). By contrast, some LPFGM experiments show smaller RMSEs than the LETKF (Fig. 10 b). In this experimental setting, the resampling parameter of $N_0 = 2.0$ shows the most stable performance. With $N_0 = 2.0$, the LPFGM outperforms the LETKF slightly for $\alpha = 0.5$–0.7 (green line of Fig. 10 b).

The time series of the background RMSEs and effective particle sizes show that the LPFGM with every-time resampling ($N_0 = 40$; Fig. 11 a, magenta), and relatively-frequent resampling ($N_0 = 10$; Fig. 11 a, red) caused filter divergences. The times when the filters diverged seem to correspond to the times when the effective particle sizes decreased (Fig. 11 b, magenta and red). This reduction in effective particle size is a typical sign of filter divergence for PF (Snyder et al., 2008). One may speculate the assumption of $\widehat{\mathbf{R}}_t \approx \mathbf{R}_t$ in the LPFGM (section 2.3) would be a reason for the filter divergence. However, adopting the appropriate norm ($\widehat{\mathbf{R}}_t = \mathbf{R}_t + \mathbf{H}_t \widehat{\mathbf{P}}_t^b \mathbf{H}_t^T$) for the LPFGM did not eliminate the filter divergence (not shown). An effective solution for avoiding the filter divergence with every-time resampling is still unknown, but a possible reason is the method of creating a transform matrix using the Monte-Carlo method that approximates two ideal conditions (Eq. 19 and a spatially smooth transition). For deeper understanding, using other resampling techniques such as the optimal transport (see section 4.3.2) would be useful. In contrast, the LPFGM experiments with infrequent resampling ($N_0 = 5$ and 2) show stable RMSEs and maintain a larger effective particle size. These stable LPFGM experiments maintain the amplitude of the ensemble spreads similar to that of the RMSEs.

We further explore regions where the LPFGM outperforms the LETKF. Figure 12 compares the space-based patterns of the background RMSEs for T (K) at the fourth model level for LETKF and LPFGM. Errors in sparsely recognized regions, such as the South Pacific Ocean, Indian Ocean, and polar regions, tend to be larger (Figs. 12 a and b). The LPFGM outperforms the LETKF in such sparsely observed regions (Fig. 12 c). The regions showing improvements correspond to the regions where the ensemble spreads and first-guess non-Gaussianity are larger (Fig. 13). Here, we measure the non-Gaussianity by Kullback–Leibler divergence (Kullback and Leibler, 1951).

The first update of the LPFGM (the Gaussian mixture) is similar to the LETKF update as discussed. Therefore, the improvements of the LPFGM with respect to the LETKF would be owing to the resampling step. Previous studies demonstrated that the LPF outperformed the LETKF for non-Gaussian DA with non-Gaussian observation and prior PDFs (Poterjoy et al. 2016a, 2016b, Penny and Miyoshi 2016). Therefore, the improvements of the LPFGM would come from the consideration of non-Gaussian prior PDF in sparsely observed regions.

Finally, we investigate the region where the difference in RMSE between the LETKF and LPFGM is large (120°W-60°W and 70°S-30°S; indicated by black dashed rectangles in Figs. 12 and 13). Figure 14 compares the time series of RMSE, ensemble spread, effective particle size, and Kullback–Leibler divergence averaged over the region for the LETKF and LPFGM. Here, we conducted an additional LPFGM experiment that employs no resampling (i.e., $N_0 = 0$) to investigate the importance of resampling. Vertical red lines in Fig. 14 represent the cases when the LETKF has large RMSEs greater than 0.8 K. Figure 14 (a) indicates that the LPFGM tends to mitigate large RMSEs in contrast to the LETKF. Since no significant

increase is seen in Kullback–Leibler divergence at the four cases (Fig. 14 c), the large RMSEs of LETKF would not be caused by analyses with highly non-Gaussian first-guess ensemble. For the last three cases, the LPFGM has significantly smaller RMSEs than the LETKF. The improvement of the LPFGM would be partially led by larger ensemble spread than the LETKF (Fig. 14 b). However, the LPFGM with resampling ($N_0 = 2$) outperforms the LPFGM without resampling ($N_0 = 0$), indicating that resampling improves the RMSE. The LETKF and LPFGM exhibit reduced effective particle size at the three later cases (Fig. 14 d). Because of the smaller effective particle size, fewer ensembles predict states that are closer to observations, and the LPFGM employed more resampling in the region. In addition to the particle shift, the LPFGM would reduce RMSE further by the resampling.

The LPFGM occasionally resulted in larger RMSE than the LETKF (e.g., beginning of March in Fig. 14a). However, the LPFGM has the potential to outperform the LETKF overall by reducing the RMSE in sparsely observed regions. While operational NWP systems assimilate massive observations, the LPFGM would be useful when spatially and temporally sparse observations are used, e.g., in the twentieth century reanalysis projects (e.g. Compo et al. 2011; Laloyaux et al. 2018) and paleo climate reconstructions (Acevedo et al. 2017; Okazaki and Yoshimura 2017).

4.3 Factors requiring further investigation

Finally, we discuss other issues for potential further improvement of the LPF and LPFGM in future studies.

4.3.1 Inflation

This study used the RTPS to inflate the posterior perturbation for the LPF and LPFGM. As shown in Figs. 5, 7 and 10, the RMSEs of the LPF and LPFGM are sensitive to the RTPS parameter. There is a need to investigate methods that estimate the RTPS parameter adaptively, as in the EnKF (Ying and Zhang, 2015; Kotsuki et al., 2017b). However, the adaptive relaxation methods used in the EnKF cannot be applied directly to the LPF because they use the innovation statistics (Desroziers et al., 2005) that assume analysis updates by the Kalman filter. Therefore, substantially different adaptive relaxation methods should be investigated for the LPF and LPFGM.

An alternative method of inflation is adding random noise to the transform matrix (Potthast et al., 2019). Regulating the amplitude of the random noise was not trivial in the authors' preliminary experiments with L96 (not shown). Too small random noise results in loss of diversity of posterior particles, whereas excessively large random noise results in an overly dispersive ensemble. Potthast et al. (2019) also suggested estimating the amplitudes of the random noise based on the innovation statistics. However, determining the minimum and maximum values of the amplitudes was not trivial from the authors' experience with L96. Therefore, a method to determine the optimum amplitude of random noise should be investigated as well. Moreover, additive inflation methods, as used in the EnKF, can be beneficial for producing diverse prior particles (Mitchell and Houtekamer, 2000; Corazza et al., 2003). For example, Penny and Miyoshi (2016)'s LPF applied an additive

inflation Gaussian noise whose variance is scaled to a magnitude of local analysis error variance. Investigating better and adaptive inflation methods is very important to stabilize LPFs.

4.3.2 Transform matrix for LPF

The transform matrix significantly affects the filter performance. Farchi and Bocquet (2018) demonstrated that the optimal transport method of Reich (2013) resulted in lower RMSE than the commonly used stochastic uniform resampling approach with the L96 model. However, it is not clear if optimal transport is beneficial for high-dimensional systems such as the SPEEDY model.

In this study, all-time resampling was detrimental for the LPFGM experiments (Figs. 7 and 10). This may be due to the MC resampling that satisfies Eq. (19) ($\bar{\mathbf{x}}_t^b = \sum_{i=1}^m \mathbf{x}_t^{b(i)} \cdot w_t^{a(i)}$) not exactly, but approximately. With a small effective particle size, the MC resampling almost satisfies Eq. (19) since the resampling is not affected by the sampling noise in the selection of particles. In contrast, satisfying Eq. (19) is more difficult when the effective particle size is larger due to the sampling noise. A possible solution to this problem is to use the optimal transport method (Reich 2013) that satisfies Eq. (19) exactly. Another essential property for local PFs is the spatially smooth transition of the transform matrix. The optimal transport method would also be useful for ensuring the spatially smooth transition of the transform matrix because the posterior weights in nearby grids are typically similar. It is necessary to investigate better methods for generating resampling matrices.

4.3.3 Tunable parameters of LPF and LPFGM

The tunable parameters of the LPF and LPFGM should be investigated further. For example, parameter N_0, which controls the resampling frequency, significantly affected the filter accuracy and stability. Adaptive determination of this parameter can prevent time-consuming parameter tuning.

The sensitivity to parameter γ should also be investigated for the LPFGM. WP22 proposed using $\gamma = 2.5$. Since γ controls the amplitude and width of Gaussian kernels, the filter accuracy and stability of the LPFGM would be sensitive to this parameter. We briefly examined the sensitivity to this parameter with REG2 and RAOB (Fig. 15; $\gamma = 2.5$, 1.5, and 0.5 for red, blue, and green lines). The results indicate that the filter accuracy of the LPFGM is sensitive to this parameter, especially for spatially-inhomogeneous RAOB observing network. For our experimental settings, the best performing γ was 1.5 for both REG2 and RAOB. However, the optimal parameter would differ for different models, observing networks, and observing frequency. Also, there might be a relation between the optimal γ and ensemble size. Therefore, adaptive determination of this parameter is helpful.

The LPFGM without weight succession ($\tau = 1.0$) resulted in lower RMSE than that with weight succession ($\tau = 0.0$). However, the optimal parameter τ would be somewhere between 0.0 and 1.0. In addition, the weight succession at each model grid point would be suboptimal. Hence, the method for weight succession should also be explored further.

5 Summary

This study aims to develop a software platform for the LETKF, LPF, and LPFGM with the intermediate global atmospheric model SPEEDY. The main results of this investigation are briefly listed as follows:

1) The LPF and LPFGM were developed by only minor modifications to the existing LETKF system with the SPEEDY model.

2) With dense observations (REG2), the LPF showed stable filter performance with large inflation by the RTPS. The best performing localization scale of the LPF was identical to that of the LETKF. The LPF forecast was less accurate than the LETKF forecast. With sparse observations (RAOB), the LPF did not work.

3) The LPFGM showed more stability and lower forecast RMSEs than the LPF. In addition to the RTPS parameter, regulating the resampling frequency and the amplitude of Gaussian kernels was important for the LPFGM. The LPFGM without weight succession resulted in more stability and lower RMSEs than that with weight succession. With RAOB, the LPFGM forecast was more accurate than the LETKF forecast in sparsely observed regions where the background ensemble spread and non-Gaussianity are larger.

As discussed in section 4.3, there is much room for improvement in the LPF and LPFGM. While the LPFGM potentially provides a more accurate forecast than the LETKF, the LPFGM has more tunable parameters than the LETKF. Manually tuning these parameters by trying numerous experiments is computationally expensive. Therefore, adaptive methods for determining such tuneable parameters need to be explored. Also, it is important to investigate computationally more efficient methods to generate resampling matrices for the LPF and LPFGM.

The SPEEDY-based LETKF, LPF, and LPFGM used in this study are available as open-source software on Github (https://github.com/skotsuki/speedy-lpf). This can act as a useful platform to investigate the LPF and LPFGM further in comparison with the well-known LETKF.

Appendix A: Derivations of Kalman gain and analysis error covariance

Here we describe derivations of Kalman gain and analysis error covariance (Eqs. 5 and 6). The Kalman gain $\mathbf{K}_t = \mathbf{P}_t^b \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t)^{-1}$ can be changed as follows:

$$\mathbf{K}_t = [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t] \mathbf{P}_t^b \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{A1}$$

$$= [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} [\mathbf{H}_t^T + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T] (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{A2}$$

$$= [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} \mathbf{H}_t^T \mathbf{R}_t^{-1} [\mathbf{R}_t + \mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T] (\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{A3}$$

$$= [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} \mathbf{H}_t^T \mathbf{R}_t^{-1}. \tag{A4}$$

Using $\mathbf{K}_t = [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} \mathbf{H}_t^T \mathbf{R}_t^{-1}$, analysis error covariance $\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^b$ can be changed as follows:

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^b = \mathbf{P}_t^b - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_t^b \tag{A5}$$

$$= \mathbf{P}_t^b - [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t \mathbf{P}_t^b \tag{A6}$$

$$= \mathbf{P}_t^b - [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} [((\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t) - (\mathbf{P}_t^b)^{-1}] \mathbf{P}_t^b \tag{A7}$$

$$= \mathbf{P}_t^b - \{\mathbf{I} - [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1} (\mathbf{P}_t^b)^{-1}\} \mathbf{P}_t^b \tag{A8}$$

$$= [(\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t]^{-1}. \tag{A9}$$

540 Consequently, we can derivate $(\mathbf{P}_t^a)^{-1} = (\mathbf{P}_t^b)^{-1} + \mathbf{H}_t^T \mathbf{R}_t^{-1} \mathbf{H}_t$ from Eq. (A9), and $\mathbf{K}_t = \mathbf{P}_t^a \mathbf{H}_t^T \mathbf{R}_t^{-1}$ from Eq. (A4).

