Sueki *et al.*
Progress in Earth and Planetary Science (2022) 9:47
https://doi.org/10.1186/s40645-022-00504-4

Progress in Earth and Planetary Science

RESEARCH ARTICLE

Open Access

Precision and convergence speed of the ensemble Kalman filter-based parameter estimation: setting parameter uncertainty for reliable and efficient estimation

Kenta Sueki[1,2*], Seiya Nishizawa[1], Tsuyoshi Yamaura[1,3] and Hirofumi Tomita[1,2]

Abstract

Determining physical process parameters in atmospheric models is critical to obtaining accurate weather and climate simulations; estimating optimal parameters is essential for reducing model error. Recently, automatic parameter estimation using the ensemble Kalman filter (EnKF) has been tested instead of conventional manual parameter tuning. To maintain uncertainty for the parameters to be estimated and avoid filter divergence in EnKF-based methods, some inflation techniques should be applied to parameter ensemble spread (ES). When ES is kept constant through the estimation using an inflation technique, the precision and convergence speed of the estimation vary depending on the ES assigned to estimated parameters. However, there is debate over how to determine an appropriate constant ES for estimated parameters in terms of precision and convergence speed. This study examined the dependence of precision and convergence speed of an estimated parameter on the ES to establish a reliable and efficient method for EnKF-based parameter estimation. This was carried out by conducting idealized experiments targeting a parameter in a cloud microphysics scheme. In the experiments, there was a threshold value for ES where any smaller values did not result in any further improvements to the estimation precision, which enabled the determination of the optimal ES in terms of precision. On the other hand, the convergence speed accelerates monotonically as ES increases. To generalize the precision and convergence speed, we approximated the time series of parameter estimation with a first-order autoregression (AR(1)) model. We demonstrated that the precision and convergence speed may be quantified by two parameters from the AR(1) model: the autoregressive parameter and the amplitude of random perturbation. As the ES increases, the autoregressive parameter decreases, while the random perturbation amplitude increases. The estimation precision was determined based on the balance between the two values. The AR(1) approximation provides quantitative guidelines to determine the optimal ES for the precision and convergence speed of the EnKF-based parameter estimation.

Keywords: Parameter estimation, Data assimilation, Ensemble Kalman filter, Atmospheric model, Cloud microphysics, Deep convective system

1 Introduction

The accuracy of numerical weather forecasting and climate projections is affected by various factors. In terms of short-term weather forecasting, accurate initial conditions are produced by utilizing a sufficient volume and spatial coverage of observations, alongside effective data assimilation (DA) methods. Additionally, applying

*Correspondence: kenta.sueki@riken.jp

[1] RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
Full list of author information is available at the end of the article

smaller grid spacing and accurate discretization schemes to fluid dynamics equations reduce the discretization error in atmospheric simulations. Although these efforts are critical in terms of improving forecasting accuracy, they are not sufficient. Atmospheric models parameterize many physical processes, such as cloud microphysics, turbulence, surface flux, radiation, and moist convection. Such parameterizations contain numerous physical and numerical parameters having their own uncertainties, potentially causing severe forecasting errors. For example, Gilmore et al. (2004) showed that uncertainties for intercept parameters and particle densities in a cloud microphysics scheme largely affected the development of convective clouds. Sanderson et al. (2008) showed that the entrainment coefficient in a convection scheme impacted climate sensitivity for general circulation model (GCM) simulations.

Ideally, physical model uncertainties should be solved by improving the model itself based on an understanding of actual physical processes, and determining model parameters based on experiments and measurements. However, to reduce prediction error, we often "tune" parameters in the existing physical model instead of improving model equations. The conventional parameter tuning method involves parameter sensitivity experiments and comparing each performance for different parameter sets (Mauritsen et al. 2012; Hourdin et al. 2017). This is a meaningful procedure as researchers can directly check the impact of varying parameter values on the prediction and evaluate its effects on the climate. It is important as, in general, the time scale of climate is quite longer than the time scale of physical processes directly governed by the parameters. However, such manual parameter tuning is limited in terms of its evaluation efficiency. As optimal parameter values in weather prediction vary depending on the region, season, and phenomenon of interest, it is not practical to manually tune parameters for various conditions. As such, an advanced optimization method that automatically processes numerous observational data and quantitatively incorporates this information into a physical model is more advantageous than tuning fixed values.

To date, many DA methods have been proposed, such as the four-dimensional variational method (4DVar; Sasaki 1969; Thompson 1969), ensemble Kalman filter (EnKF; Evensen 1994), and four-dimensional ensemble-variational method (4DEnVar; Buehner et al. 2013). These methods have mainly been used to produce the most likely initial conditions, incorporated into numerical weather forecasting systems (e.g., Houtekamer et al. 2005; Lien et al. 2017; JMA 2019; NOAA 2019; ECMWF 2020). These methods have also been applied to parameter optimization for physical models (Ruiz et al. 2013a). Zhu and Navon (1999) attempted to optimize model parameters in a global spectral model using 4DVar with the adjoint model; they reported that optimized parameters reduced the forecast error over a few days. Aksoy et al. (2006b) tested the parameter estimation in a two-dimensional (2D) sea breeze model using the ensemble square root filter (EnSRF; Whitaker and Hamill 2002). They were successful in estimating the true parameter values of the nature run. Aksoy et al. (2006a) also estimated a vertical eddy mixing coefficient in a non-hydrostatic mesoscale model. Tong and Xue (2008) successfully estimated the parameters of a cloud microphysics scheme by applying EnSRF to convective-scale radar data assimilation. Ruiz et al. (2013b) tested the estimation of convective scheme parameters incorporated into a GCM using an ensemble transform Kalman filter (ETKF; Bishop et al. 2001). Kotsuki et al. (2018) optimized the parameter of a convective scheme by applying the ETKF to global precipitation data assimilation. Recently, Kotsuki et al. (2020) attempted to optimize a spatially varying parameter field by using a local ensemble transform Kalman filter (LETKF; Hunt et al. 2007). Ruckstuhl and Janjić (2020) tested the estimation of the roughness length in a convection-permitting model using the LETKF; they showed estimating the roughness length led to the improvement of cloud and precipitation forecasts.

In the EnKF-based parameter estimation, perturbed parameter values are assigned to each ensemble member to artificially generate parameter uncertainty. Unlike state variables such as wind, temperature, and moisture, the parameter values generally remain constant during the forecast and are only updated during the analysis step. The parameter ensemble spread (ES), defined as a standard deviation of the parameter ensemble, is a key metric of parameter uncertainty. This ES also remains constant during the forecast; as such, the updated analysis ES for the parameter is always smaller than its previous one. This leads to filter divergence, in which the parameter ES gradually decreases, and the parameter value finally loses its sensitivity to observations. To avoid filter divergence, previous studies have set a threshold value for the standard deviation of the parameter ensemble, where the standard deviation was adjusted back to the threshold when it fell below the threshold (e.g., Aksoy et al. 2006a, b; Tong and Xue 2008). Kotsuki et al. (2018) applied an inflation technique to parameter ensemble at every analysis step to ensure that the parameter ES was constant.

Thus, the issue centers on determining the optimal threshold or fixed value of the parameter ES. Tong and Xue (2005) stated that an excessively large ES generates oscillations around the true value, reducing estimation precision. By contrast, an ES that is too small leads to slow convergence, while the use of further smaller ES can

cause filter divergence. Kotsuki et al. (2018) showed that the estimation reaches a specific value using different ES values for the estimated parameter; however, it takes longer to converge to that value with a smaller ES. These studies suggest that an ES that is not too large nor too small is appropriate for precise and rapidly converging parameter estimation. A possible compromise is to use larger ES for the first few cycles prioritizing fast convergence and then to use smaller ES expecting high precision like Aksoy et al. (2006a, b) and Tong and Xue (2008). In this case, however, the issue about how we set the lower limit of ES value to achieve high precision remains.

Ruiz et al. (2013b) proposed a method to estimate the optimal ES for each analysis step. In their method, the inflation factor for the ES of the parameter was determined based on the analysis error covariance matrix in the ensemble space, including information on state variables and parameters. Ruiz et al. (2013b) stated that regardless of the initial ES, the method can optimize the parameter ES for each analysis step depending on model sensitivity to parameters. While this method provides optimized parameter ES, there is some spin-up time until the ES reaches the optimal value. This is unlikely to be problematic when estimating parameters in GCMs during the integration time of several tens of days, as carried out by Ruiz et al. (2013b). However, this could be fatal if, for example, we want to estimate parameters in cloud microphysics schemes by assimilating radar reflectivity data for less than an hour. Therefore, a suitable ES value is required for the ensemble from the initial time to estimate parameters effectively. Ordinarily, model sensitivity to each parameter is unknown a priori, making it unclear what constitutes a suitable ES value. To establish a reliable and effective estimation method, there is a need to advance the current understanding of how the parameter ES affects the precision and convergence speed of the estimation.

This study aims to clarify how the precision and convergence speed varies depending on the prescribed constant ES. This will help develop a procedure to determine the optimal ES that provides precise and rapidly converging estimation. An observing system simulation experiment (OSSE) which estimates a cloud microphysics parameter with the assimilation of radar reflectivity data was conducted. Different constant ES values were set for an estimated parameter, and we investigated the dependence of precision and convergence speed on the ES value. Finally, we approximated the estimation time series using a simple mathematical model to quantify the precision and convergence speed of the parameter estimation. We also checked the accuracy of parameter estimation because, in general, there is no guarantee that the true value can be estimated accurately in the OSSE. Note that accuracy and precision are different metrics: The accuracy means how close the estimated value is to the true value; the precision means how close the estimated values for different trials are to one another. Thus, it is possible that the estimation shows high precision but poor accuracy due to some systematic errors in the estimated values.

We focused on cloud microphysics parameters as they significantly affect the prediction of moist convective phenomena. Deep moist convection often results in severe weather, such as heavy rain and strong wind gusts. The optimization of cloud microphysics parameters may help preserve life and property by reducing the forecasting error of severe storms and providing accurate weather information. The direct assimilation of convective cloud observations (e.g., radar reflectivity) continues to be challenging because of the possible discrepancy for convection appearance between a numerical model and nature. A quantitative understanding of the precision and convergence speed of parameter estimation for cloud microphysics is essential to expanding the current knowledge on convective-scale data assimilation based on the EnKF.

This paper has been organized into four additional sections. Section 2 describes the methodology of our experiment, Sect. 3 presents the results, Sect. 4 discusses the results, and Sect. 5 presents the conclusions of this study.

