

Chapter 12

Model Error Representations Using the Covariance Inflation Methods in Ensemble Data Assimilation System



Sujeong Lim and Seon Ki Park

Abstract Ensemble data assimilation estimates the initial conditions and the flow-dependent background error covariance using observations and ensemble forecasts. The ensemble background error covariance represents the model uncertainty, but it is usually underestimated due to insufficient ensemble size and model errors. Consequently, analysis overtrusts the model forecasts and ignores observations. To solve this problem, we implemented the stochastically perturbed hybrid tendencies scheme to the local ensemble transform Kalman filter in a global numerical weather prediction model—the Korean Integrated Model. It describes the model uncertainties from the computational representations of underlying partial differential equations and the imperfect physical parameterizations, simultaneously. As a result, the new stochastic perturbation scheme leads to an increase in ensemble spread and a decrease in the ensemble mean error, especially in the troposphere.

Keywords Ensemble data assimilation · Background error covariance · Model uncertainty · Inflation methods · Stochastic perturbations

12.1 Introduction

The ensemble data assimilation (EDA) system produces the uncertainties in initial conditions (ICs) and model forecasts. An ensemble of analyses, computing the flow-dependent background error covariance (BEC), improves the assimilation of observations (Palmer et al. 2009). The EDA system often generates underestimated BEC due to the limited ensemble size and model errors (Miyoshi 2011; Kotsuki et al.

S. Lim

Center for Climate/Environment Change Prediction Research, Ewha Womans University, Seoul 03760, Republic of Korea
e-mail: sujeong.lim@ewha.ac.kr

S. K. Park (✉)

Department of Climate and Energy System Engineering, Ewha Womans University, Seoul 03760, Republic of Korea
e-mail: spark@ewha.ac.kr

2017). Therefore, covariance inflation (CI) methods that artificially increase the error covariance (Luo and Hoteit 2013) are required to prevent the filter divergence whose analyses diverge from the real state (Houtekamer and Mitchell 1998; Lim et al. 2020). The CI methods in the EDA system have two approaches with respect to the prior (i.e., background) and posterior (i.e., analysis) (Duc et al. 2020): the prior CI methods inflate BEC, and then it is digested into the EDA system to determine the analysis mean and analysis error covariance while the posterior CI methods use BEC in the DA process, and the resulting analysis error covariance is inflated.

In detail, the prior CI methods indirectly inflate the analysis error covariance by inflating BEC. For example, *multiplicative inflation* (e.g., Anderson and Anderson 1999) multiplies BEC (\mathbf{P}^b) with an inflation factor (γ). It increases the amplitude of the error covariance without modifying the structure as follows:

$$\mathbf{P}_{\text{inf}}^b = \gamma \mathbf{P}^b, \quad (12.1)$$

where $\mathbf{P}_{\text{inf}}^b$ is the inflated BEC and γ is larger than 1. Next, *additive inflation* adds the perturbation samples with zero mean from a given model error distribution to analysis ensemble members (e.g., Mitchell and Houtekamer 2000; Whitaker and Hamill 2012). It helps to depict the heterogeneous model uncertainties to analysis perturbations. Last, *stochastic perturbation methods* estimate the model uncertainties due to the imperfect processes (e.g., Buizza et al. 1999; Shutts 2005; Berner et al. 2009) such as the dynamical and physical approximations: the stochastically perturbed parameterization tendency (SPPT) scheme assumes uncertainty in the parameterized physical tendency (Palmer et al. 2009); the stochastically perturbed dynamical tendency (SPDT) scheme assumes uncertainty in the dynamical tendency (Koo and Hong 2014); and the stochastic kinetic energy backscatter (SKEB) scheme assumes uncertainty from the unresolved scale interactions of NWP model (Shutts 2005). It is known that they are effective in increasing ensemble spread and improving probability skills (Leutbecher et al. 2017). Previous studies simulated the SPPT and stochastic backscatter (SPBS) schemes to investigate the impacts of model uncertainty on the EDA system (Isaksen et al. 2010). Results showed that the SPPT scheme increased the ensemble spread of temperature at the top of the planetary boundary layer (PBL) and wind in the tropics near 700 hPa, whereas SPBS increased the ensemble spread of wind in the PBL. Therefore, it is recommended to use a combination of stochastic perturbation schemes in the EDA system because they complement each other by expressing different model uncertainties.