Appendix B: The number of required samples for the MC resampling

This appendix investigates the number of required samples for the MC resampling to obtain accurate transform matrices stably. Here we assume that 10,000 samples are sufficient to obtain accurate transform matrix ($\mathbf{T}_{N_{MC}=10,000}$). An absolute error (AE) of transform matrix with G samples ($\mathbf{T}_{N_{MC}=G}$) is defined to measure accuracy of the transform matrix:

545 $$AE = \sum_{i=1}^m \sum_{j=1}^m |\mathbf{T}_{N_{MC}=G}^{(i,j)} - \mathbf{T}_{N_{MC}=10,000}^{(i,j)}|. \tag{B1}$$

Figure B1 shows the AE as a function of the number of samples for ensemble members 10, 20, 40 and 80. For this investigation, we obtained 1,000 independent weights generated by uniform random numbers, and average, minimum and maximum AEs of 1,000 cases are shown. For four ensembles, AEs decrease rapidly for the first 1,000 samples, followed by gradual decrease until AE=0. To reach to 10 % of initial error at $N_{MC} = 1$, MC resampling requires about 70, 300, 400, and 600 samples for 550 ensemble members 10, 20, 40, and 80. It suggests that more ensemble requires more samples to obtain accurate transform matrix stably by the MC resampling.

Author contributions

SK developed the data assimilation system, conducted the experiments, analyzed the results and wrote this manuscript. TM is the PI and directed the research with substantial contribution to the development of this paper. KK developed the data 555 assimilation system with SK. RP contributed substantially to the introduction of the LPFGM to the standard local particle filter.

Data and code availability

Developed data assimilation system is available at (https://github.com/skotsuki/speedy-lpf), which is based on the exiting SPEEDY-LETKF system (https://github.com/takemasa-miyoshi/letkf). The data assimilation system used in this manuscript is archived on Zonodo (https://zenodo.org/record/6586309; doi: 10.5281/zenodo.6586309). Due to the large 560 volume of data (> 6 TB) and limited disk space, processed data and scripts for visualization are also shared on Zenodo at the same link.

Acknowledgements

The authors thank the members of the Data Assimilation Research Team, RIKEN Center for Computational Science (R-CCS), for valuable suggestions and discussions. SK thanks Mr. Ken Oishi of Chiba University for useful discussions. This 565 study was partly supported by the Japan Aerospace Exploration Agency (JAXA) Precipitation Measuring Mission (PMM),

advancement of meteorological and global environmental predictions utilizing observational 'Big Data' of the social and scientific priority issues (Theme 4) to be tackled by using post K computer of the FLAGSHIP2020 Project of the Ministry of Education, Culture, Sports, Science and Technology Japan (MEXT), the Program for Promoting Researches on the Supercomputer Fugaku" (Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation) of MEXT JPMXP1020200305, the Initiative for Excellent Young Researchers of MEXT, JST AIP Grant Number JPMJCR19U2, the Japan Society for the Promotion of Science (JSPS) KAKENHI grants JP18H01549, 21H04571 and 21H05002, JST PRESTO MJPR1924, and IAAR Research Support Program of Chiba University. The experiments were performed using the Supercomputer for earth Observation, Rockets, and Aeronautics (SORA) at JAXA.

References

Acevedo, W., Fallah, B., Reich, S., and Cubasch, U.: Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model. Clim. Past, 13, 545-557. doi: 10.5194/cp-13-545-2017, 2017

Ades, M., and van Leeuwen, P. J.: An exploration of the equivalent weights particle filter. Q. J. R. Meteorol. Soc., 139, 820–840. doi: 10.1002/qj.1995, 2013.

Ades, M., and van Leeuwen, P. J.: The equivalent-weights particle filter in a high-dimensional system. Q. J. R. Meteorol. Soc., 141, 484–503. doi: 10.1002/qj.2370, 2015

Anderson J. L.: An Ensemble Adjustment Kalman Filter for Data Assimilation. Mon. Wea. Rev., 129, 2884–2903. doi: 10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2, 2001

Anderson, J. L., and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev., 127, 2741–2758. doi: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, 1999

Bengtsson, T., Snyder, C., and Nychka, D.: Toward a nonlinear ensemble filter for high-dimensional systems. J. Geophys. Res., 108. doi: 10.1029/2002JD002900, 2003

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

Burgers G., van Leeuwen P. J., and Evensen G.: Analysis scheme in the ensemble Kalman filter. Mon. Wea. Rev., 126, 1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2, 1998

Compo, G. P. and Coauthors: The twentieth century reanalysis project. Q. J. R. Meteorol. Soc., 137, 1-28. doi: 10.1002/qj.776, 2011.

Corazza M., and Coauthors.: Use of the breeding technique to estimate the structure of the analysis "errors of the day." Nonlin. Processes Geophys., 10, 233–243. doi:10.5194/npg-10-233-2003, 2003

Desroziers G., Berre L., Chapnik B., and Poli P.: Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc., 131, 3385–3396. doi:10.1256/qj.05.108, 2005

Evensen G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., 99, 10143–10162. doi:10.1029/94JC00572, 1994

600 Farchi, A., and Bocquet, M.: Comparison of local particle filters and new implementations. Nonlin. Processes Geophys., 25, 765–807. doi: /10.5194/npg-25-765-2018, 2018

Gordon, N. J., Salmond, D. J., and Smith, A. F.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. 140F, 107–113. doi: 10.1049/ip-f-2.1993.0015, 1993

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and Ensemble Kalman Filter Localization
605 Techniques. Mon. Wea. Rev., 139, 511–522, doi: 10.1175/2010MWR3328.1, 2011

Hoteit, I., Pham, D. T., Triantafyllou, G., and Korres, G.: A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. Mon. Wea. Rev., 136, 317–334. doi: 10.1175/2007MWR1927.1, 2008

Houtekamer P. L., and Mitchell H. L.: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 796–811. doi: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998

610 Houtekamer, P. L., and Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev., 144, 4489–4532. doi: 10.1175/MWR-D-15-0440.1, 2016

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos : A local ensemble transform Kalman filter. Physica D, 230, 112–126. doi: 10.1016/j.physd.2006.11.008, 2007

Kondo, K., Miyoshi, T., and Tanaka, H.L.: Parameter sensitivities of the dual-localization approach in the local ensemble
615 transform Kalman filter. SOLA, 9, 174–178. doi: 10.2151/sola.2013-039, 2013

Kondo, K., and Miyoshi, T.: Impact of removing covariance localization in an ensemble Kalman filter: Experiments with 10 240 members using an intermediate AGCM. Mon. Wea. Rev., 144, 4849–4865. doi: 10.1175/MWR-D-15-0388.1, 2016

Kondo, K., and Miyoshi, T.: Non-Gaussian statistics in global atmospheric dynamics: a study with a 10240-member ensemble Kalman filter using an intermediate AGCM. Nonlin. Processes in Geophys., 26, 211–225. doi: 10.5194/npg-26-211-2019,
620 2019

Kong, A., Liu, J. S., and Wong, W. H.: Sequential imputations and Bayesian missing data problems. J. Am. Stat. Assoc., 89, 278–288. doi: 10.1080/01621459.1994.10476469, 1994

Kotsuki, S., Miyoshi, T., Terasaki, K., Lien, G.-Y., and Kalnay, E.: Assimilating the global satellite mapping of precipitation data with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM). J. Geophys. Res., 122, 631–650. doi:
625 10.1002/2016JD025355, 2017a

Kotsuki, S., Ota. Y., and Miyoshi, T.: Adaptive covariance relaxation methods for ensemble data assimilation : experiments in the real atmosphere, Q. J. R. Meteorol. Soc., 143, 2001–2015. doi: 10.1002/qj.3060, 2017b

Kotsuki, S., Kurosawa, K., and Miyoshi, T.: On the Properties of Ensemble Forecast Sensitivity to Observations. Q. J. R. Meteorol. Soc., 145, 1897–1914. doi: 10.1002/qj.3534, 2019a

630   Kotsuki S., Terasaki K., Kanemaru K., Satoh M., Kubota T. and Miyoshi T.: Predictability of Record-Breaking Rainfall in Japan in July 2018: Ensemble Forecast Experiments with the Near-real-time Global Atmospheric Data Assimilation System NEXRA. SOLA, 15A, 1–7. doi: 10.2151/sola.15A-001, 2019b

Kotsuki, S., Pensoneault, A., Okazaki, A. and Miyoshi, T.: Weight Structure of the Local Ensemble Transform Kalman Filter: A Case with an Intermediate AGCM. Q. J. R. Meteorol. Soc. doi: 10.1002/qj.3852, 2020

635   Kotsuki, S., and Bishop, H. C.: Implementing Hybrid Background Error Covariance into the LETKF with Attenuation-based Localization: Experiments with a Simplified AGCM. Mon. Wea. Rev., doi: 10.1175/MWR-D-21-0174.1, 2022

Kullback, S., and Leibler, R. A.: On information and sufficiency. Ann. Math. Statist., 22, 79–86. doi: 10.1214/aoms/1177729694, 1951

Laloyaux, P. and Coauthors: CERA-20C: A coupled reanalysis of the twentieth century. J. Adv. Modeling Earth Syst., 10,
640   1172-1195. doi: 10.1029/2018MS001273, 2018

Lien, G.-Y., Kalnay, E. and Miyoshi, T.: Effective assimilation of global precipitation : simulation experiments. Tellus A, 65, 1–16. doi: 10.3402/tellusa.v65i0.19915, 2013

Lien, G.-Y., Kalnay, E., Miyoshi, T., and Huffman, G. J.: Statistical Properties of Global Precipitation in the NCEP GFS Model and TMPA Observations for Data Assimilation. Mon. Wea. Rev., 144, 663–679. doi: 10.1175/MWR-D-15-0150.1, 2016

645   Lorenz, E.: Predictability - A problem partly solved. Proc. Seminar on Predictability, Reading, United Kingdom, ECMWF, 1–18., 1996

Lorenz, E. and Emanuel, K. A.: Optimal Sites for Supplementary Weather Observations : Simulation with a Small Model. J. Atmos. Sci., 55, 399–414. doi: 10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2, 1998

Mitchell H. L., and Houtekamer P. L.: An adaptive ensemble Kalman filter. Mon. Wea. Rev., 128, 416–433. doi:10.1175/1520-
650   0493(2000)128<0416:AAEKF>2.0.CO;2, 2000

Miyoshi, T.: Ensemble Kalman filter experiments with a primitive-equation global model. Ph.D. dissertation, University of Maryland, College Park, 197pp, 2005

Miyoshi, T., and Yamane, S.: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. Mon. Wea. Rev., 135, 3841–3861. doi: 10.1175/2007MWR1873.1, 2007

655   Miyoshi T.: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. Mon. Wea. Rev., 139, 1519–1535. doi:10.1175/2010MWR3570.1, 2011

Miyoshi, T., Kondo, K., and Imamura, T.: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. Geophys. Res. Lett., 41, 5264–5271. doi: 10.1002/2014GL060863, 2014

Molteni, F.: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and
660   variability in multi-decadal experiments. Clim. Dyn., 20, 175–191. doi: 10.1007/s00382-002-0268-2, 2003

Oczkowski, M., Szunyogh, I., and Patil, D. J.: Mechanisms for the development of locally low-dimensional atmospheric dynamics. J. Atmos. Sci., 62, 1135–1156. doi:10.1175/JAS3403.1, 2005

Okazaki, A., and Yoshimura, K.: Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction. Clim. Past, 13, 379-393. doi: 10.5194/cp-13-379-2017, 2017

665   Ota, Y., Derber, J. C., Miyoshi, T. and Kalnay, E.: Ensemble-based observation impact estimates using the NCEP GFS. Tellus, 65A, 20–38, doi:10.3402/tellusa.v65i0.20038, 2013

Ott E., Hunt B. R., Szunyogh I., Zsimin A. V., Kostelich E. J., Corazza M., Kalnay E. Patil D. J., and Y. J. A.: A local ensemble Kalman filter for atmospheric data assimilation. Tellus A, 56, 415–428. doi: 10.1111/j.1600-0870.2004.00076.x, 2004

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E.: Local Low Dimensionality of Atmospheric Dynamics. Phys.