2 Methodology

2.1 Overview

This study carried out OSSE within the framework of a perfect-model twin experiment. The nature run was an idealized squall-line simulation using a bulk cloud microphysics scheme with fixed parameter values. One of the microphysics parameters in the nature run was considered the true value and a target of parameter estimation. Figure 1 shows the schematic of the experiment; the DA cycle using the LETKF had commenced from $T=3$ h of the nature run. The initial fields of the ensemble simulation were generated by adding random perturbations to

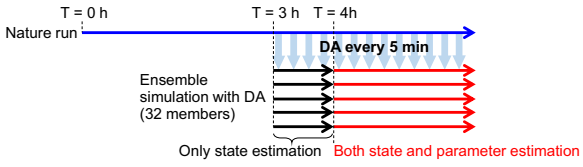

Fig. 1 Schematic of the experiment. First, the nature run was performed (horizontal blue arrow). State estimation using the LETKF with 32 ensemble members commenced at $T=3$ h (black arrows). Pseudo-radar reflectivity fields calculated from the nature run were assimilated into the ensemble simulation every 5 min. Parameter estimation commenced at $T=4$ h (red arrows)

the state fields of the nature run at $T=3$ h. Pseudo-radar reflectivity fields calculated from the nature run were used as observations for the DA cycle. During the first 1 h DA cycle, only the state estimation was conducted without parameter estimation, where the ensemble used the same parameter value as the true value of the nature run. During this period, the ES of state fields gradually increased and then became steady. Following this, perturbation was added to the target parameter in the ensemble and parameter estimation commenced from $T=4$ h of the nature run.

2.2 Nature run

A 2D squall-line simulation was carried out to produce the nature run using a non-hydrostatic atmospheric model of the Scalable Computing for Advanced Library and Environment-Regional Model (SCALE-RM; Nishizawa et al. 2015; Sato et al. 2015) version 5.3.3. The horizontal and vertical grid spacing was 5 km and 250 m, respectively. The size of the calculation domain was 2160 and 17.5 km in the horizontal and vertical directions, respectively. The integration time interval for the advection scheme was 2.5 s. Periodic conditions were applied horizontally with 100 km buffer regions from lateral boundaries for potential temperature, wind speed, and water vapor mixing ratio. In the buffer regions, gravity waves and cold air outflow by the squall line were damped. A six-class single-moment microphysics scheme (Tomita 2008) was utilized to simulate moist convective processes, the target of parameter estimation. Sub-grid scale turbulence parameterization was not used to avoid estimation complexity, and only numerical diffusion was added for model stability. Analytical thermodynamic and wind profiles proposed by Weisman and Klemp (1982) were used to provide a favorable environment for the development of the squall line. Figure 2 presents the environmental profiles of the potential temperature, water vapor mixing ratio, and horizontal wind speed. The maximum mixing ratio near the surface was set to 14 g kg^{-1}, while the maximum wind speed at the top of the atmosphere was set to 10 m s^{-1}. Figure 3 shows the development of the cold pool and snapshots of the pseudo-radar reflectivity. This simulation successfully produced a squall line that endured for more than one day.

2.3 Initial ensemble

The initial ensemble was generated by adding random perturbations to potential temperature, horizontal and vertical wind speeds, and the mixing ratios of water vapor and hydrometeors of the nature run at $T=3$ h. Each perturbation was generated by summing 2D Gaussian kernels whose centers were at each model grid, and horizontal and vertical width (1σ) was 5 km and 250 m, respectively. The amplitude of the Gaussian kernel was randomly sampled from a normal distribution with a zero mean and a specific standard deviation. The standard deviation of the kernel amplitude differed for each variable: 0.2 K for potential temperature, 0.2 m s^{-1} for wind speed, and 0.06 g kg^{-1} for mixing ratio. Note that perturbations were added only to a specific domain; 400 km

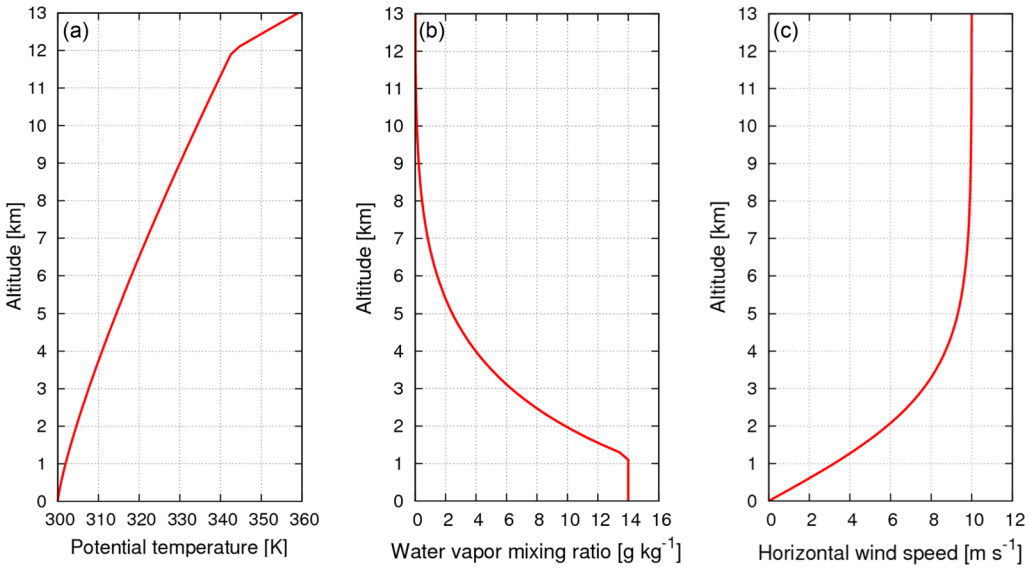

Fig. 2 Environmental profile for squall-line simulation: **a** potential temperature; **b** water vapor mixing ratio; and **c** horizontal wind speed

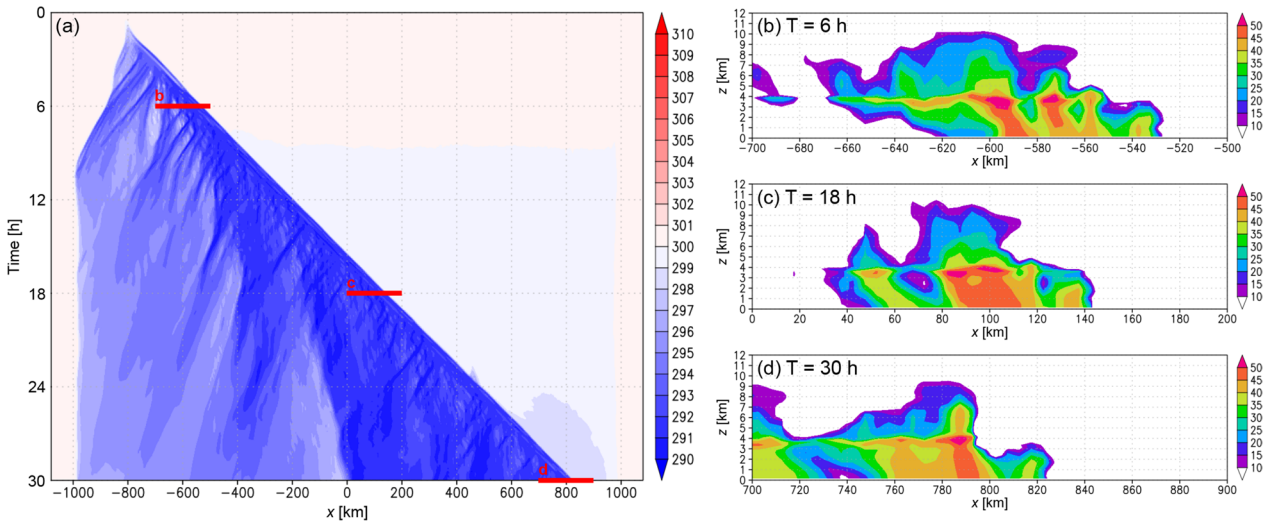

Fig. 3 Cold pool and radar reflectivity of the simulated squall line in the nature run: **a** Hovmöller diagram of potential temperature at the lowest model level; **b**–**d** snapshots of radar reflectivity, where (**b**) $T=6$ h; (**c**) $T=18$ h; and (**d**) $T=24$ h. The three red line segments in (**a**) correspond to the times and locations of (**b**–**d**)

horizontally around the squall line and 10 km vertically from the surface.

2.4 State and parameter estimation

The SCALE-LETKF system (Lien et al. 2017) which incorporates the DA system using the LETKF into the SCALE-RM was used. We performed the simultaneous estimation of state variables and a parameter with different localization scales; the ensemble size was 32. Pseudo-radar reflectivity fields obtained from the nature run were assimilated into the SCALE-RM ensemble simulation every 5 min. Model grids with a reflectivity of ≥ 10 dBZ in the nature run were considered precipitation data, while those < 10 dBZ were considered non-precipitation data and replaced with 5 dBZ uniformly prior to assimilation. Radar reflectivity observation error was assumed to be 1 dBZ. Note that precipitation data in the nature run were only assimilated if at least half the ensemble members (16 members) at corresponding grids also experienced precipitation. In contrast, non-precipitation data were assimilated if there is even one member experiencing precipitation to remove precipitation that does not exist in the nature run from the ensemble simulation. Reflectivity data at grids that did not meet these conditions were not assimilated. The localization length scales for state estimation were 4 and 2 km in the horizontal and vertical directions, respectively. The model dynamics, physics, and domain in the SCALE-LETKF system were the same as the nature run. During the first 1 h of simulation (i.e., 12 DA cycles), the ensemble used common parameter values with the nature run. During this period, only state variables were estimated, such as wind speed, temperature, pressure, water vapor, cloud water, rain, cloud ice, snow, and graupel. The inflation technique, known as the relaxation to prior spread (RTPS) method (Whitaker and Hamill 2012), was used with a fixed relaxation parameter of 0.95 for the state estimation. Within this 1 h integration, the ES for the radar reflectivity field gradually increased and reached an almost steady state (Fig. 4). We did not perform a long spin-up period before starting DA but the grows of spread for every 5 min seem to be large enough to increase the spread gradually before starting the parameter estimation. The parameter estimation for the microphysics scheme commenced at $T=4$ h of the nature run (Fig. 1). Unlike the state estimation, a non-localized ETKF was applied to estimate a spatially uniform parameter; this was a LETKF with an infinite localization scale. For the parameter ensemble, a relaxation parameter of one was utilized for the RTPS inflation to maintain the ES as a constant parameter throughout the experiment. This was the same approach as adopted by Kotsuki et al. (2018).