As for the posterior CI methods, it directly inflates the analysis error covariance including the following methods. First, *relaxation-to-prior perturbation* (RTPP; Zhang et al. 2004) inflates the analysis perturbations (\mathbf{X}^a) with relaxation factor (α) as follows:

$$\mathbf{X}_{\text{inf}}^a = (1 - \alpha)\mathbf{X}^a + \alpha\mathbf{X}^b, \quad (12.2)$$

where $\mathbf{X}_{\text{inf}}^a$ is the inflated analysis perturbations, \mathbf{X}^b is the forecast perturbations, and α approaches 1.0. It relaxes the posterior perturbations back toward the prior per-

turbations. Second, *relaxation-to-prior spread* (RTPS; Whitaker and Hamill 2012) inflates the analysis ensemble standard deviation (spread), and thus, it relaxes the ensemble standard deviation back to the prior by multiplication as below:

$$\mathbf{X}_{\text{inf}}^a = \left(1 + \alpha \left(\frac{\sigma^b}{\sigma^a} - 1 \right) \right) \mathbf{X}^a, \quad (12.3)$$

where σ^b and σ^a are the prior and posterior ensemble standard deviation at each analysis grid point. Note that RTPP works on large-scale processes and RTPS works on small-scale processes (Bowler et al. 2017; Duc et al. 2020). Last, *adaptive inflation* adaptively estimates the multiplicative inflation parameters (α) using the posterior innovation statistics (Ying and Zhang 2015).

In this study, we focused on the stochastic perturbation method, which is one of the prior CI methods, to account for the model uncertainties in the EDA system. We assume that the model uncertainties come from dynamical and physical tendencies and have implemented the stochastic perturbation hybrid tendencies scheme to a global numerical weather prediction model (Lim et al. 2020). We anticipate that this new stochastic perturbation scheme increases the ensemble spread and reduces the ensemble mean error by solving the underestimated BEC problems. Section 12.2 presents the methodology, and the experimental design is described in Sect. 12.3. Sections 12.4 and 12.5 provide the results and discussion and conclusion, respectively.

12.2 Methodology

12.2.1 Forecast Model and Ensemble Data Assimilation System

We used the Korean Integrated Model (KIM) (Hong et al. 2018; Kim et al. 2021), which is an operational global atmospheric model at the Korea Meteorological Administration (KMA) since April 2020. It consists of a spectral-element non-hydrostatic dynamical core on a cubed sphere and advanced physics parameterization packages (see Hong et al. 2018). The ensemble forecast implemented to the EDA system has a 50 km horizontal resolution and 91 vertical levels up to 0.01 hPa in the hybrid sigma-pressure vertical coordinate.

For the EDA system, we used the four-dimensional local ensemble transform Kalman filter (4D-LETKF) system (Hunter et al. 2007; Liu et al. 2008; Shin et al. 2016). The 4D-LETKF system finds the ensembles of analysis obtained by assimilating the observations within a local region (Hunter et al. 2007; Shin et al. 2016, 2018). Control variables are the zonal wind, meridional wind, potential temperature, humidity mixing ratio, and surface pressure. The initial 50 ensemble members are produced by modifying the analysis with the lagged forecast differences. Assimilated observations

are sonde, surface, aircraft, Global Positioning System-Radio Occultation (GPS-RO), Infrared Atmospheric Sounding Interferometer (IASI), Advanced Microwave Sounding Unit-A (AMSU-A), Cross-track Infrared Sounder (CrIS), Microwave Humidity Sounder (MHS), Advanced Technology Microwave Sounder (ATMS), and Atmospheric Motion Vector (AMV) (Kang et al. 2018).