670   Rev. Lett., 86, 5878–5881. doi: 10.1103/PhysRevLett.86.5878, 2001

Penny, S. G., and Miyoshi, T.: A local particle filter for high-dimensional geophysical systems. Nonlin. Processes Geophys., 23, 391–405. doi: 10.5194/npg-23-391-2016, 2016

Poterjoy, J.: A localized particle filter for high-dimensional nonlinear systems. Mon. Wea. Rev., 144, 59–76. doi: 10.1175/MWR-D-15-0163.1, 2016

675   Poterjoy, J., and Anderson, J. L.: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. Mon. Wea. Rev., 144, 2007–2020. doi: 10.1175/MWR-D-15-0322.1, 2016

Potthast, R., Walter, A., and Rhodin, A.: A Localized Adaptive Particle Filter within an operational NWP framework. Mon. Wea. Rev., 147, 345–362. doi: 10.1175/MWR-D-18-0028.1, 2019

Reich, S.: A nonparametric ensemble transform method for Bayesian inference. SIAM J. Sci. Comput. 35, A2013–A2024. doi:

680   10.1137/130907367, 2013

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to high-dimensional particle filtering. Mon. Wea. Rev., 136, 4629–4640. doi: 10.1175/2008MWR2529.1, 2008

Snyder, C., Bengtsson, T., and Morzfeld, M.: Performance bounds for particle filters using the optimal proposal. Mon. Wea. Rev., 143, 4750–4761. doi: 11/4750/72218, 2015

685   Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., and Vallès, B.: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. Comput. Geosci., 15, 293–305. doi: 10.1007/s10596-010-9207-1, 2011

Terasaki, K., Kotsuki, S., and Miyoshi, T.: Multi-year analysis using the NICAM-LETKF data assimilation system. SOLA, 15, 41–46. doi: 10.2151/sola.2019-009, 2019

van Leeuwen, P. J,: A consistent interpretation of the stochastic version of the Ensemble Kalman Filter. Q. J. R. Meteorol.

690   Soc., 146, 2815–2825. doi: 10.1002/qj.3819, 2020

van Leeuwen, P. J.: Particle filtering in geophysical systems. Mon. Wea. Rev., 137, 4089–4114. doi: 10.1175/2009MWR2835.1, 2009

van Leeuwen, P. J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter. Q. J. R. Meteorol. Soc., 136, 1991–1999. doi: 10.1002/qj.699, 2010

695   van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., and Reich, S.: Particle filters for high-dimensional geoscience applications: A review. Q. J. R. Meteorol. Soc., 145, 2335–2365. doi: 10.1002/qj.3551, 2019

Wang, X., Bishop, C. H., and Julier, S. J.: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble?. Mon. Wea. Rev., 132, 1590–1605. doi: 10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2, 2004

700 Walter A. and Potthast, R.: Particle Filtering with Model Error -a Localized Mixture Coefficients Particle Filter (LMCPF). (submitted to J. Meteor. Soc. Japan), 2022

Whitaker J. S., and Hamill T. M.: Ensemble Data Assimilation without Perturbed Observations. Mon. Wea. Rev., 130, 1913–1924. doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002

Whitaker, J. S., and Hamill, T. M.: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. Mon.
705   Wea. Rev., 140, 3078–3089. doi: 10.1175/MWR-D-11-00276.1, 2012

Yang, S.-C., Kalnay, E., Hunt, B., and Bowler, N. E.: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. Q. J. R. Meteorol. Soc., 135, 251–262. doi: 10.1002/qj.353, 2009

Ying Y., and Zhang F.: An adaptive covariance relaxation method for ensemble data assimilation. Q. J. R. Meteorol. Soc., 141, 2898–2906. doi:10.1002/qj.2576, 2015

710 Zhang F., Snyder C., and Sun J.: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. Mon. Wea. Rev., 132, 1238–1253. doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2, 2004

Zhu, M., van Leeuwen, P. J., and Amezcua, J.: Implicit equal-weights particle filter. Q. J. R. Meteorol. Soc., 142, 1904–1919. doi: 10.1002/qj.2784, 2016

715 Algorithms

Algorithm 1: Generation of the resampling matrix **T**

Requirement: r$_{sorted}$ and **w**$_{t,acc}$

1:	All components of **T** and **z** are initialized to be 0	
	#z: an indicator of resampled particles	
2:	**do** j=1 to m	# index of posterior particle
3:	**do** i=1 to m	# index of prior particle
4:	**if** $\{w_{t,acc}^{(i-1)} < r_{sored}^{(j)} \leq w_{t,acc}^{(i)}\}\ z^{(j)} = i$	# resampled particle
5:	**end do**	
6:	**end do**	
7:	**do** j=1 to m	
8:	$i = z^{(j)}$	
9:	**if** $\{\mathbf{T}^{(i,i)} = 0\}$ **then**	
10:	$\mathbf{T}^{(i,i)}$ =1	# diagonal components
11:	$z^{(j)} = 0$	
12:	**end if**	
13:	**end do**	
14:	**do** j=1 to m	# (loop for rows of **T**)
15:	**if** $\{z^{(j)} \neq 0\}$ **then**	
16:	**do** i=1 to m	# (loop for columns of **T**)
17:	**if** $\{sum(\mathbf{T}^{(:,j)}) = 0\}\ \mathbf{T}^{(i,j)}$ =1	# off-diagonal components
18:	**end do**	
19:	**end if**	
20:	**end do**	

720

Tables

Table 1: Computational complexities of LETKF, LPF, and LPFGM. Each step corresponds to the steps in Fig. 1. Cross-marks represent the steps used for LETKF, LPF and LPFGM.

	Step	Computational complexity	LETKF	LPF	LPFGM
C_H	A	m applications of H	X	X	X
	B	$2mp$	X	X	X
	C	$2nm$	X	X	X
C_L	D	※1	X	X	X
	E	mp_L (※2)	X		X
	F	$\leq 2m^2 p_L$	X		X
	G	$O(m^3)$	X		X
	H	$O(m(m+p_L))$	X		
	I	$O(m(m+p_L))$			X
	J	$O(mp_L)$ (※2)		X	X
	K	$O(m^2 \cdot N_{MC})$		X	X
	L	$\leq 2m^3$			X
C_H	M	$\leq 2nm^2$	X	X	X
	N	$O(nm)$	X	X	X

725 ※1: This computation depends on the localization scale

※2: These computations assume the diagonal **R** (i.e., uncorrelated observation error)

n: system size

m: ensemble size

p: the number of observations

730 p_L: the number of local observations

N_{MC}: the number of times resampling matrices to be generated by **Algorithm 1**

735

Table 2: List of experiments

Obs	DA	Purpose	ρ_h (km)	α	γ	N_0	τ	Figures
REG2	LETKF	control experiment	600	※1	/	/	/	
	LPF	sensitivity to ρ_h and α	400/500/600/700	0.0-1.0	/	/	/	Figs. 5, 6, 9
	LPFGM	sensitivity to α, N_0, and τ	600	0.0-1.0	1.5	2/4/10/40	0.0/0.1/0.2/0.3/0.5/1.0	Figs. 7, 8, 9
	LPFGM	sensitivity to α and γ	600	0.0-1.0	0.5/1.5/2.5	2	1.0	Figs. 15 (a)
RAOB	LETKF	control experiment	1,100	※1				
	LPF	sensitivity to ρ_h and α	1,100	0.0-1.0	/	/	/	not shown
	LPFGM	sensitivity to α and N_0	1,100	0.0-1.0	1.5	2/4/10/40	1.0	Figs. 10, 11, 12, 13
	LPFGM	sensitivity to N_0	1,100	0.60	1.5	0/2	1.0	Fig. 14
	LPFGM	sensitivity to α and γ	1,100	0.0-1.0	0.5/1.5/2.5	2	1.0	Figs. 15 (b)

740   ρ_h: horizontal localization scale (Eq. 38)

α: relaxation parameter of RTPS (Eq. 8)

γ: amplitude of Gaussian kernel of the LPFGM (Eq. 25)

N_0: parameter controls resampling frequency (Eq. 22)

τ: forgetting factor of weight (Eq. 23)

745   ※1: Instead of RTPS, the LETKF uses Miyoshi (2011)'s adaptive multiplicative inflation

750

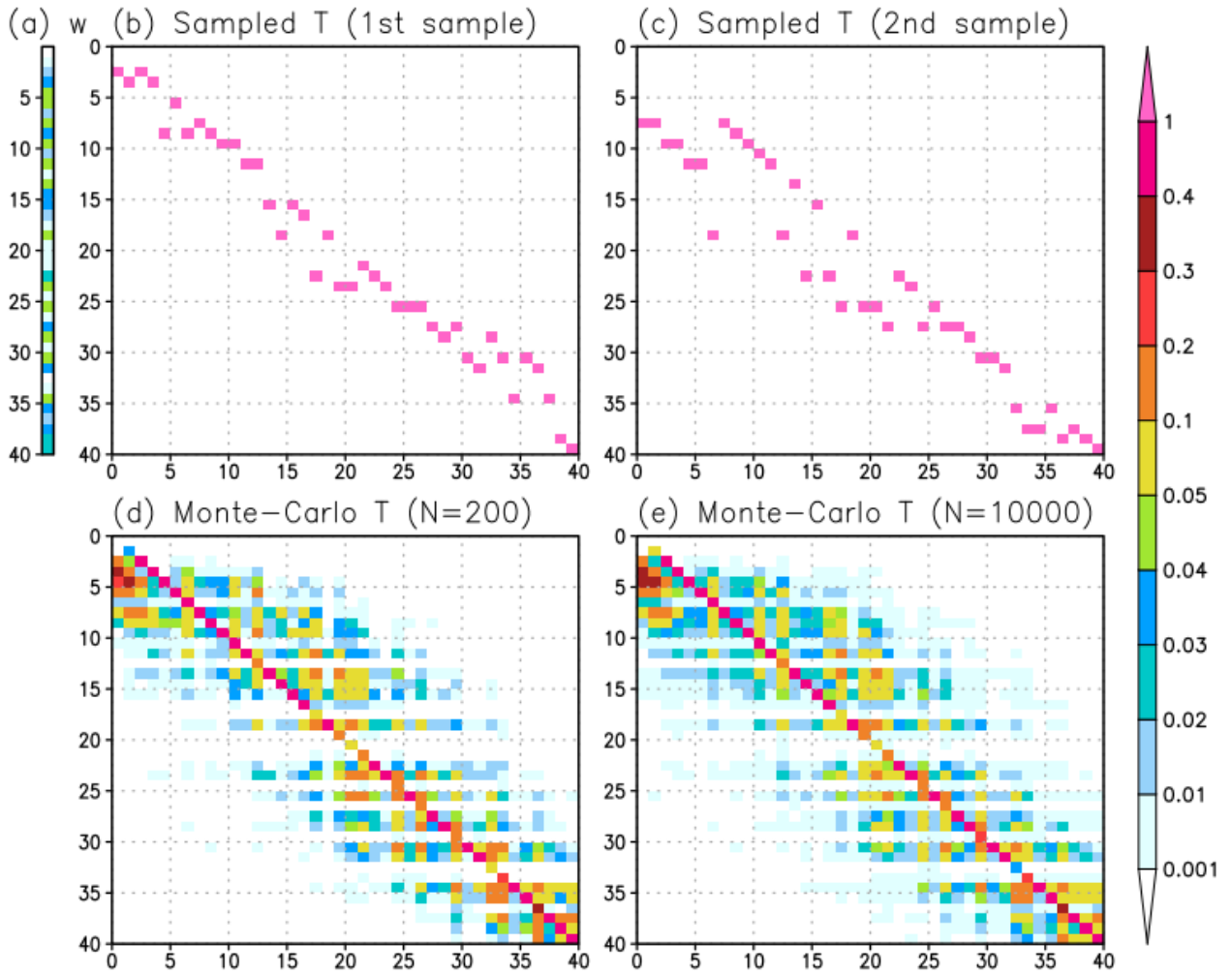
Figures

755

Figure 1: Examples of the resampling matrix in the case of 40 particles with a given weight generated by uniform random numbers. (a) Weight for 40 particles. (b, c) Examples of sampled resampling matrices by **Algorithm 1** using different random numbers **r**. (d, e) Resampling matrices using the Monte Carlo stochastic approach with 200 and 10,000 sampled matrices, respectively.