2.5 Parameter to be estimated

The estimation target was a parameter in a six-class single-moment microphysics scheme (Tomita 2008). This scheme predicts mass exchange among six categories of water substances: water vapor, cloud water, rain, cloud ice, snow, and graupel. The mass exchange processes consist of saturation adjustment, auto-conversion, accretion, evaporation, sublimation, deposition, melting, freezing, and the Bergeron process. These processes contain many

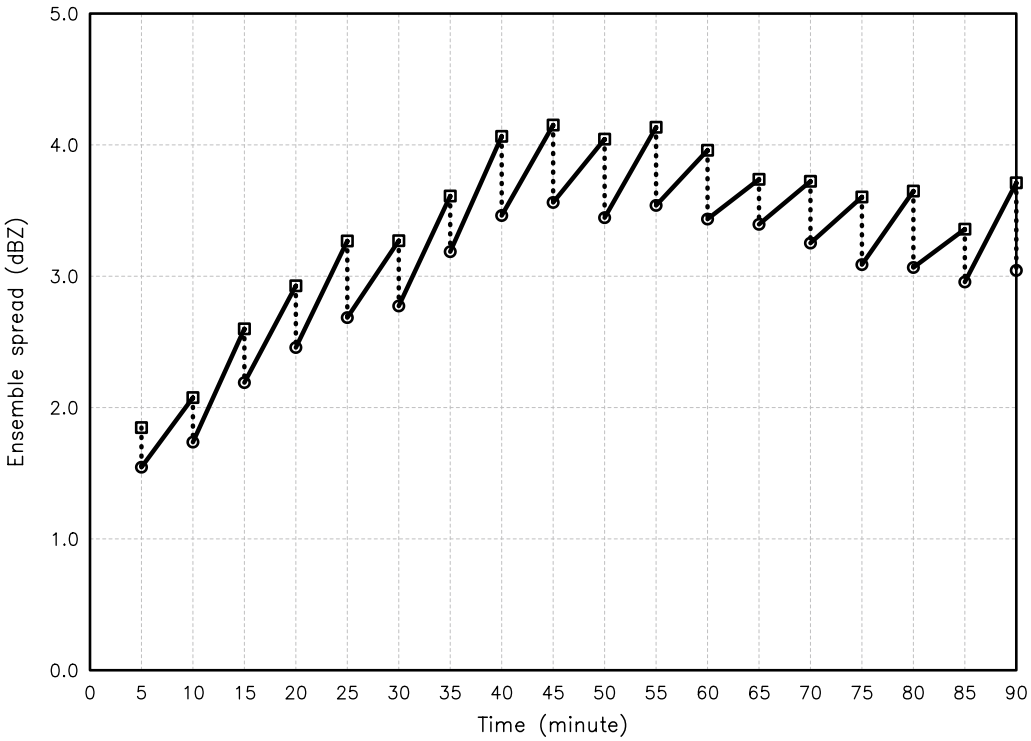

Fig. 4 Time series of the analysis ensemble spread for radar reflectivity. The ensemble spread was defined as the mean sample standard deviation of the radar reflectivity averaged over model grids where the observations to be assimilated existed. Open squares and open circles indicate the spreads for the first guess and the analysis, respectively. The origin of the time axis corresponds to $T=3$ h for the nature run

tunable parameters, such as intercept and slope parameters for the Marshall–Palmer distribution of precipitating hydrometeors (Marshall and Palmer 1948), terminal velocity coefficients for these hydrometeors, collection efficiencies of accretion processes, and parameters controlling auto-conversion rates. As this study investigates the precision and convergence speed of parameter estimation in idealized settings, we selected parameters that were highly sensitive to observation (radar reflectivity). As such, the terminal velocity coefficient of a raindrop, c_{R}, was selected as the unknown parameter:

$$v_{t\mathrm{R}} = c_{\mathrm{R}}\left(D\frac{\rho_0}{\rho}\right)^{1/2}, \tag{1}$$

where $v_{t\mathrm{R}}$ is the terminal velocity of the raindrop; D is the raindrop diameter; ρ is air density; and ρ_0 is the reference air density of 1.28 kg m^{-3}. In this experiment, the true value of c_{R} is 130 in the nature run. We selected c_{R} as the estimated parameter as it controls the falling speed of rain. If c_{R} becomes larger, the falling speed of rain becomes faster (Eq. (1)), and thus, the area of intense radar reflectivity moves downward more quickly. To prevent c_{R} from becoming a value that is too large or too small to be physically practical, a variable transformation was introduced (as per Eq. (2)), and the variable, x, was estimated and converted back to c_{R} after the estimation:

$$c_{\mathrm{R}}(x) = \frac{1}{2}(c_{\mathrm{R\,max}} + c_{\mathrm{R\,min}}) + \frac{1}{2}(c_{\mathrm{R\,max}} - c_{\mathrm{R\,min}})\tanh x. \tag{2}$$

Kotsuki et al. (2018) also used this transformation. Note that c_{R} was always between c_{Rmin} and c_{Rmax} for any x, where c_{Rmin} was set to zero as a negative c_{R} causes the ascent of raindrops (Eq. (1)). On the other hand, c_{Rmax} was set to 260 such that the transformation of Eq. (2) was symmetric with respect to the true value of $c_{\mathrm{R}}=130$. Figure 5 shows the $c_{\mathrm{R}}(x)$ curve of Eq. (2); in the x space, $x=0$ is the true value corresponding to $c_{\mathrm{R}}=130$ in the nature run.

2.6 Experimental settings

At the initial time of parameter estimation ($T=4$ h in Fig. 1), different parameter values were assigned to each member in the ensemble. Parameter values for each member were randomly sampled from a normal distribution with a mean of μ_{init} and a standard deviation of σ in the x space. Note that the parameter ensemble only satisfied a normal distribution in the x space, and not the c_{R} space as the transformation between c_{R} and x was not

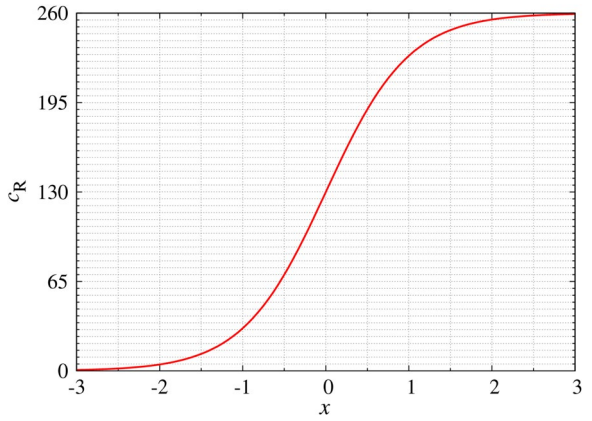

Fig. 5 A variable transformation between c_R and x. The curve of c_R is plotted against x as per Eq. (2)

linear (Eq. (2) and Fig. 5). The parameter perturbation was updated, and the μ of the ensemble was modified at every DA step. An updated μ was considered an estimated parameter value for each step. The standard deviation, σ, is a measure of the parameter ES. As stated in Sect. 2.4, the RTPS method with a coefficient of one was applied to the parameter ensemble; as such, σ was kept constant throughout the estimation.

The time-independent σ value is a key parameter of the experiment. We investigated the dependence of precision and convergence speed of the estimated parameter value (μ) on the parameter ES (σ). Four different ES values were tested: $\sigma=0.05$, 0.1, 0.2, and 0.4 in the x space, corresponding to approximately 6.5, 13, 26, and 49 in the c_R space, respectively, at $\mu=0$ ($c_R=130$). Note that the spread in the c_R space becomes smaller when the c_R value approaches c_{Rmin} or c_{Rmax} even though σ in the x space is kept constant (see Fig. 5). However, to discuss the dependence of precision and convergence speed on the parameter ES as simply as possible, we fixed the parameter ES in the x space throughout the estimation. Under the set σ values, three different initial ensemble means were tested: $\mu_{init}=0$, 0.55, and -0.55 in the x space, corresponding to $c_R=130$, 65, and 195. As the OSSE was a perfect-model twin experiment, it was expected that all experiments may be able to finally attain the true value of the nature run ($x=0$), as per previous studies such as Tong and Xue (2008). Thus, experiments where $\mu_{init}=0$ were expected to provide a statistically stationary time series of μ fluctuating around the true value and illustrate the estimation precision. We also expected that experiments where $\mu_{init}=0.55$ and -0.55 would result in convergence from the wrong values to the true value, illustrating the convergence speed of the estimation. All experimental settings are listed in Table 1.

Table 1 List of experimental settings

Experimental ID	μ_{init}	σ	Initial time
Exp. 1	0 (Cr = 130)	0.05 (ΔCr ~ 6.5)	$T=4$ h
Exp. 2	0	0.1 (ΔCr ~ 13)	$T=4$ h
Exp. 3	0	0.2 (ΔCr ~ 26)	$T=4$ h
Exp. 4	0	0.4 (ΔCr ~ 49)	$T=4$ h
Exp. 5	0.55 (Cr = 195)	0.05	Every 1 h from $T=7$ h
Exp. 6	0.55	0.1	Every 1 h from $T=7$ h
Exp. 7	0.55	0.2	Every 1 h from $T=7$ h
Exp. 8	0.55	0.4	Every 1 h from $T=7$ h
Exp. 9	−0.55 (Cr = 65)	0.05	Every 1 h from $T=7$ h
Exp. 10	−0.55	0.1	Every 1 h from $T=7$ h
Exp. 11	−0.55	0.2	Every 1 h from $T=7$ h
Exp. 12	−0.55	0.4	Every 1 h from $T=7$ h

The left-most column shows the experimental ID; μ_{init} represents the parameter ensemble mean at the initial time; and σ represents the time-independent standard deviation of the parameter ensemble (parameter ensemble spread). The right-most column details the initial time of each experiment corresponding to the simulation time of the nature run.

3 Results

3.1 Overview

Figure 6 shows the time series of the parameter ensemble mean for all experiments listed in Table 1; Fig. 6a–d presents the experiments for $\mu_{init}=0$ (Exps. 1–4), where 50 trials with different initial parameter ensemble perturbations were carried out for each setting. This figure shows that the parameter ensemble mean successfully remains around the true value of the nature run ($x=0$) with some fluctuation; these time series appeared to be statistically stationary. Figures 6a–d clearly illustrate that the fluctuation frequency varies depending on σ: the time series with a small ES fluctuated slowly, while a large ES fluctuated rapidly. The amplitude of the fluctuation was indicative of the precision; the time series with the largest ES ($\sigma=0.4$) fluctuated with the largest amplitude, indicating low precision. The difference in the fluctuation amplitude between $\sigma=0.05$ and 0.2 was unclear, while some extreme values appeared for the experiment of $\sigma=0.2$.