12.2.2 Stochastic Perturbation Hybrid Tendencies Scheme

We simultaneously perturb dynamical and physical tendencies and define it as the stochastic perturbation hybrid tendencies (SPHT) scheme (i.e., SPDT + SPPT) (Lim et al. 2020). The dynamical tendency is related to the explicitly resolved dynamics and horizontal diffusion, and the physical tendency is related to the physical parameterization schemes. As a result, the SPHT scheme accounts for model uncertainties associated with computational representations of the underlying partial differential equations and imperfect physical parameterizations. In the SPHT scheme, the dynamical and physical tendencies are perturbed using the randomly generated multiplicative perturbation at each model time step and grid point, i.e.,

$$x^{n*} = x^n + (1 + \mu r) \left(\frac{\partial x^n}{\partial t} \right)_{\text{dyn}} \Delta t, \quad (12.4)$$

and

$$x^{n+1} = x^{n*} + (1 + \mu r) \left(\frac{\partial x^{n*}}{\partial t} \right)_{\text{phy}} \Delta t, \quad (12.5)$$

where $\left(\frac{\partial x}{\partial t} \right)_{\text{dyn}}$ and $\left(\frac{\partial x}{\partial t} \right)_{\text{phy}}$ are the dynamical and physical tendencies, respectively, n is time step, t is time, and $*$ is provisional solution of the dynamical process; $\mu \in \{0, 1\}$ represents the vertical tapering function ($e^{\eta-1}$) in the generalized vertical coordinate (η), and r is the random forcing. Note that the model variable (x) consists of temperature and humidity mixing ratio only. Since the physics and dynamics in KIM are coupled by a time-splitting method, this approach differs from the method of perturbing total model tendency.

The random forcing determining the perturbations depends on the following tuning parameters: (1) the horizontal correlation length scale (L) determines how much perturbed errors propagate in a horizontal direction; (2) the de-correlation time scale (τ) determines how long the perturbed errors will be sustained; (3) the standard deviation of perturbation (σ) controls the amplitudes of random forcing; and (4) the tapering function (μ) determines whether the error would exponentially decrease by tapering to zero in the lowermost and the uppermost vertical layers or remain the same to avoid numerical instability. If the SPPT and SPDT schemes use the same random forcing, the latter produces larger perturbations; thus, we have designed the SPDT scheme to use smaller σ in order to suppress spurious instability (Koo and Hong 2014).

Table 12.1 List of random forcing tuning parameters used in the SPDT and SPPT schemes, respectively

	Length scale (L ; in km)	Time scale (τ ; in s)	Standard deviation (σ)	Tapering function (μ)
SPDT	500	10,800	0.5	On
SPPT	500	21,600	1.0	Off

12.3 Experimental Design

We performed two sets of EDA runs to assess the SPHT scheme: (1) CTRL is the control EDA runs without any stochastic perturbation schemes, and (2) STOC runs the SPHT scheme in EDA runs. The random forcing tuning parameters used in the SPHT scheme are listed in Table 12.1. The experiments started at 12:00 UTC on 22 June 2018 and ended at 12:00 UTC on 7 July 2018. We specified the first 78 h as a spin-up period, and each cycle produced 6 h forecasts to generate ensemble BEC.

To prevent the filter divergence, the two experiments included the localization and posterior CI methods. Regarding the covariance localization in LETKF, the horizontal localization is given by a Gaussian-like piecewise fifth-order rational function (Gaspari and Cohn 1999; Miyoshi 2011) and varies from 660 to 1800 km in radius of influence depending on vertical levels (Kleist and Ide 2015). The vertical localization differs as to the observation type: the conventional data are defined by a Gaussian-like rational function as $2\sqrt{\frac{10}{3}}\sigma_v$ where σ_v depends on pressure (p) (e.g., σ_v is $0.2 \ln(p)$ for wind and surface pressure and $0.1 \ln(p)$ for mass variables), and the radiance data are defined by the gradient of transmittance of the measured radiance (Thépaut 2003). Regarding the posterior CI methods, both experiments used the additive inflation and the RTPS with the relaxation parameter of 0.95.