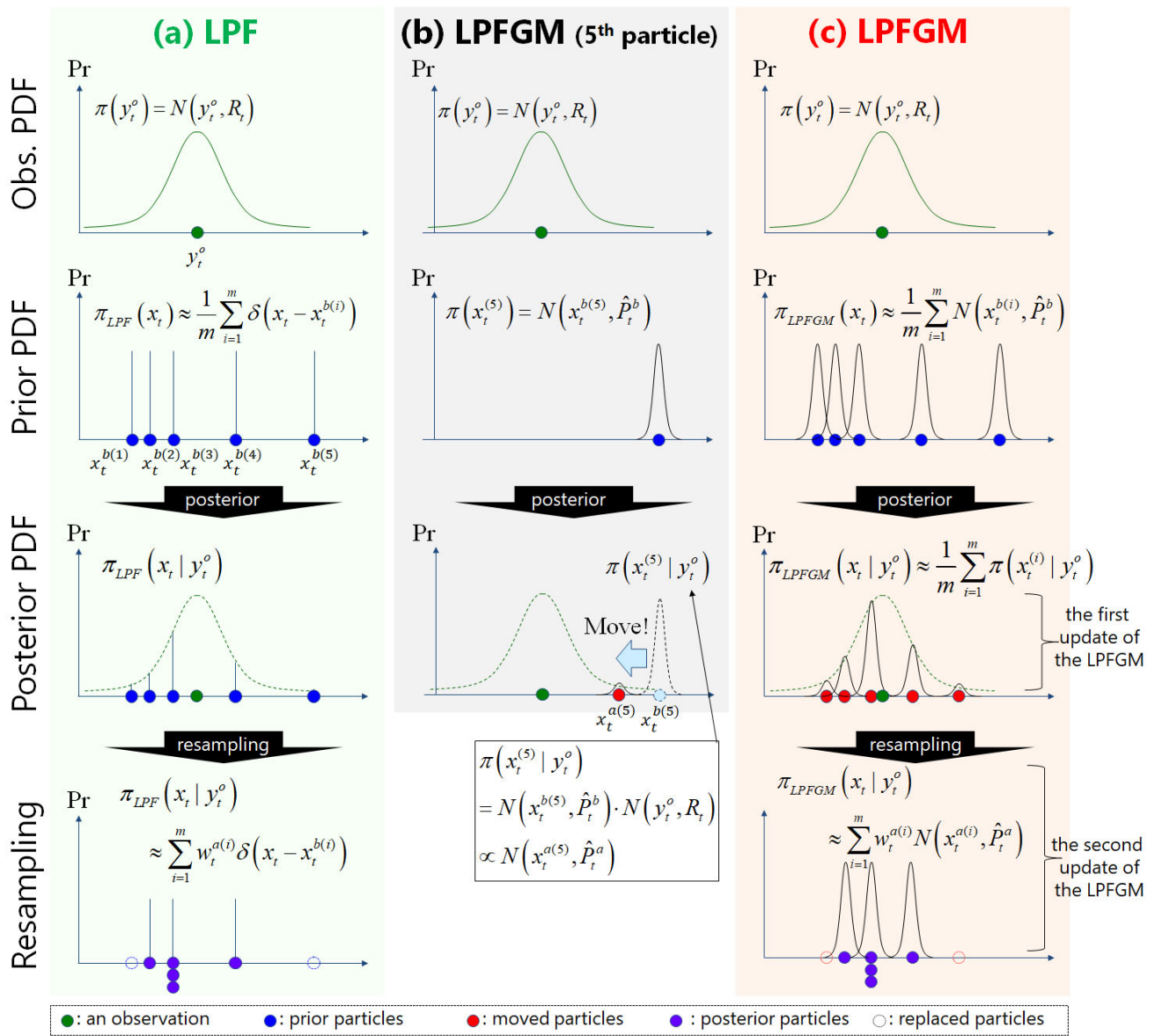



Figure 2: A simple scalar example of observation, prior and posterior PDFs for (a) LPF, (b) 5th particle of LPFGM and (c) LPFGM. Green, blue, red, purple dashed circles represent an observation and prior particles, moved particles, posterior particles, and replaced particles respectively.

760

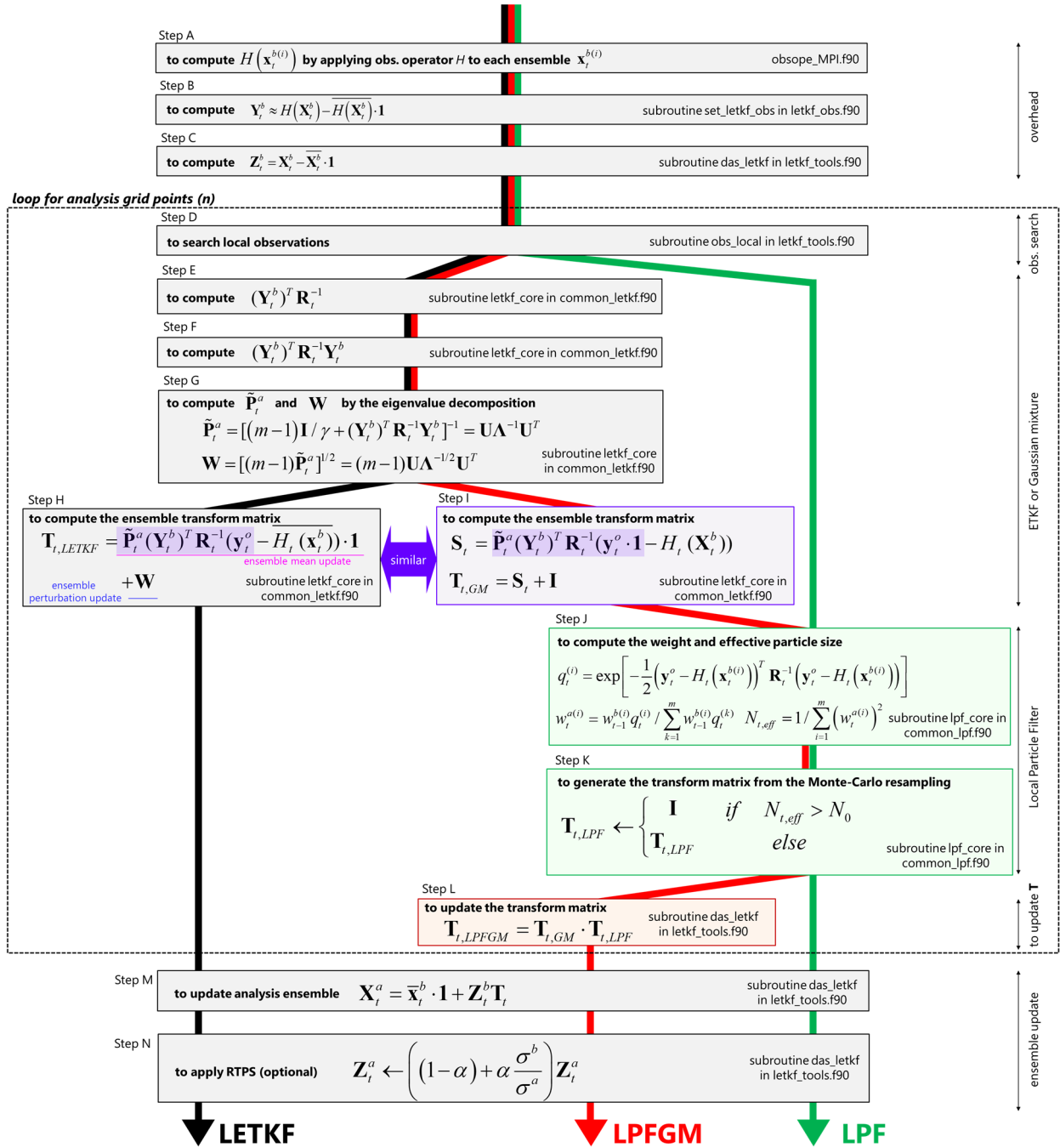



Figure 3: Workflows of LETKF, LPFGM, and LPF. Gray, purple, green, and red boxes represent the components of LETKF, a similar component as in the LETKF, the components of the LPF, and a unique component of the LPFGM, respectively. Subroutines, source files, and computational complexities are also described in boxes, where n, m, p, and p' denote the model dimension, ensemble size, number of observations, and number of local observations at analysis grid points, respectively.

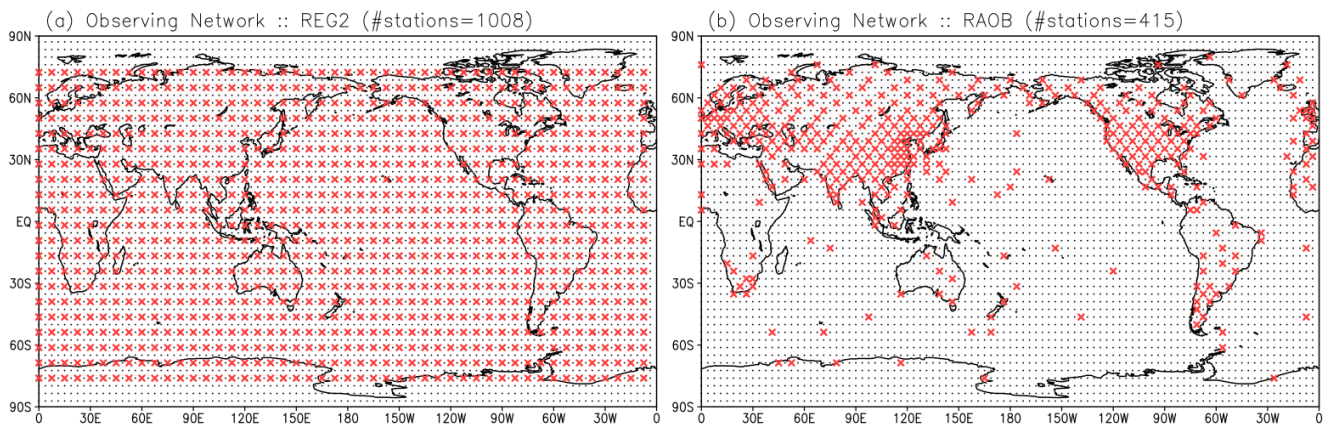



Figure 4: Observing networks for (a) REG2 and (c) RAOB experiments. Small black dots and red crosses represent model grid points and observing points, respectively.