Figure 7 shows the power spectral density of the time series for Exps. 1–4. For the high-frequency range, the spectral amplitude increased with σ. This corresponded

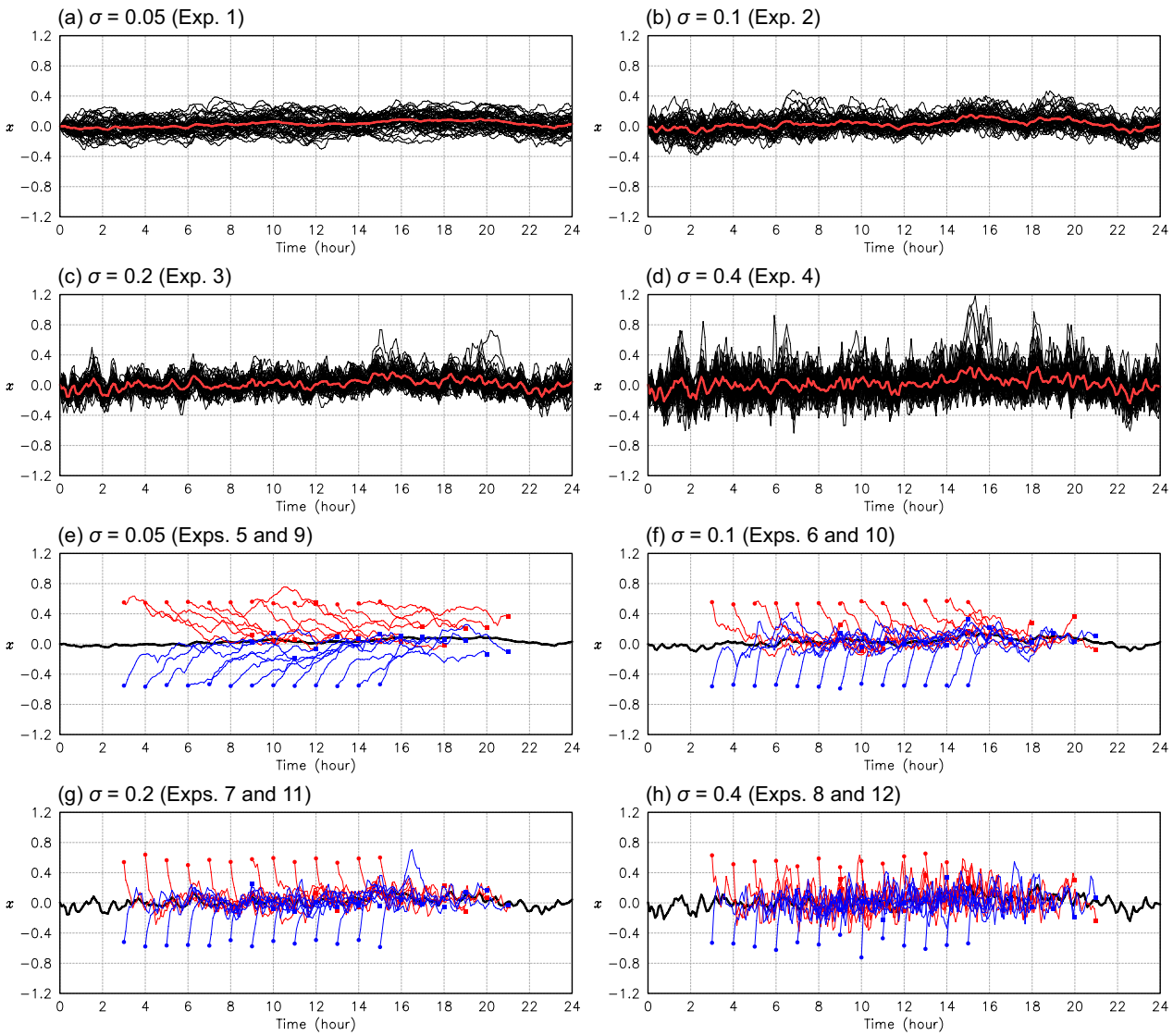

Fig. 6 Time series of parameter ensemble mean for all experiments: **a**–**d** experiments with an initial ensemble mean μ_{init} of 0 (Exps. 1–4). The parameter ensemble spread (σ) for each experiment was: **a** 0.05; **b** 0.1; **c** 0.2; and **d** 0.4. Black lines represent the time series for a total of 50 trials for each setting. Red lines represent averaged time series of all trials. The origin of the time axis corresponds to $T=4$ h for the nature run; **e**–**h** experiments where $\mu_{\text{init}}=0.55$ (Exps. 5–8: red lines), and − 0.55 (Exps. 9–12: blue lines); **e** experiments where $\sigma=0.05$ (Exps. 5 and 9); **f** $\sigma=0.1$ (Exps. 6 and 10); **g** $\sigma=0.2$ (Exps. 7 and 11); and **h** $\sigma=0.4$ (Exps. 8 and 12). Initial times of the experiments were every 1 h from 3 h following the commencement of Exps.1–4. The estimation duration was 6 h for each experiment. The beginning and end points of estimation are denoted by circles and squares, respectively. Black lines represent the averaged time series of Exps.1–4, which are the same as the red lines in (**a**–**d**)

to the frequently fluctuating time series for the large σ, as shown in Fig. 6. In all cases, the spectral amplitude decreased with increasing frequency, showing the "red noise" spectrum characteristics. The red noise spectrum is a feature that indicates the first-order autoregression process. In the later section, we approximate the experimental time series as the first-order autoregression model (AR(1) model) in order to understand the behavior of the time series clearly.

Figure 6e–h shows the results from experiments with $\mu_{\text{init}}=0.55$ (Exps. 5–8: red lines) and − 0.55 (Exps. 9–12: blue lines). For this set of experiments, there were 13 trials conducted for each setting. The initial times of the experiments were every 1 h from 3 h following the commencement of Exps. 1–4; the initial time of these experiments was shifted from Exps. 1–4 because the time series of the parameter ensemble mean for each trial of Exps. 1–4 (black lines in Fig. 6a–d) during the early simulation

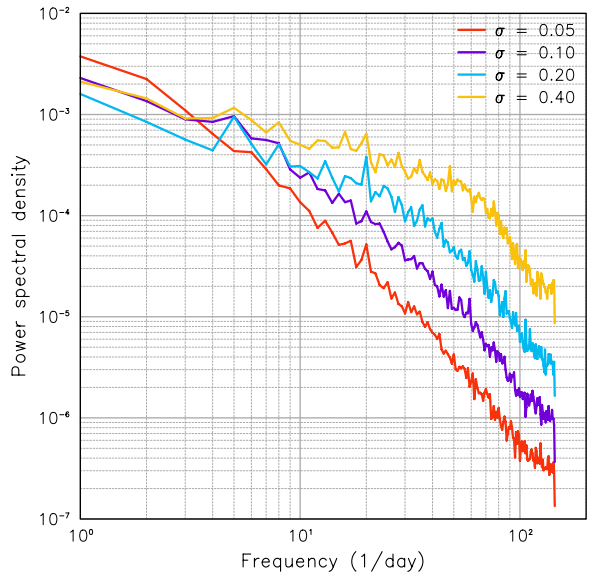

Fig. 7 Power spectral density of the time series of parameter ensemble mean for Exps. 1–4. The red line is for the parameter ensemble spread (σ) of 0.05, the purple line is for $\sigma = 0.1$, the cyan line is for $\sigma = 0.2$, and the yellow line is for $\sigma = 0.4$. Each spectrum was calculated by applying the discrete Fourier transformation to the time series of each trial and averaging these values

time did not sufficiently spread out from the initial value provided. As such, it was not appropriate to carry out a statistical comparison of the time series during this period. We also modified the initial times for each trial of Exps. 5–12, as the convergence speed may vary depending on the simulation time. The estimation duration was 6 h for each experiment. The parameters successfully converged toward the true value in the $\mu_{\text{init}} = 0.55$ and -0.55 series; the convergence speed accelerated as σ became larger, which supports the findings from Kotsuki et al. (2018).

3.2 Accuracy

Although all experiments appeared to attain the true value of the nature run successfully, it was not clear whether the true value was accurately estimated for the entire simulation time. Figure 8 shows the averaged time series for all 200 trials of Exps. 1–4. Red open circles on the black line indicate that the mean of the estimates differs from the true value ($x = 0$) with a significance level of 1% based on the Student's t test. From 14 to 20 h following the commencement of parameter estimation, the estimated parameters tended to be higher than $x = 0$. During this period, time steps where the mean estimates significantly differed from the true value continued intermittently. This indicates that the estimation accuracy varies depending on the observations for each simulation time, despite carrying out the estimation under the twin experiment framework. In this study, we considered the 3 h moving average of the mean time series (blue line in Fig. 8) as the bias component common to all experiments. Sections 3.3 and 3.4 discuss the precision and convergence speed of the estimation, accounting for this estimation bias.

3.3 Precision

The estimation precision is defined as how close the estimated values for different trials are to one another. In this study, the metric of estimation precision was the sample standard deviation of estimation time series with the bias component removed. The sample standard deviation of the estimated parameter was calculated as:

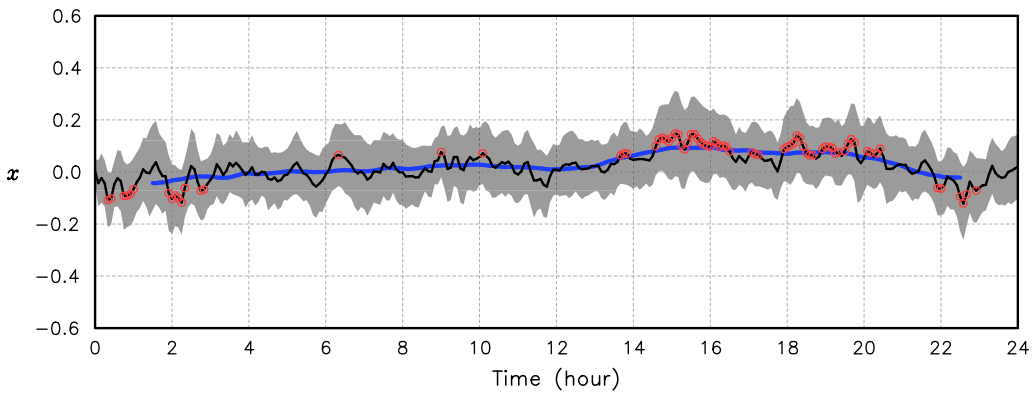

Fig. 8 Averaged time series for all the trials of Exps. 1–4. The black line is the averaged time series. Color shade represents the standard deviation of all trials. Red open circles indicate that the mean of the estimates differs from the true value ($x = 0$) with a significance level of 1% based on the Student's t test. The blue line shows the 3 h moving average of the black line

$$\sigma_\mu = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\{\mu(t) - x_{\mathrm{bias}}(t)\}^2}, \quad (3)$$

where $\mu(t)$ is the time series of the parameter ensemble mean for each time step, t; $x_{\mathrm{bias}}(t)$ is the bias component (blue line in Fig. 8); and T is the number of time steps of the estimation. Here, a smaller σ_μ indicates higher precision; Fig. 9 presents the mean σ_μ, averaged for all 50 trials for each parameter ES (σ). The σ_μ became smallest when σ was 0.1 or less; this implied that $\sigma = 0.1$ was the threshold for the parameter ES. There were no further improvements to the estimation precision even if σ was further reduced from a value of 0.1. As the convergence speed accelerates as σ decreases (see Sect. 3.4), a σ of 0.1 was optimal to provide the best precision and better convergence speed in this experimental framework. Section 3.5 presents the quantification of the dependency between precision and the parameter ES using the AR(1) model.

3.4 Convergence speed

Figure 10 illustrates the composite time series of the parameter ensemble mean for Exps.5–12. Note that the bias component shown in Fig. 8 was removed from each time series prior to compositing. The convergence speed accelerated as σ became larger. The open circles in Fig. 10 indicate that the set of time series did not differ from that for Exps. 1–4, with a significance level of 1% based on the Student's t test. With the exception of Exp. 5 (red line in Fig. 10a), all experiments converged to statistically stationary time series within 6 h.