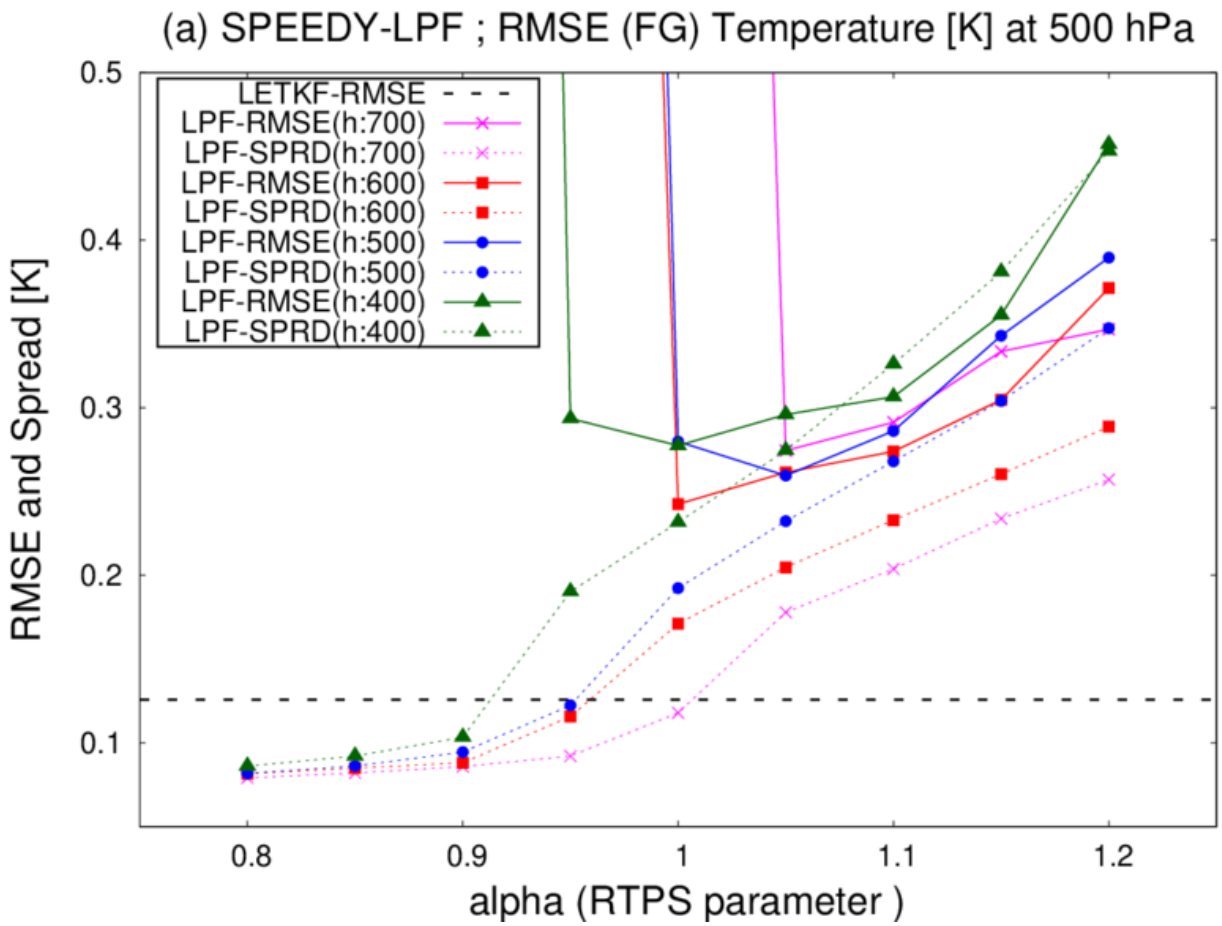



Figure 5: Time-mean background RMSEs (solid lines) and ensemble spreads (dashed lines) for T (K) at the fourth model level
(~500 hPa) as a function of RTPS parameter α averaged over three months of the third year (February-April) with REG2
observations. Magenta, red, blue, and green lines are LPF experiments with horizontal localization scales of 700, 600, 500,
and 400 km, respectively. Dashed lines represent RMSE of LETKF (0.1257 K) with adaptive multiplicative inflation instead
of RTPS.

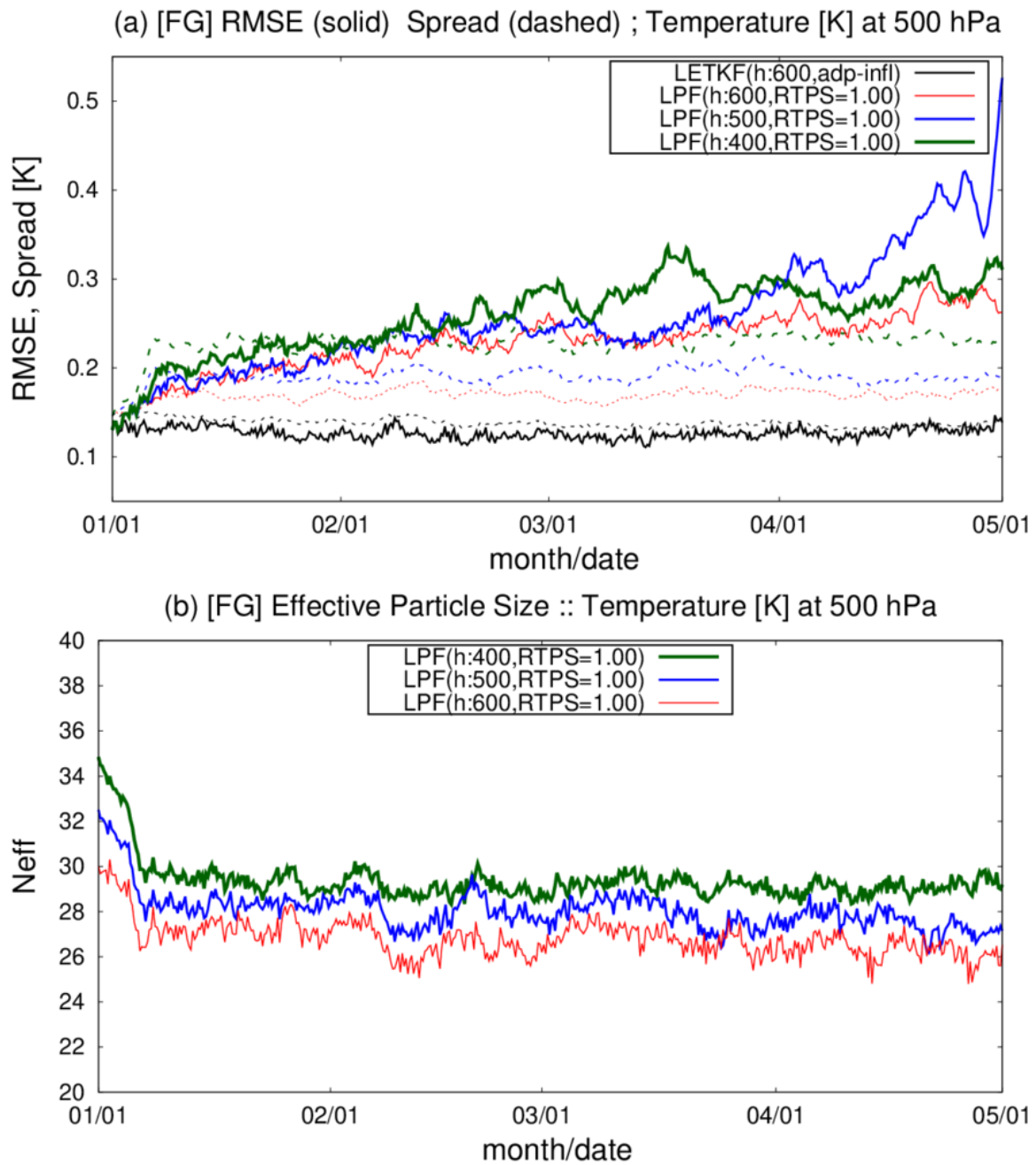



Figure 6: Time series of globally-averaged background (a) RMSEs (solid lines) and ensemble spreads (dashed lines) for T (K) and (b) effective particle size N_{eff} at the fourth model level (~500 hPa) with REG2 observations. Black lines show the LETKF. Red, blue, and green lines are the LPF experiments with localization scales of 600, 500, and 400 km, respectively. RTPS parameter α is set to 1.00 in all three LPF experiments. The abscissa shows month/day.

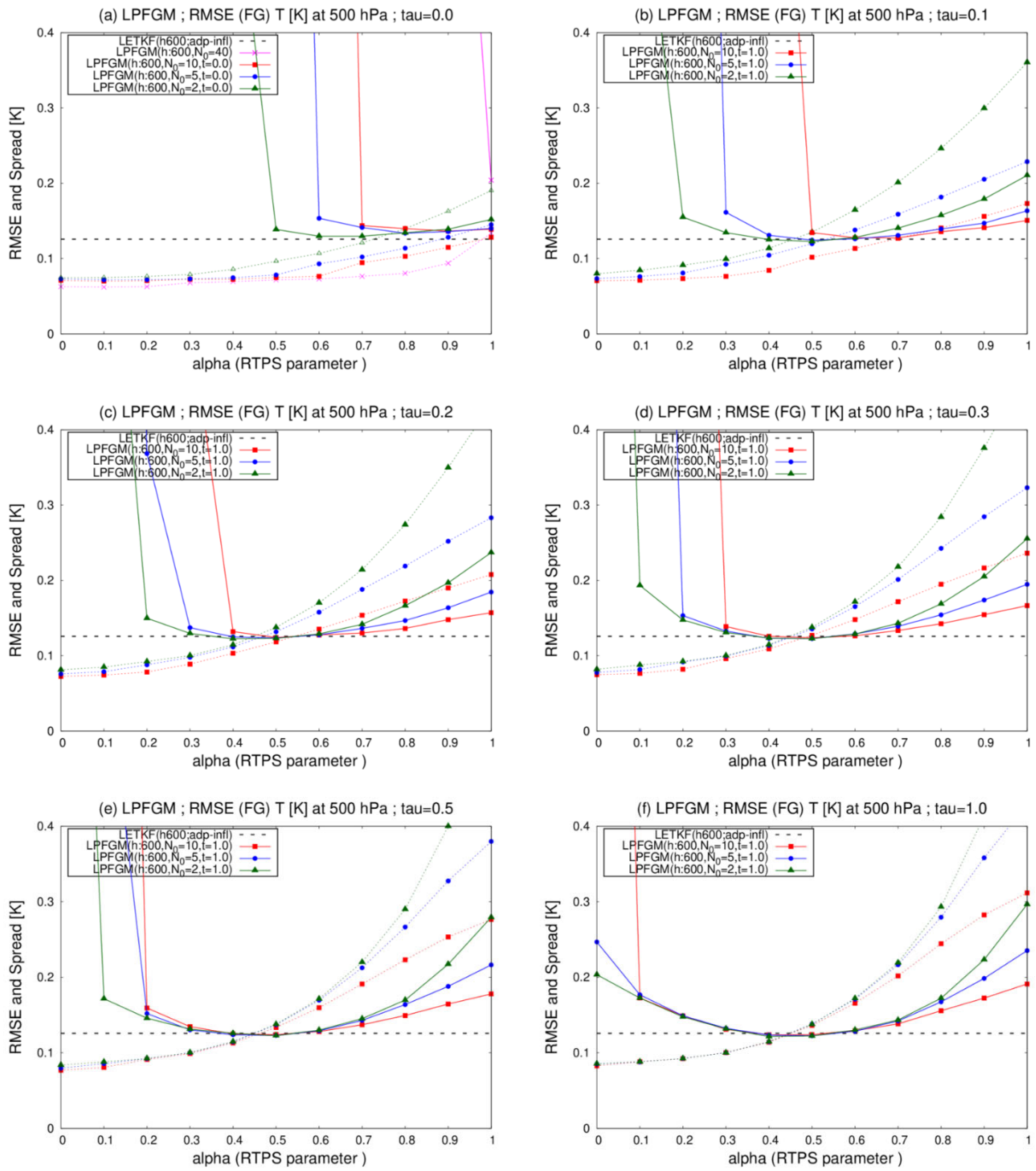
780

785

Figure 7: Similar to Figure 5, but for the LPFGM with REG2 and 600-km localization scale. Magenta, red, blue, and green lines show the cases with N_0 = 40, 10, 5, and 2, respectively. Dashed lines represent RMSE of LETKF (0.1257 K). Panels show the LPFGM with forgetting factors for (a) τ = 0.0, (b) τ = 0.1, (c) τ = 0.2, (d) τ = 0.3, (e) τ = 0.5 and (f) τ = 1.0, respectively.

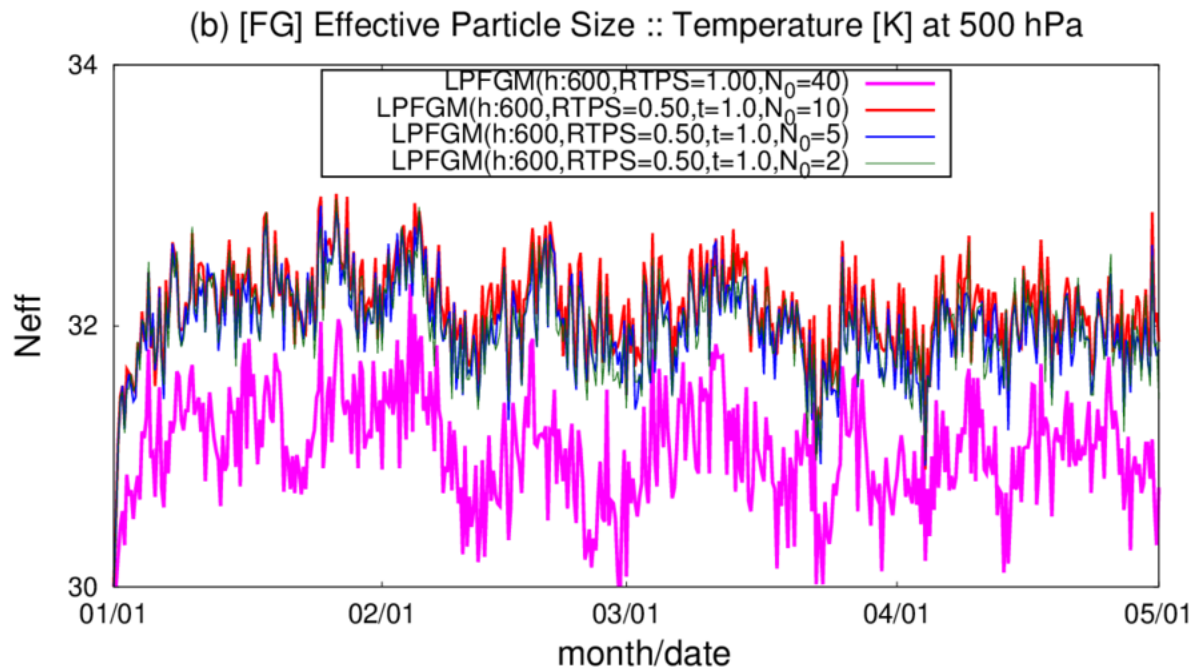
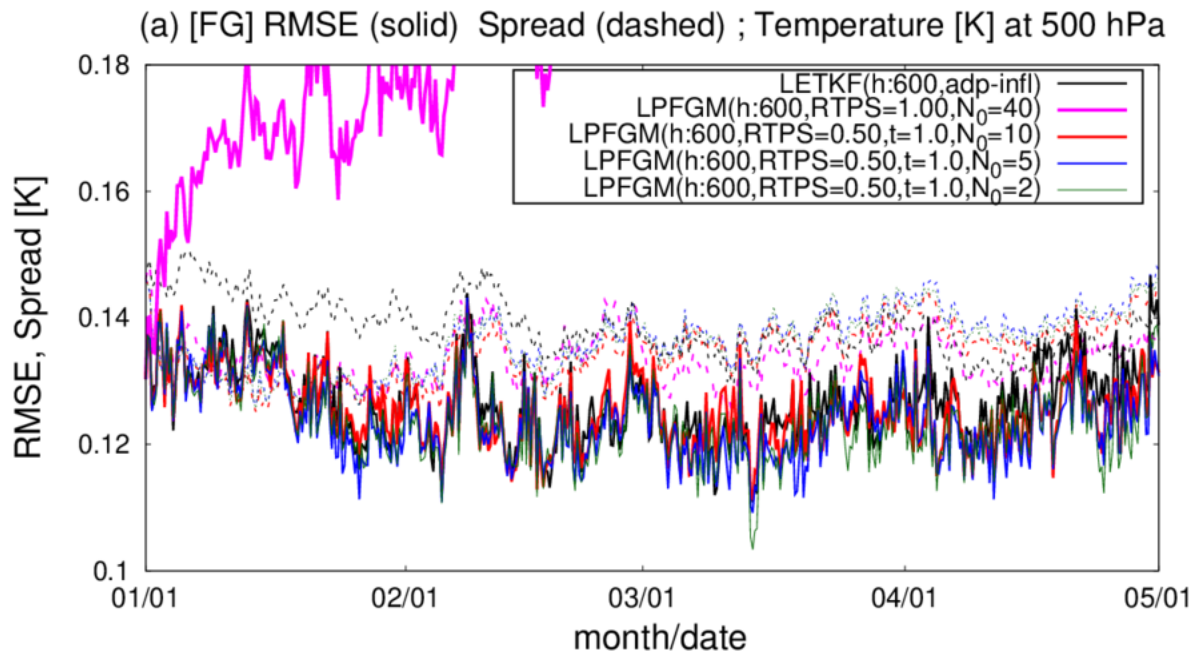



Figure 8: Similar to Figure 6, but showing the LPFGM experiments with REG2 and 600-km localization scale. Magenta, red, blue, and green lines are LPFGM experiments with N_0 = 40, 10, 5, and 2, respectively. Forgetting factor τ is 1.0. The RTPS parameter is set to 1.00 in the experiment with N_0 = 40 (i.e., all-time resampling) and to 0.50 for in other experiments.

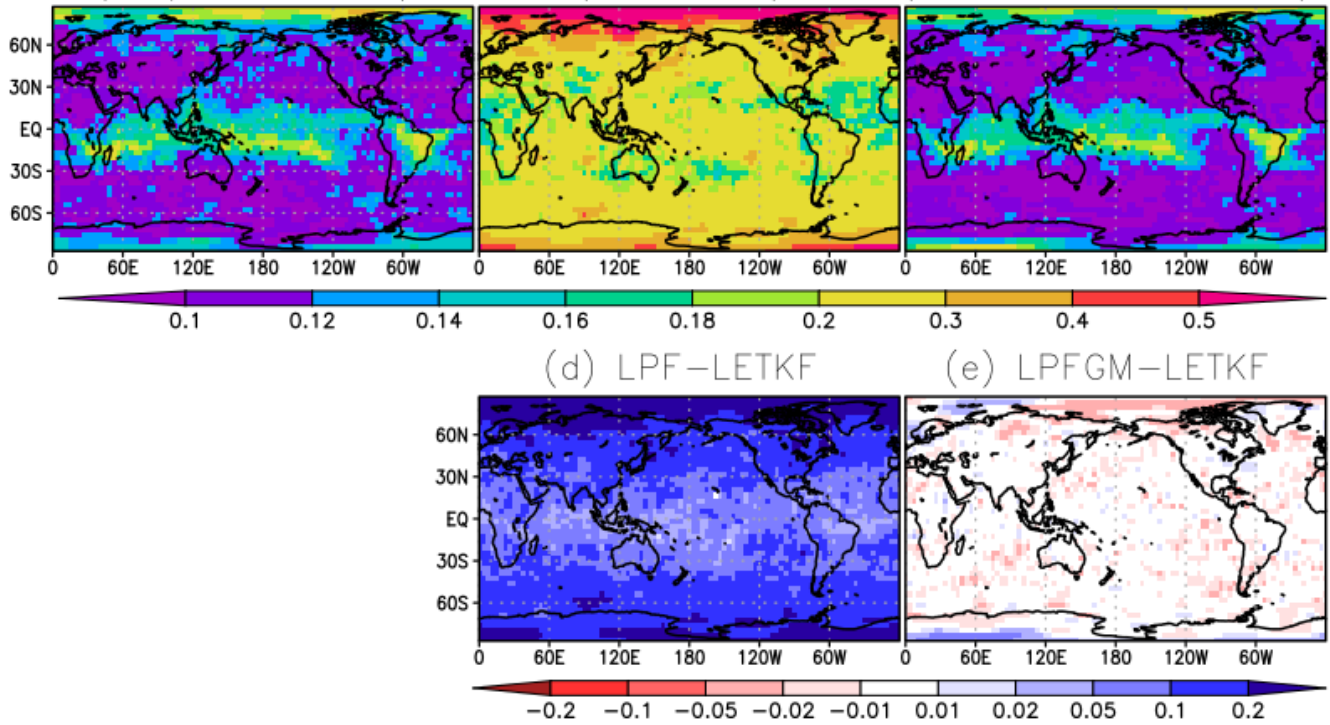



Figure 9: Spatial patterns of background RMSE for T (K) at the fourth model level (~500 hPa) for (a) LETKF, (b) LPF, and (c) LPFGM with best performing localization scale and RTPS parameter, averaged over February-April with REG2 and 600-km localization scale. Panels (d) and (e) show the differences LETKF–LPF and LETKF–LPFGM, respectively. The warm (cold) color indicates that the LPF or LPFGM is better (worse) than the LETKF.

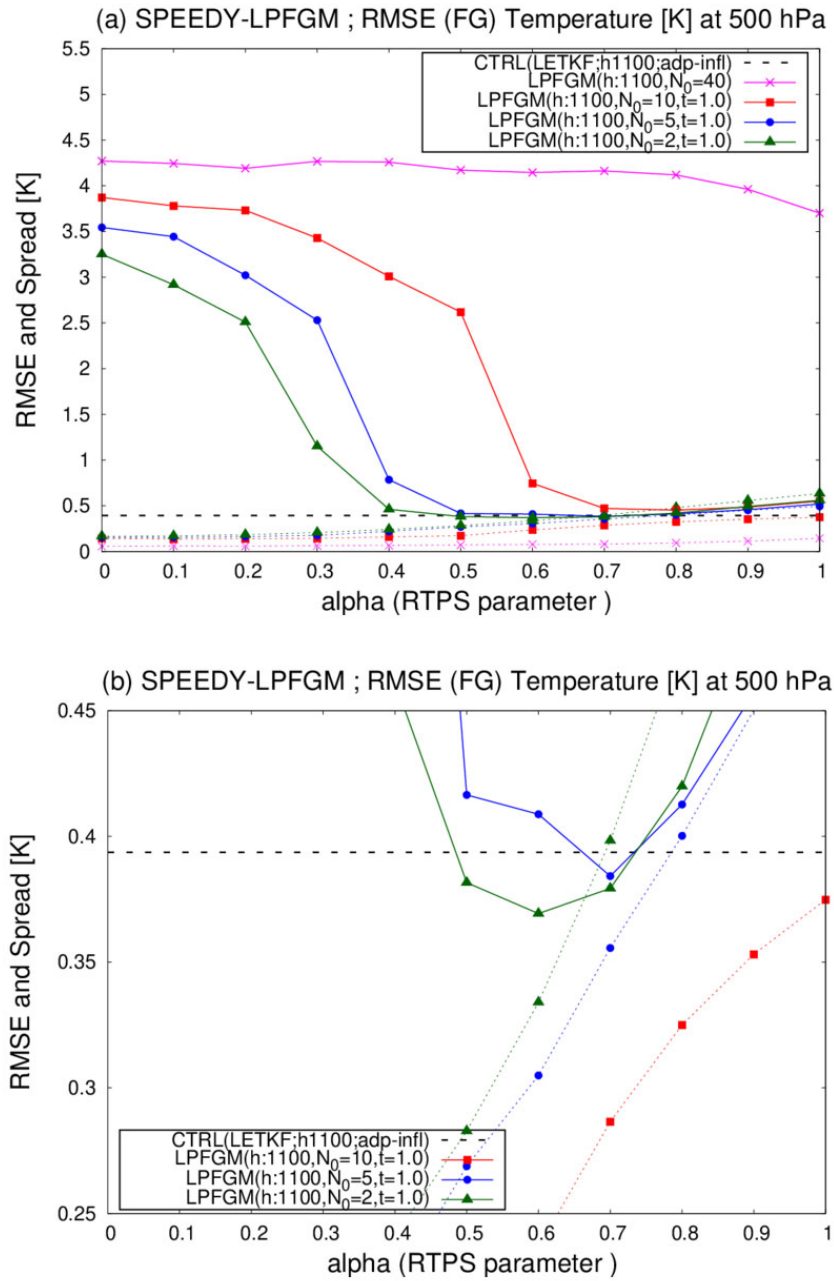



Figure 10: Similar to Figure 5, but showing the LPFGM experiments with the RAOB and 1100-km localization scale. Magenta, red, blue, and green lines are the LPFGM experiments with N_0 = 40, 10, 5, and 2, respectively. Forgetting factor τ is fixed at 1.0. Dashed black lines represent the RMSE of the LETKF (0.4762 K). (b) enlarges (a) for the range of the RMSE of 0.25–0.45.