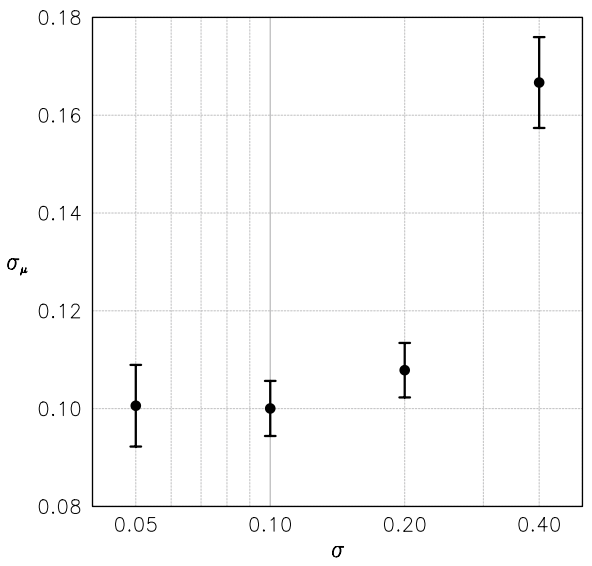

Fig. 9 Sample standard deviation of estimated parameters for Exps. 1–4. The standard deviation for each trial was calculated after removing the bias component (blue line in Fig. 8). Vertical bars represent a confidence interval of 99%

Figure 11 shows the convergence times for each experiment; here, the convergence time was defined as the period from the initial time to the first open circle in Fig. 10. For experiments where $\sigma = 0.4$, the parameter converged within only a few DA cycles; this occurred at ~ 10 min. This means that if σ was not too small, successful parameter estimation occurred within a shorter time than the time scale of mesoscale convective systems (a few hours). The convergence speed is related to the persistence of the time series. This was discussed in quantitative terms by considering the AR(1) model in Sect. 3.5. Notably, when $\sigma \leq 0.2$, the convergence speed for $\mu_{\mathrm{init}} = 0.55$ (Exps. 5–7) was slower than for $\mu_{\mathrm{init}} = -0.55$ (Exps. 9–11). This indicates that the convergence speed differs depending on whether parameter estimation begins from a larger or smaller value than the true value. This is likely to imply that model sensitivity to the c_{R} value was not symmetric to $c_{\mathrm{R}} = 130$. In addition, it is possible that the parameter likelihood itself was not symmetric around the true value.

3.5 Mathematical model of parameter estimation time series

Thus far, this study has demonstrated that although the precision of parameter estimation may be maximized for $\sigma \leq 0.1$ (Fig. 9), the convergence speed for estimation accelerates as σ becomes larger, and as such, the convergence time is shortened (Fig. 11). To quantify the precision and convergence speed of parameter estimation, we attempted to model the temporal variation of the estimated parameter. The spectral amplitude that decreases with increasing frequency in Fig. 7 suggests the estimation time series was close to the AR(1) time series. Here, we approximated the time series of the parameter ensemble mean using the AR(1) model:

$$\mu'(t) = \phi\mu'(t-1) + \varepsilon(t), \quad (4)$$

where t is each DA time step; $\mu'(t) \equiv \mu(t) - x_{\mathrm{bias}}(t)$ is the time series of the parameter ensemble mean where the bias component is removed; ϕ is the autoregressive parameter; and $\varepsilon(t)$ is a random perturbation. The expected value of $\varepsilon(t)$ is assumed as zero (i.e., $\mathrm{E}[\varepsilon] = 0$), and its variance is referred to as ${\sigma_\varepsilon}^2 = \mathrm{E}[\varepsilon^2]$, where $\mathrm{E}[a]$ indicates the expected value of a. When the time series by Eq. (4) is in a statistically stationary state, $\mathrm{E}[\mu'] = 0$. Under this condition, the variance of μ' is defined as, ${\sigma_\mu}^2 = \mathrm{E}[\mu'^2]$. As the AR(1) model assumes that μ' and ε have no correlation (i.e., $\mathrm{E}[\mu'\varepsilon] = 0$), the following relationship between ${\sigma_\mu}^2$ and ${\sigma_\varepsilon}^2$ is satisfied:

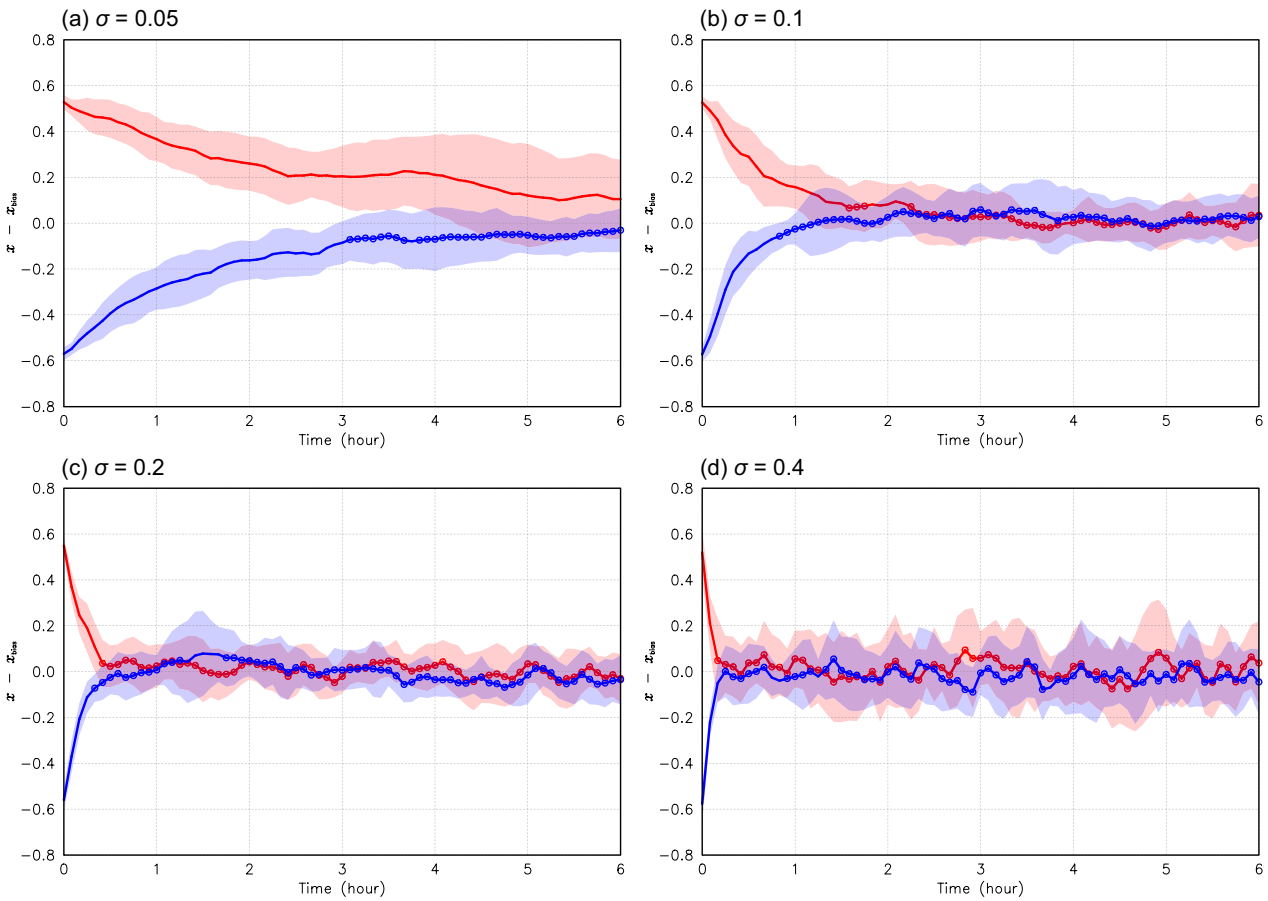

Fig. 10 Composite time series of parameter ensemble mean for Exps. 5–12. Red (blue) lines are for the initial ensemble mean of $\mu_{init} = 0.55$ (− 0.55); **a** experiments where parameter ensemble spread (σ) was 0.05 (Exps. 5 and 9); **b** $\sigma = 0.1$ (Exps. 6 and 10); **c** $\sigma = 0.2$ (Exps. 7 and 11); and **d** $\sigma = 0.4$ (Exps. 8 and 12). Note that the bias component (x_{bias}) shown as the blue line in Fig. 8 was removed from each time series prior to compositing. Color shade represents the standard deviation of all trials. The origin of the time axis corresponds to the initial times of each trial. Open circles on each line indicate that the set of time series does not differ from Exps.1–4, with a significance level of 1% based on the Student's *t* test

$$\sigma_\mu^2 = \phi^2 \sigma_\mu^2 + \sigma_\varepsilon^2. \tag{5}$$

Then, the lag-1 autocorrelation coefficient of μ', $\rho(1)$, satisfies the following:

$$\rho(1) = \frac{E\left[\mu'(t)\mu'(t-1)\right]}{\sigma_\mu^2} = \phi. \tag{6}$$

Note that the range of ϕ is $(-1, 1)$. In the AR(1) model, ϕ is a metric of the time series persistence, where if ϕ is close to 1, $\mu'(t)$ is highly correlated to the previous value $\mu'(t-1)$. In terms of parameter estimation, this means that if a parameter value at one step deviates considerably from the true value, the parameter at the next step also tends to be highly deviating, resulting in a slow convergence speed.

For each trial, ${\sigma_\mu}^2$ was estimated from Eq. (3); then, using Eq. (6), ϕ was estimated as:

$$\phi = \frac{1}{T-1} \sum_{t=2}^{T} \mu'(t)\mu'(t-1)/\sigma_\mu^2. \tag{7}$$

Finally, using Eq. (5), we estimated the standard deviation of the random perturbation, σ_ε, as:

$$\sigma_\varepsilon = \sigma_\mu \sqrt{1-\phi^2}. \tag{8}$$

Figure 12 shows the ϕ and σ_ε values, estimated by calculating Eqs. (7) and (8) for all trials and averaged for each experimental setting; here, ϕ approaches 1 as σ decreases. This trend corresponds to a slow convergence speed of the estimated parameter for the experiment with a smaller σ. At the same time, σ_ε increases with σ; this trend corresponds to a larger amplitude of the time series fluctuation for experiments with a larger σ.