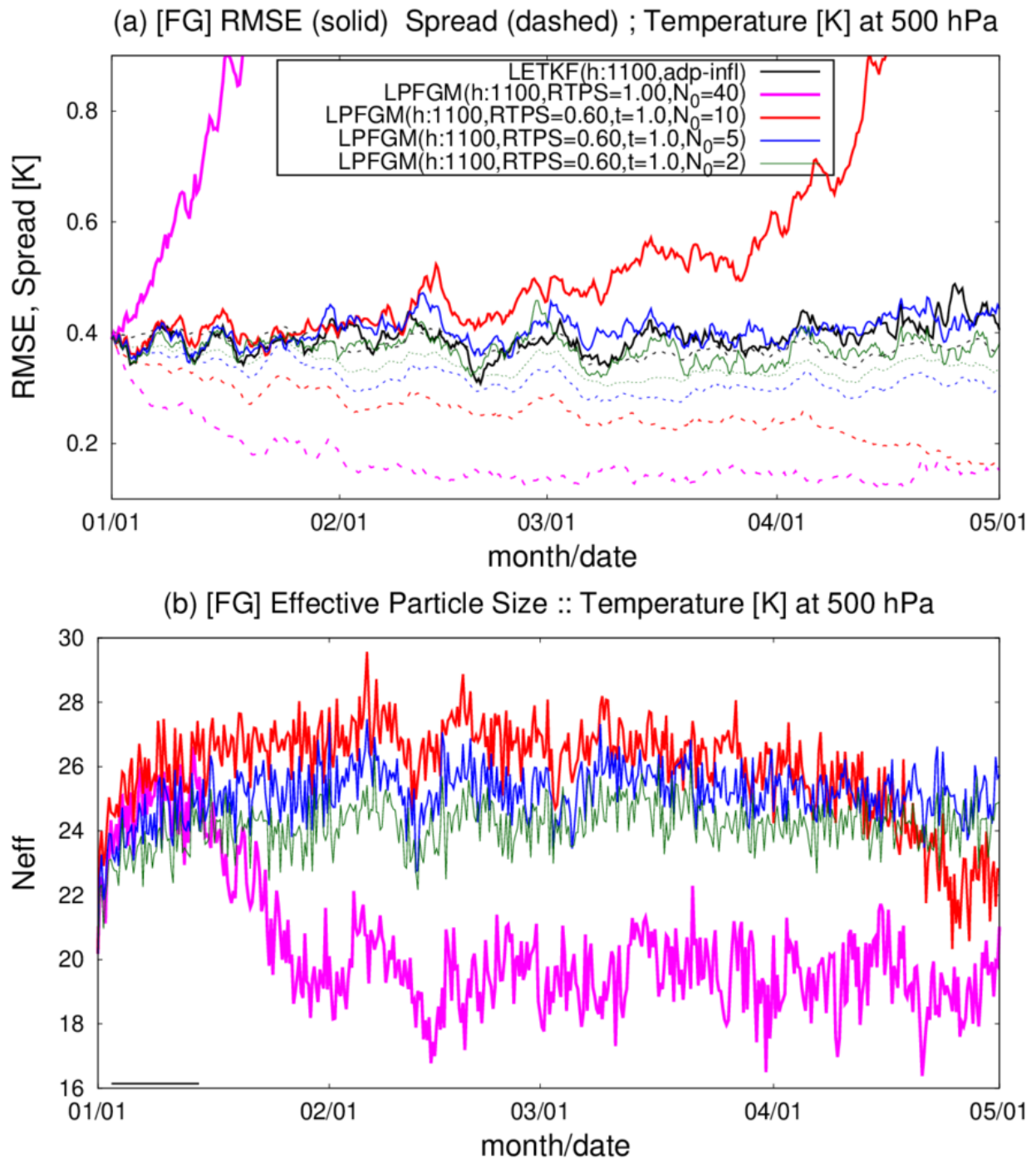
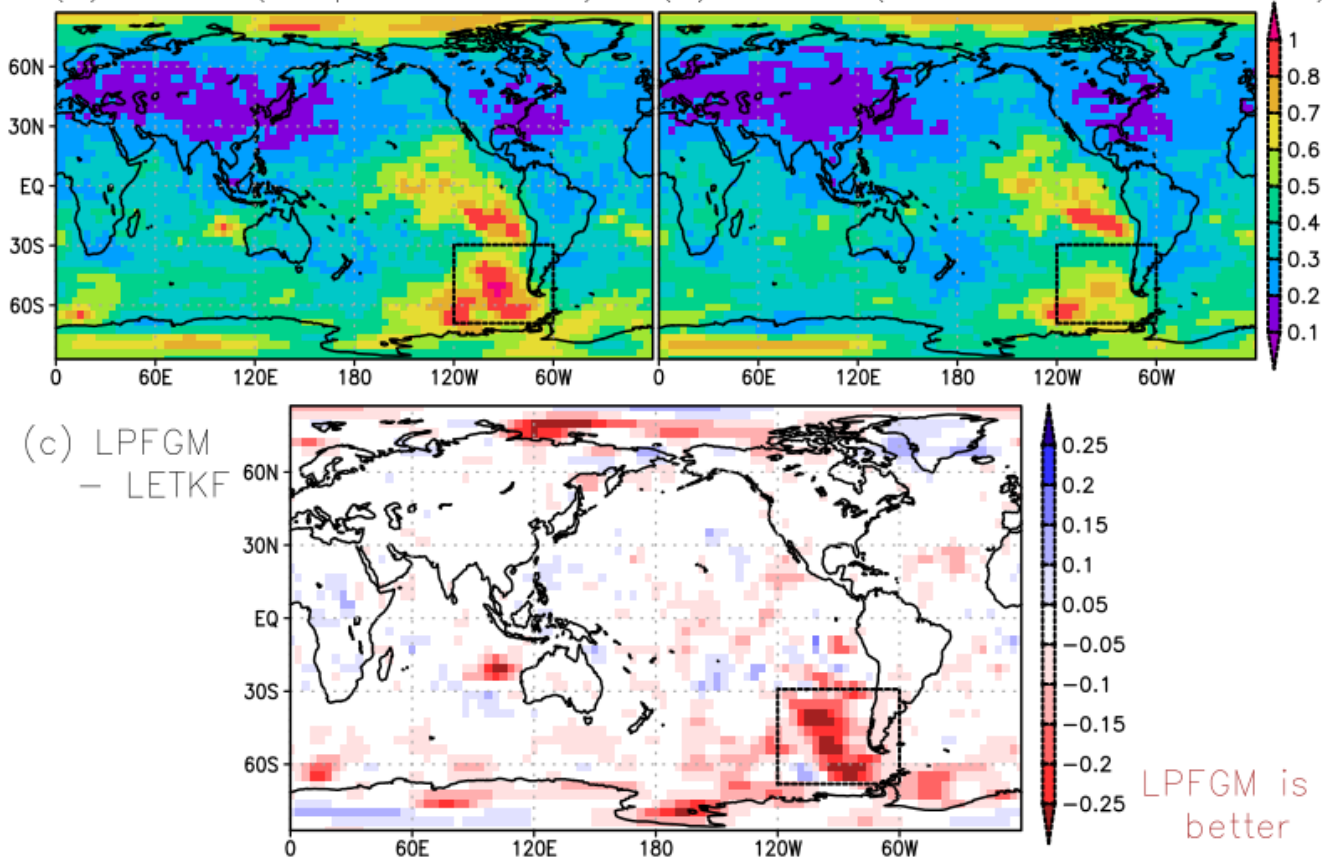
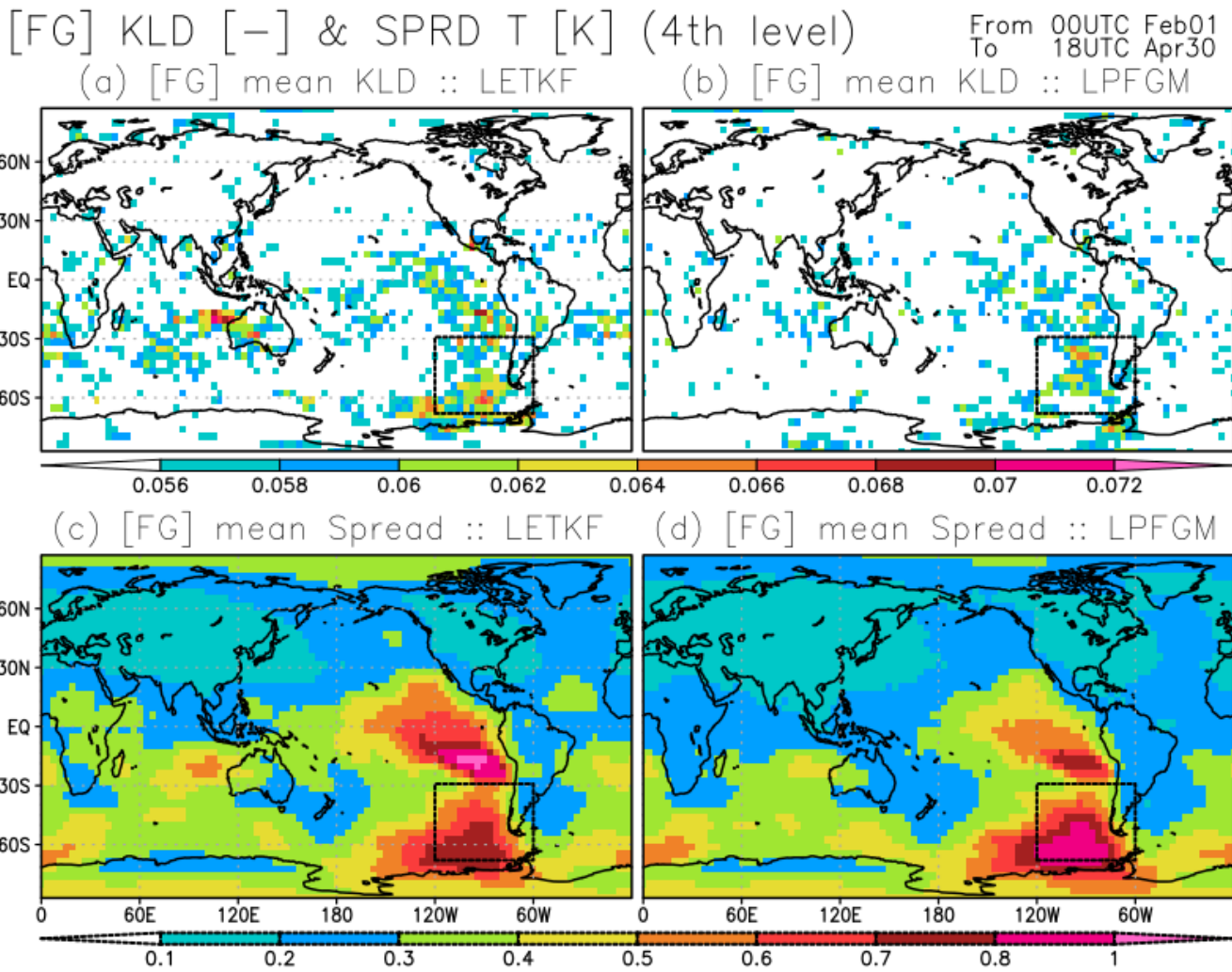



Figure 11: Similar to Figure 8, but showing the LPFGM experiments with the RAOB and 1100-km localization scale. Magenta, red, blue, and green lines are LPFGM experiments with $N_0 = 40$, 10, 5, and 2, respectively. Forgetting factor τ is 1.0.

810 **Figure 12:** Spatial patterns of background RMSE for T (K) at the fourth model level (~500 hPa) for (a) LETKF and (b) LPFGM, averaged over February-April with RAOB and 1100-km localization scale. Panel (c) shows the difference between LETKF and LPFGM. Warm (cold) color indicates that the LPFGM is better (worse) than the LETKF. Black dashed rectangles show the region (120°W-60°W and 70°S-30°S) where difference between LETKF and LPFGM are investigated in Figure 14.

815 **Figure 13:** Spatial distributions of the background (a, b) Kullback–Leibler divergence and (c, d) ensemble spread for T (K), at the fourth model level (~500 hPa) averaged over February-April with RAOB and 1100-km localization scale. (a, c) and (b, d) show the LETKF and LPFGM, respectively. Black dashed rectangles show the region (120°W-60°W and 70°S-30°S) where difference between LETKF and LPFGM are investigated in Figure 14.

820

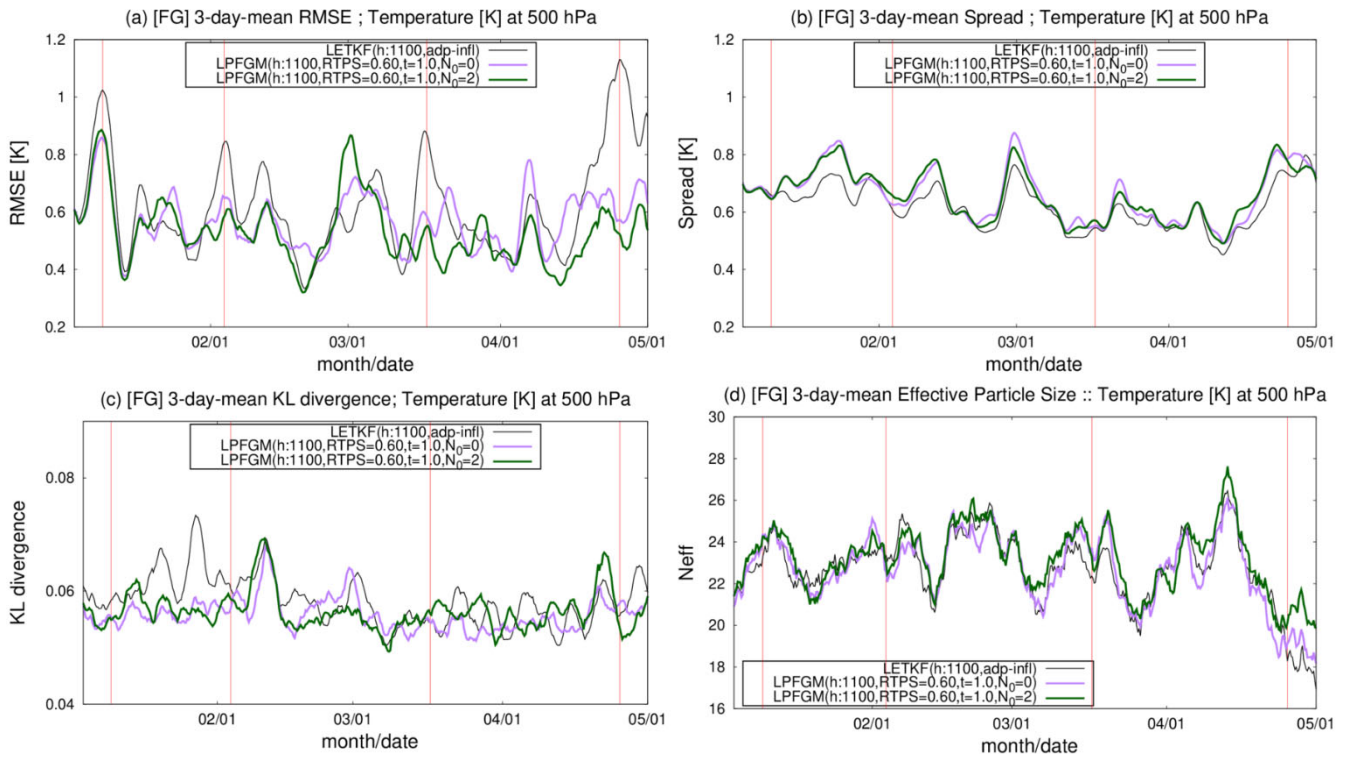



Figure 14: Time series of 3-day-mean (a) RMSE (K), (b) ensemble spread (K), (c) Kullback–Leibler divergence, and (d) effective particle size N_{eff} at fourth model level temperature T, averaged over the region indicated by rectangles in Figs. 12 and 13 (120°W-60°W and 70°S-30°S). Black line represents the LETKF. Green and purple lines are the LPFGM whose resampling frequencies N_0 are 2 and 0, respectively. Vertical red lines represent the cases when the LETKF show large RMSE greater than 0.8 K. In (d), the effective particle size is also computed for the LETKF using the first-guess ensemble.

825

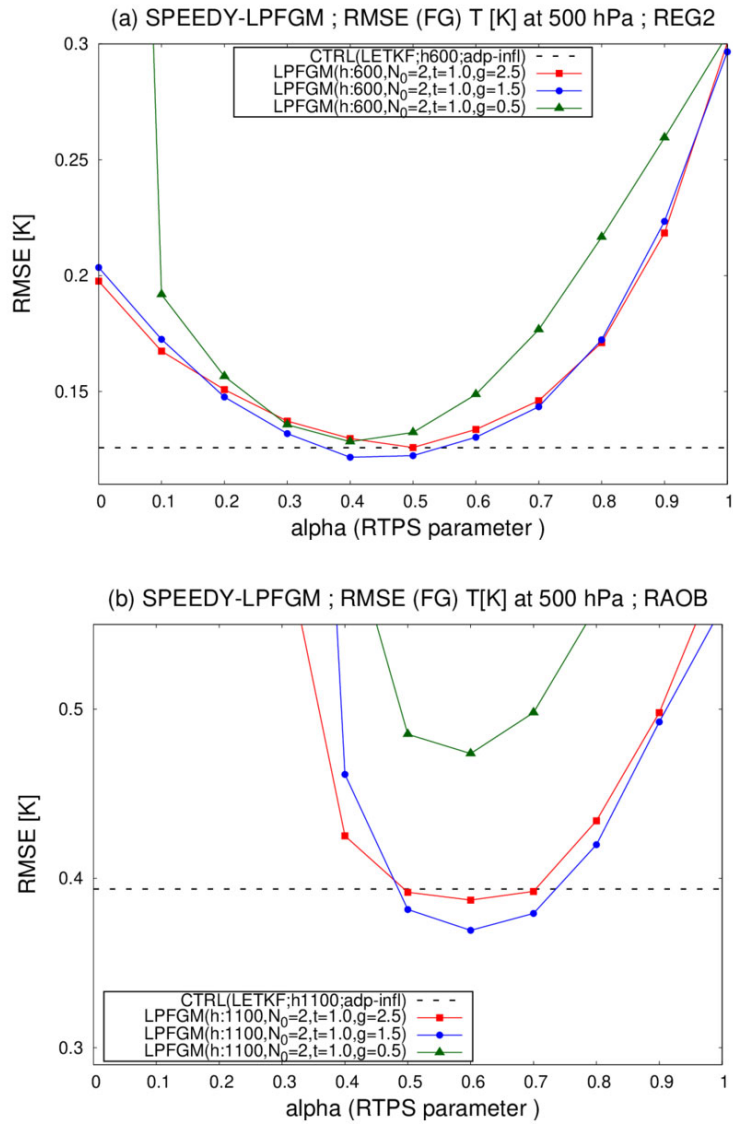



Figure 15: Time-mean background RMSEs for T (K) at the fourth model level (∼500 hPa) as a function of RTPS parameter α averaged over three months of the third year (February-April). (a) shows REG2 with 600-km localization scale, and (b) shows RAOB with 1100-km localization scale. Red, blue, and green lines are LPFGM experiments with being γ=2.5, 1.5, and 0.5, respectively. Black dashed lines represent RMSE of LETKF (0.1257 K) with adaptive multiplicative inflation instead of RTPS. Other parameters are set to be the best performing parameters: $N_0 = 2$ and $\tau = 1.0$.

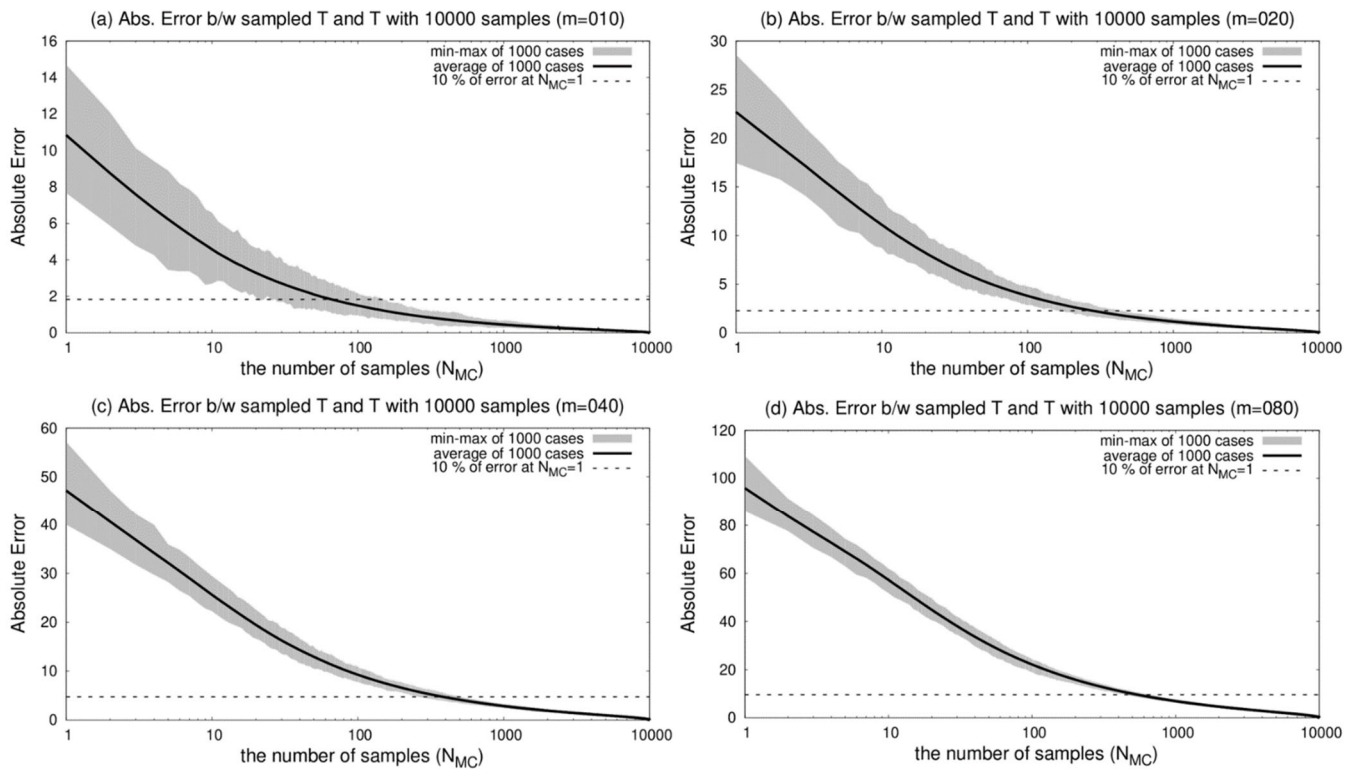



Figure B1: Absolute errors of the transform matrix of MC resampling as a function of the number of samples (N_{MC}) verified
against the transform matrix with 10,000 samples. Absolute errors are computed for 1,000 independent cases that have different
weights generated by uniform random numbers. Black bold lines show the average of 1,000 cases, and gray shades represent
minimum and maximum errors of the 1,000 cases. Panels (a-d) show experiments with 10, 20, 40, and 80 ensembles,
respectively. Dashed line shows the 10 % of initial error at $N_{MC} = 1$.