The power spectral density function of AR(1) model is expressed as:

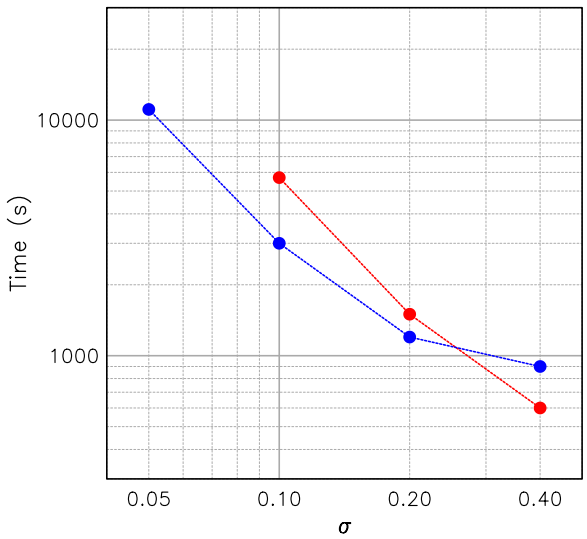

Fig. 11 Convergence time plotted against the parameter ensemble spread (σ). Red line shows Exps. 5–8 where the initial ensemble mean of $\mu_{\text{init}} = 0.55$; blue line shows Exps. 9–12 where $\mu_{\text{init}} = -0.55$. The time series of Exp. 5 does not converge within 6 h, and as such, was not plotted

$$S(f) = \frac{4\sigma_\varepsilon^2 / T}{1 + \phi^2 - 2\phi \cos(2\pi f)}, \tag{9}$$

where f is frequency (Wilks 2011). Here, f was discretized as $f = n/T$ for $n = 0, 1, \ldots, T/2$; $f = 1/2$ was the Nyquist frequency corresponding to a 10 min period. Figure 13 presents the power spectral density function of Eq. (9) using the estimates of ϕ and σ_ε in Fig. 12 (dotted lines), alongside the power spectral density directly derived from the experimental time series (solid lines). Note that the experimental power spectral density slightly differed from Fig. 7 because the bias component was removed from each time series. For $\sigma = 0.05$, the spectrum function was in good agreement with the experimental spectrum, except for the right end of the spectrum (red lines). We inferred that the AR(1) model could suitably approximate the parameter estimation time series for the target parameter c_R. As σ increased, the spectrum function deviated slightly from the experimental spectrum, particularly for low-frequency components. This disagreement may be related to the relatively large uncertainty of estimated ϕ, when σ is large (Fig. 12a). As for the disagreement in the low-frequency range, it is also possible that the frequency components having a period of several hours and more were not detected accurately since the simulation time was just 24 h. From the perspective of the AR(1) model, the rapidly fluctuating time series for a large σ is explained in terms of the dependence of power spectral density function on ϕ. As the parameter ES (σ) becomes larger, the autoregressive parameter ϕ value declines (Fig. 12a), resulting in the increase of power spectral density for the high-frequency (larger f) range as per Eq. (9).

3.6 Quantitative understanding of precision and convergence speed

This section quantifies the precision and convergence speed of parameter estimation based on the AR(1) model. Figures 9 and 11 present the experimental results, illustrating the dependence of precision and convergence

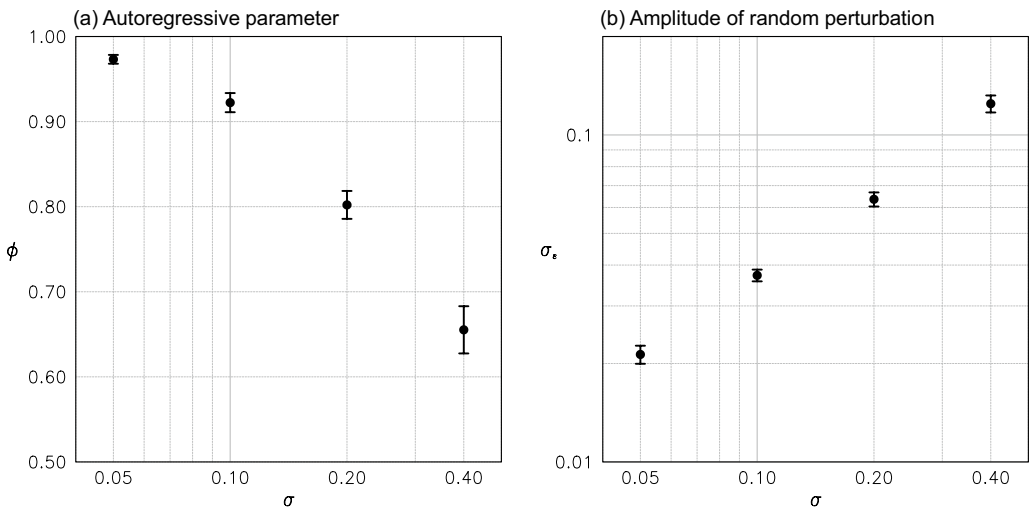

Fig. 12 Two AR(1) model parameters estimated from the time series of Exps. 1–4: **a** estimated autoregressive parameter φ of Eq. (7); and **b** estimated standard deviation of the random perturbation σ_ε of Eq. (8). Each plot represents the mean value for all the trials of Exp. 1 ($\sigma = 0.05$), Exp. 2 ($\sigma = 0.1$), Exp. 3 ($\sigma = 0.2$), and Exp. 4 ($\sigma = 0.4$). The vertical bars represent a confidence interval of 99% for each estimated value

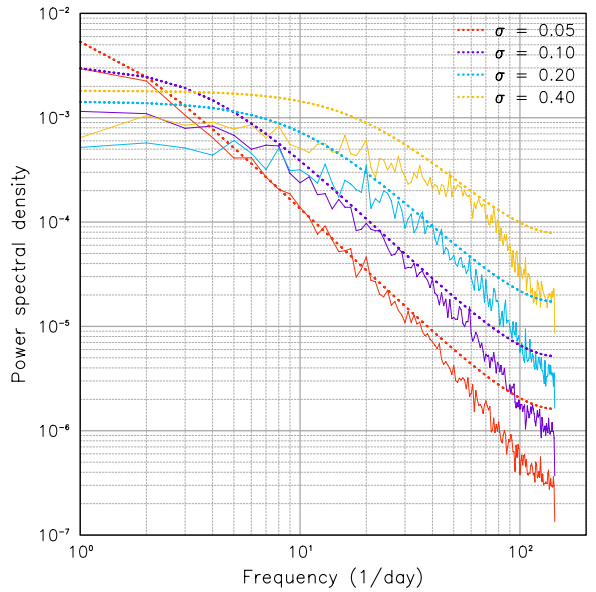

Fig. 13 Power spectral density function assuming the AR(1) model. The lines of spectral functions were based on Eq. (9) using estimated values of autoregressive parameter φ and standard deviation of the random perturbation σ_ε for experiments where $\sigma = 0.05$ (red), 0.1 (purple), 0.2 (cyan), and 0.4 (yellow) (dotted lines). Solid lines show power spectral density directly derived from the experimental time series. The experimental power spectral density differed slightly from Fig. 7 because the bias component (blue line in Fig. 8) was removed from each time series prior to calculating power spectral density

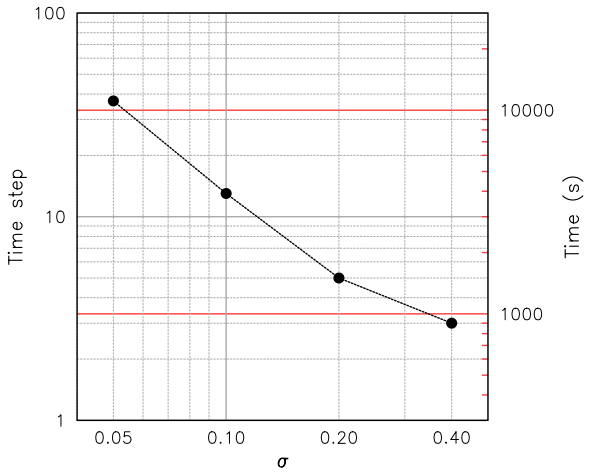

Fig. 14 Time step until the expected deviation from the true value becomes $< 1/e$ assuming the AR(1) model. Each value was calculated based on Eq. (11) using the estimated autoregressive parameter, φ, for experiments where $\sigma = 0.05$, 0.1, 0.2, and 0.4 (Fig. 12a). The right vertical axis shows the actual time in correspondence where one time step is 5 min

time on the parameter ES (σ). This section reconsiders this in terms of the dependence of ϕ and σ_ε in Eq. (4) for σ. Based on Eq. (4), the convergence from the initial wrong value to the true value is expressed as:

$$\left[\mu'(t)\right] = \phi^t \mu'(0), \tag{10}$$

where $\mu'(0)$ is the initial parameter ensemble mean; we used $\mathrm{E}[\varepsilon] = 0$. The e-folding time was estimated by substituting $\mu'(0)/e$ into the left-hand side of Eq. (10) and solving it for t:

$$t = -\frac{1}{\ln(\phi)}.$$

Here, we assumed $\phi > 0$, which is reasonable for this experiment (Fig. 12a). As the DA time step, t, should be an integer, the actual time step where the expected deviation from the true value is $< 1/e$ of its initial deviation assuming the AR(1) model, is:

$$t_{\mathrm{AR}(1)} = \left\lceil -\frac{1}{\ln(\phi)} \right\rceil. \tag{11}$$

Here, $\lceil a \rceil$ indicates the least integer greater than or equal to a. Figure 14 shows the $t_{\mathrm{AR}(1)}$ calculated using Eq. (11), where the estimated ϕ was adopted from Fig. 12a. The right vertical axis of the figure presents the actual time in the correspondence, where one time step is 5 min. The AR(1)-based convergence time in Fig. 14 was in good agreement with the convergence time shown in Fig. 11, indicating that the AR(1) model approximation is reasonable. The convergence speed accelerates as σ increases. In terms of convergence time, a larger σ was more advantageous for parameter estimation; however, the acceptable convergence time was dependent on the time scale of the phenomena (e.g., the time scale of deep convective clouds). If convergence time is within an acceptable range, the optimal parameter ES should be determined to maximize precision.

The standard deviation of the estimated σ_μ is a metric of estimation precision, which is expressed by rewriting Eq. (5) for σ_μ:

$$\sigma_{\mu\mathrm{AR}(1)} = \frac{\sigma_\varepsilon}{\sqrt{1-\phi^2}}. \tag{12}$$

Figure 15 shows the $\sigma_{\mu\mathrm{AR}(1)}$ obtained using Eq. (12), where the estimated ϕ and σ_ε values are shown in Fig. 12; here, $\sigma_{\mu\mathrm{AR}(1)}$ was minimized at $\sigma = 0.05$. However, the $\sigma_{\mu\mathrm{AR}(1)}$ values for $\sigma = 0.05$ and 0.1 did not differ greatly; $\sigma_{\mu\mathrm{AR}(1)} = 9.63 \times 10^{-2}$ for $\sigma = 0.1$, was only 4% larger than $\sigma_{\mu\mathrm{AR}(1)} = 9.27 \times 10^{-2}$ for $\sigma = 0.05$. By contrast, $\sigma_{\mu\mathrm{AR}(1)} = 1.06 \times 10^{-1}$ for $\sigma = 0.2$, was 11% larger than for $\sigma = 0.1$. Thus, we suppressed the estimation uncertainty by setting σ to ≤ 0.1. With a decrease in σ, ϕ and σ_ε were asymptotic to 1 and 0, respectively (Fig. 12). This

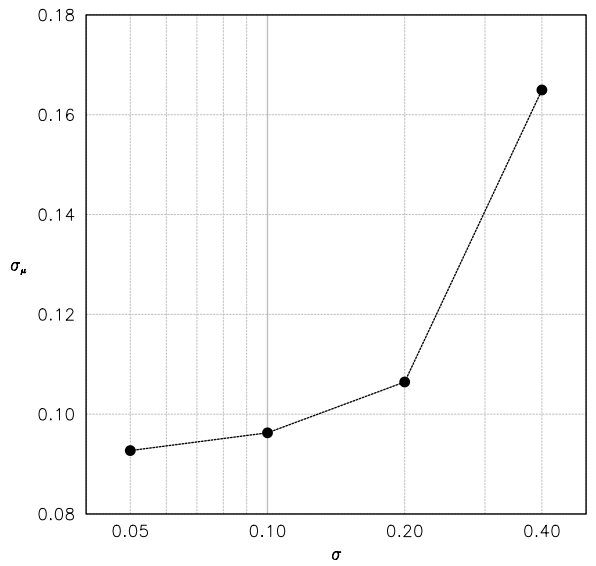

Fig. 15 Standard deviation of estimated values assuming the AR(1) model. This standard deviation was based on Eq. (12) using the estimated φ and σ_ε for experiments where $\sigma = 0.05$, 0.1, 0.2, and 0.4 (Fig. 12)

means that the numerator and denominator on the right-hand side of Eq. (12) becomes zero in the $\sigma = 0$ limit. By decreasing σ, the behavior of $\sigma_{\mu\mathrm{AR}(1)}$ is determined by how the numerator and denominator of Eq. (12) approach zero. For the current target parameter, the ratio of the speed at which the denominator and numerator were asymptotic to zero was almost constant for $\sigma \leq 0.1$; $\sigma_{\mu\mathrm{AR}(1)}$ was minimized for this condition. This demonstrates that the parameter ES of $\sigma \leq 0.1$ achieves the best estimation precision.

Following this, we determined the minimum ES for an acceptable convergence time. The convergence time for $\sigma = 0.1$ was approximately 1 h, as shown in Fig. 14. This time scale was comparable or shorter than long-lasting convective systems. On the other hand, the convergence time for $\sigma = 0.05$ was approximately 3 h. This is likely to be excessive duration for practical parameter estimation to target deep convection. In terms of convergence speed, we accepted a value of $\sigma \geq 0.1$; and therefore, concluded that $\sigma = 0.1$ was the optimal parameter ES to estimate true c_R in this twin experiment. Table 2 summarizes the estimated ϕ and σ_ε values for different σ, and the convergence time step and the standard deviation of the estimated σ_μ, on the basis of the AR(1) fitting.

The first and second rows show the estimated autoregressive parameter, ϕ, and standard deviation of the random perturbation, σ_ε, respectively. The third row presents the $t_{\mathrm{AR}(1)}$ calculated using Eq. (11); the fourth row presents $\sigma_{\mu\mathrm{AR}(1)}$ obtained using Eq. (12); these values were calculated using the estimated ϕ and σ_ε.

Table 2 List of parameters assuming the AR(1) model for different parameter ensemble spread

	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.4$
Estimated φ	0.973	0.922	0.802	0.655
Estimated σ_ε	2.13×10^{-2}	3.72×10^{-2}	6.36×10^{-2}	1.25×10^{-1}
$t_{\mathrm{AR}(1)}$	37	13	5	3
$\sigma_{\mu\mathrm{AR}(1)}$	9.27×10^{-2}	9.63×10^{-2}	1.06×10^{-1}	1.65×10^{-1}

4 Discussion

As the previous section illustrated, the AR(1) model was able to sufficiently approximate the time series for c_R value estimation. As such, the optimal ES may be discussed by considering the autoregressive parameter and the amplitude of random perturbation in the AR(1) model. However, there were also discrepancies with the simple AR(1) time series; for example, for the target parameter, c_R, Fig. 11 shows that the convergence speed varies slightly, depending on whether the estimation starts from a value larger than or smaller than the true value. This may be because the autoregressive parameter ϕ in Eq. (4) is asymmetrically distributed around the true value instead of being a constant. If the autoregressive parameter under constant ES varies depending on model sensitivity to the parameter, it is expected that the estimation time series will follow the AR(1) model more closely by using the method in Ruiz et al. (2013b). Their method modified the parameter ES depending on model sensitivity to the parameter. It also raises an interesting question as to whether the optimal parameter ES estimated by their method approaches the ES determined based on our AR(1) model approach. This issue should be addressed in the future by applying the method in Ruiz et al. (2013b) to this twin experiment and comparing the results.

In the present study, we chose the AR(1) model because of its simplicity in interpretation. However, in general, we can approximate the time series as an autoregression model with an arbitrary order of p (AR(p) model) by considering the autocorrelation from lag-1 to lag-p. In this case, we can construct a model fitting better with the experimental time series by using larger p, but we should choose reasonable p value based on some information criteria (e.g., Akaike's information criterion). It is possible that the relative magnitudes for each autoregressive parameter of the AR(p) model are related to the time scales of various physical processes involved in a target phenomenon, in this case a squall-line convection system. In the present study, both state variables and

the parameter were updated at each DA step, and thus, the 2nd to *p*th autoregressive parameters might become smaller than the 1st one. However, if we conduct the state estimation with a longer time interval than that for the parameter estimation, the magnitudes of second and subsequent autoregressive parameters may become larger depending on the time scale of physical processes.

Note that the method of parameter transformation affects the estimation results. In the case of the transformation of Eq. (2), when the c_R value approaches its upper or lower limit, the spread in the c_R space becomes smaller even though σ in the x space does not change. In this case, the persistence of the estimate could become larger due to lower parameter sensitivity, leading to less convergence to the true value. If we apply more sophisticated method, such that σ in the x space becomes larger when the c_R approaches its limit values to keep the c_R spread approximately constant, the estimation time series might better fit the AR(1) model. Furthermore, we need to keep in mind the possible effects of constraining the lower limit of the parameter to zero. This could skew the parameter likelihood distribution, which might result in the estimation bias and/or the different convergence speed between larger and smaller initial parameter values. While this issue did not appear to have a significant impact on our results, it can cause some problems in the case of more complex experimental settings. Effects of parameter transformation settings on the precision and convergence speed need to be investigated in future research.

The results for c_R estimation do not guarantee that any other parameter estimation follows the AR(1) time series. Parameters that have little or an indirect effect on observed values might show almost random-walk time series that is also the AR(1) process with $\phi = 1$. Conversely, it may be suggested that parameters with a direct impact on observed values follow the autoregressive process and may easily be estimated. For parameter optimization using real observations (e.g., operational radar observations), the "true value" cannot be known a priori, unlike the twin experiment in this study. Furthermore, it is possible that a parameter value does not converge to one specific value in a real situation; it is unclear whether the time series of parameter optimization follows an autoregressive process. Thus, we suggest the following steps. First, we attempt some short-period optimization experiments beginning from different initial parameter ensemble means. Here, a relatively large parameter ES should be utilized, because, as shown in Fig. 12a, it is expected that the larger the ES is, stronger the autoregression (smaller the ϕ in Eq. (4)) becomes. Suppose there are some signs of autoregression in a time series during the experiment, such as a tendency for the parameter to begin at different initial values to approach the same value. In this case, we can apply the idea of optimal ES discussed in this study to the parameter to be optimized. Then, by estimating the dependence of the autoregressive parameter and the amplitude of the random perturbation on the parameter ES based on some experiments, we can identify its threshold in terms of precision. Finally, we determine the optimal ES by comparing the expected convergence time and the time scale of the target phenomena. If we determine the optimal ES, we can conduct a longer period estimation with sufficient reliability and efficiency. If the estimated parameter value is trapped near prescribed lower or upper limit of parameter transformation, the true value might be located outside the limit. In this case, we should extend the limit value.

This study also highlighted an interesting albeit problematic issue that the estimated value contains some bias; this was also true for the twin experiment framework using the perfect model (Fig. 8). The overestimation of the c_R for $T = 14$–20 h in the experiment implies that a slightly larger value than the true $c_R = 130$ would reduce prediction error during this period. This is likely a result from the observations during this period. While the nature run of the experiment provided one realization under $c_R = 130$, it is not necessarily the most likely realization in the state space under this condition. The time evolution of the nature run for $T = 14$–20 h probably went through the path which was the most likely realization when c_R value was slightly larger than 130. This estimation bias may be removed if the parameter estimation was carried out over a longer simulation time or the estimates were averaged over many different simulations. For parameter optimization using real observations, the "true value" is likely to vary depending on time and space. The bulk cloud microphysics scheme used was highly simplified compared with real microphysics. Thus, we would be unable to determine the unique optimal parameter values that produce the best prediction accuracy under various atmospheric conditions. The optimized parameter values for the cloud microphysics scheme will vary depending on the type of target convective system, model resolution, and combination with other physical models. In parameter optimization for practical numerical weather prediction, there is a need to clarify the phenomenon to be predicted, the parameter that requires optimization for such a phenomenon, and how the parameter interacts with observed values.

Note that there were also many issues to be considered, even for the simple twin experiment in this study. For example, altering the time interval of the DA may also change the precision and convergence speed of parameter estimation. This study assimilated the radar

reflectivity every 5 min; if the time interval is reduced, the effects of different parameter values on the radar reflectivity distribution may not be apparent during the ensemble forecast, making parameter estimation difficult. Conversely, if the time interval is lengthened, nonlinearities of the microphysical processes collapses a clear correlation between the target parameter and radar reflectivity, also making parameter estimation difficult. In this case, however, the possible detrimental effect of nonlinearities may be mitigated by conducting the state estimation with a time interval shorter than that for the parameter updates. Ensemble size is also an important factor of data assimilation; increasing ensemble size may improve the precision and convergence speed of the estimation without altering other settings. Additionally, the simultaneous estimation of multiple parameters may change some results. In reality, the optimal parameter values estimated concurrently may differ from those estimated separately due to interactions among multiple parameters. The estimation of multiple parameters is an area that should be addressed in future research.

In addition to these issues, how to determine the optimal ES for the estimation of time-dependent parameters needs further discussion. When parameters to be estimated are time-dependent, slow convergence speed could be more fatal than low precision. The parameter ensemble mean might never reach the true value at each time step because the estimate cannot follow the time variation of it due to the slow convergence speed. In this case, it is better to prioritize the convergence speed over the precision in determining the better parameter ES. Other compromise solution is to estimate the parameter for a longer period and average the result to determine one optimal value.

Finally, we note that the dependence of the precision and convergence speed on the parameter ES should be explained from the mathematical view of the data assimilation. Large persistence of the time series for the small parameter ES might reflect low parameter sensitivity due to the small parameter uncertainty. On the other hand, large fluctuation amplitude for the large parameter ES seems to be more difficult to be explained. It could reflect some interaction among the parameter uncertainty, the observational errors, and the state variables' uncertainties. In addition, it is important to investigate how the settings for the parameter uncertainty affects the accuracy of state estimation to understand the relation between parameter and state estimations. Theoretical studies on these issues should be addressed in future research.

5 Conclusions

This study carried out an OSSE, focusing on model parameter estimation; the dependence of the precision and convergence speed of estimation on the parameter ES was investigated. We experimentally demonstrated that the precision deteriorates with increasing parameter ES, while the convergence speed accelerates. However, there is a threshold for ES, where any smaller values do not result in any further improvements to the estimation precision. By using this approach, we determined the optimal parameter ES that provides the best precision and reasonable convergence time shorter than the time scale of target phenomena. At the same time, we checked the accuracy of estimation. We found there exists some systematic errors of estimated values which depend on the observation at each time step. In contrast to the precision and convergence speed, the accuracy cannot be controlled by setting the parameter ES.

In addition, we clarified that the time series of parameter estimation follows a simple mathematical model. We approximated the estimation time series as the AR(1) model, demonstrating that the precision and convergence speed of the estimation may be quantified by two AR(1) model parameters: the autoregressive parameter and the amplitude of random perturbation. As the ES of the parameter increases, the autoregressive parameter decreases, contributing to improving the precision and accelerating convergence speed. By contrast, as the amplitude of random perturbation increases; this reduces precision. From this perspective, the estimation precision was determined based on the balance between autoregression and random perturbation. This viewpoint is also useful in determining the optimal ES for other parameters that were not the focus of this study.

The final objective of this study was to establish a practical procedure to optimize the parameters of atmospheric models by assimilating real observations and improving the accuracy of numerical weather and climate prediction. For example, parameter optimization for cloud-resolving models using state-of-the-art observation equipment, such as phased array weather radars, may improve the accuracy of severe weather predictions. However, there is still a large gap between the findings of this study and achieving this final objective. From this study, we will gradually complicate the problem settings. For example, we will attempt parameter optimization of a low-resolution model by assimilating high-resolution simulation. In addition, optimization for a single-moment cloud microphysics scheme will be carried out by assimilating simulations with a more precise double-moment scheme. The simultaneous estimation of

multiple parameters and the estimation of time-dependent parameters are also important issues; by successively building knowledge, practical applications of parameter optimization for actual numerical weather predictions may be achieved.

Abbreviations

4DEnVar: Four-dimensional ensemble-variational method; 4DVar: Four-dimensional variational method; AR(1) model: First-order autoregression model; DA: Data assimilation; EnKF: Ensemble Kalman filter; EnSRF: Ensemble square root filter; ES: Ensemble spread; ETKF: Ensemble transform Kalman filter; GCM: General circulation model; LETKF: Local ensemble transform Kalman filter; RTPS: Relaxation to prior spread; SCALE-RM: Scalable computing for advanced library and environment-regional model.

Acknowledgements

We would like to thank the two reviewers of our manuscript for their helpful suggestions. We would like to thank the Data Assimilation Research Team of RIKEN Center for Computational Science (R-CCS) for providing us with the SCALE-LETKF system and their constructive comments and suggestions. We would also like to thank Editage (www.editage.com) for English language editing.

Author contributions

KS and HT proposed the topic, conceived, and designed the study. SN proposed a basic idea for the analysis. TY prepared and improved DA system. All authors read and approved the final manuscript.

Funding

This study was supported by JST CREST Grant Number JPMJCR1312, JST AIP Grant Number JPMJCR19U2, and the Foundation for Computational Science (FOCUS) Establishing Supercomputing Center of Excellence, Japan.

Availability of data and material

The code for SCALE-RM version 5.3.x is freely available at https://scale.riken.jp under the 2-Clause BSD license. The code for the SCALE-LETKF system is available upon reasonable request from the Data Assimilation Research Team of RIKEN Center for Computational Science (R-CCS) (http://www.data-assimilation.riken.jp/index_e.html). The model configuration files and experimental results were deposited in local storage at the R-CCS, Kobe, Japan. They are available upon request from the corresponding author.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. [2]RIKEN Cluster for Pioneering Research, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. [3]Research Center for Urban Safety and Security, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan.

Received: 1 December 2021 Accepted: 13 August 2022
Published online: 06 September 2022

References

Aksoy A, Zhang F, Nielsen-Gammon JW (2006a) Ensemble-based simultaneous state and parameter estimation with MM5. Geophys Res Lett 33(12):L12801. https://doi.org/10.1029/2006GL026186

Aksoy A, Zhang F, Nielsen-Gammon JW (2006b) Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Mon Wea Rev 134(10):2951–2970. https://doi.org/10.1175/MWR3224.1

Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. Mon Wea Rev 129(3):420–436. https://doi.org/10.1175/1520-0493(2001)129%3c0420:ASWTET%3e2.0.CO;2

Buehner M, Morneau J, Charette C (2013) Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. Nonlin Process Geophys 20(5):669–682. https://doi.org/10.5194/npg-20-669-2013

ECMWF (2020)IFS Documentation CY47R1. Part II: data assimilation. https://www.ecmwf.int/node/19746. Accessed 30 Nov 2021

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res Oceans 99(C5):10143–10162. https://doi.org/10.1029/94JC00572

Gilmore MS, Straka JM, Rasmussen EN (2004) Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. Mon Wea Rev 132(11):2610–2627. https://doi.org/10.1175/MWR2810.1

Hourdin F, Mauritsen T, Gettelman A, Golaz JC, Balaji V, Duan Q, Folini D, Ji D, Klocke D, Qian Y, Rauser F, Rio C, Tomassini L, Watanabe M, Williamson D (2017) The art and science of climate model tuning. Bull Amer Meteor Soc 98(3):589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Houtekamer PL, Mitchell HL, Pellerin G, Buehner M, Charron M, Spacek L, Hansen B (2005) Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. Mon Wea Rev 133(3):604–620. https://doi.org/10.1175/MWR-2864.1

Hunt BR, Kostelich EJ, Szunyogh I (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D 230(1–2):112–126. https://doi.org/10.1016/j.physd.2006.11.008

JMA (2019) Outline of the operational numerical weather prediction at the Japan Meteorological Agency. http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2019-nwp/index.htm. Accessed 30 Nov 2021

Kotsuki S, Terasaki K, Yashiro H, Tomita H, Satoh M, Miyoshi T (2018) Online model parameter estimation with ensemble data assimilation in the real global atmosphere: a case with the nonhydrostatic icosahedral atmospheric model (NICAM) and the global satellite mapping of precipitation data. J Geophys Res Atmos 123(14):7375–7392. https://doi.org/10.1029/2017JD028092

Kotsuki S, Sato Y, Miyoshi T (2020) Data assimilation for climate research: model parameter estimation of large-scale condensation scheme. J Geophys Res Atmos 125(1):e2019JD031304. https://doi.org/10.1029/2019JD031304

Lien GY, Miyoshi T, Nishizawa S, Yoshida R, Yashiro H, Adachi SA, Yamaura T, Tomita H (2017) The near-real-time SCALE-LETKF system: A case of the september 2015 Kanto-Tohoku heavy rainfall. SOLA 13:1–6. https://doi.org/10.2151/sola.2017-001

Marshall JS, Palmer WM (1948) The distribution of raindrops with size. J Meteor 5(4):165–166. https://doi.org/10.1175/1520-0469(1948)005%3c0165:TDORWS%3e2.0.CO;2

Mauritsen T, Stevens B, Roeckner E, Crueger T, Esch M, Giorgetta M, Haak H, Jungclaus J, Klocke D, Matei D, Mikolajewicz U, Notz D, Pincus R, Schmidt H, Tomassini L (2012) Tuning the climate of a global model. J Adv Model Earth Syst 4(3):M00A01. https://doi.org/10.1029/2012MS000154

Nishizawa S, Yashiro H, Sato Y, Miyamoto Y, Tomita H (2015) Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations. Geosci Model Dev 8(10):3393–3419. https://doi.org/10.5194/gmd-8-3393-2015

NOAA (2019) The global forecast system (GFS) - global spectral model (GSM). https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php. Accessed 30 Nov 2021

Ruckstuhl Y, Janjić T (2020) Combined state-parameter estimation with the LETKF for convective-scale weather forecasting. Mon Wea Rev 148(4):1607–1628. https://doi.org/10.1175/MWR-D-19-0233.1

Ruiz JJ, Pulido M, Miyoshi T (2013a) Estimating model parameters with ensemble-based data assimilation: a review. J Meteor Soc Japan 91(2):79–99. https://doi.org/10.2151/jmsj.2013-201

Ruiz JJ, Pulido M, Miyoshi T (2013b) Estimating model parameters with ensemble-based data assimilation: parameter covariance treatment. J Meteor Soc Japan 91(4):453–469. https://doi.org/10.2151/jmsj.2013-403

Sanderson BM, Piani C, Ingram WJ, Stone DA, Allen MR (2008) Towards constraining climate sensitivity by linear analysis of feedback patterns in

thousands of perturbed-physics GCM simulations. Clim Dyn 30:175–190. https://doi.org/10.1007/s00382-007-0280-7
Sasaki Y (1969) Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. J Meteor Soc Jpn 47(2):115–124. https://doi.org/10.2151/jmsj1965.47.2_115
Sato Y, Nishizawa S, Yashiro H, Miyamoto Y, Kajikawa Y, Tomita H (2015) Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation? Prog Earth Planet Sci 2:23. https://doi.org/10.1186/s40645-015-0053-6
Thompson P (1969) Reduction of analysis error through constraints of dynamical consistency. J Appl Meteor 8(5):738–742. https://doi.org/10.1175/1520-0450(1969)008%3c0738:ROAETC%3e2.0.CO;2
Tomita H (2008) New microphysical schemes with five and six categories by diagnostic generation of cloud ice. J Meteor Soc Jpn 86A:121–142. https://doi.org/10.2151/jmsj.86A.121
Tong M, Xue M (2005) Simultaneous retrieval of microphysical parameters and atmospheric state variables with radar data and ensemble Kalman filter method. In: Extended abstract of 17th Conference on Numerical Weather Prediction, Washington, DC, pp 1–5 Aug 2005. https://ams.confex.com/ams/WAFNWP34BC/techprogram/paper_95042.htm
Tong M, Xue M (2008) Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: parameter estimation experiments. Mon Wea Rev 136(5):1649–1668. https://doi.org/10.1175/2007MWR2071.1
Weisman ML, Klemp JB (1982) The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. Mon Wea Rev 110(6):504–520. https://doi.org/10.1175/1520-0493(1982)110%3c0504:TDONSC%3e2.0.CO;2
Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Wea Rev 130(7):1913–1924. https://doi.org/10.1175/1520-0493(2002)130%3c1913:EDAWPO%3e2.0.CO;2
Whitaker JS, Hamill TM (2012) Evaluating methods to account for system errors in ensemble data assimilation. Mon Wea Rev 140(9):3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1
Wilks DS (2011) Statistical methods in the atmospheric sciences, 3rd edn. Academic Press
Zhu Y, Navon IM (1999) Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. Mon Wea Rev 127(7):1497–1517. https://doi.org/10.1175/1520-0493(1999)127%3c1497:IOPEOT%3e2.0.CO;2

Publisher's Note