12.4 Results

We examined how ensembles in CTRL describe the model uncertainty by comparing the ensemble spread and ensemble mean error. Here, we defined the ensemble mean error as the root-mean-squared error (RMSE) of the ensemble mean against the Integrated Forecast System (IFS) analysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). IFS has a 25 km resolution with 25 pressure levels from 1000 to 1 hPa, and we assumed it as a true state. Figure 12.1 shows the zonally averaged ensemble spread and ensemble mean error to diagnose the current ensemble status. The ensemble spread is expected to be similar to the ensemble mean error in terms of magnitude and patterns. For temperature (Fig. 12.1a and d), the underestimated ensemble spread is found in the lower troposphere (e.g., below 700 hPa) and the stratosphere. For specific humidity (Fig. 12.1b and e), the underestimation is

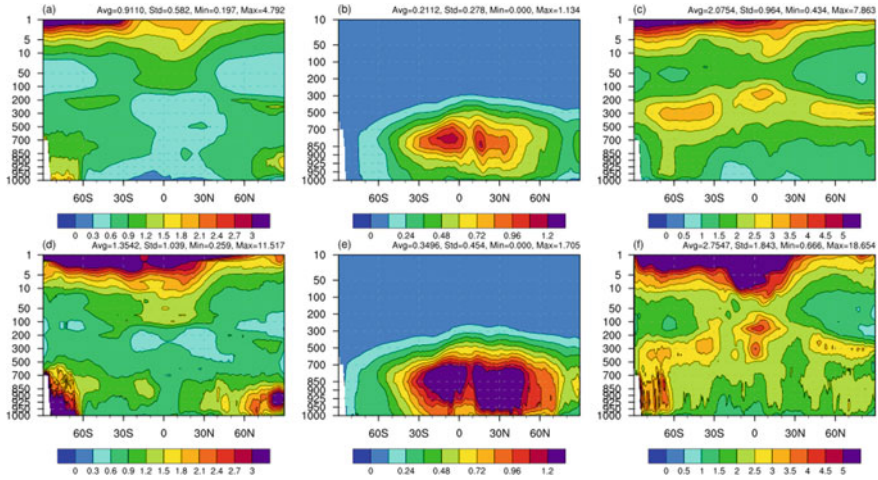


Fig. 12.1 Zonal mean ensemble spread (top panels) and ensemble mean error (bottom panels) for **a, d** temperature (in K), **b, e** specific humidity (in g kg^{-1}), and **c, f** zonal wind (m s^{-1}) in CTRL. Background (i.e., 6 h forecasts) is averaged from 1800 UTC 25 June 2018 to 1800 UTC 7 July 2018, excluding the spin-up period (first 78 h of the experiment time)

founded in the tropics and mid-latitudes below 700 hPa. For zonal wind (Fig. 12.1c and f), the ensemble spread already described the model uncertainty, but the underestimation also remained in Antarctica and most of the stratosphere. Overall, the model uncertainties in the troposphere, especially below 700 hPa, were not depicted in CTRL experiments. Although available large ensemble members (i.e., 50 ensemble members), localization methods, and posterior CI methods (e.g., additive inflation and RTPS) tried to describe the model uncertainty, they were insufficient, especially for temperature and specific humidity. Therefore, an additional CI method is necessary to increase the ensemble spread in the lower troposphere to produce the desirable ensembles in EDA cycles.

Next, we examine if the SPHT scheme as a prior CI method can increase the ensemble spread in the EDA system. Figure 12.2 shows the globally averaged vertical profiles of the difference between STOC and CTRL for temperature, specific humidity, and zonal wind. The ensemble spread of ICs in the initial cycle was identical, but they were increased as the forecast times. The main difference was founded below 700 hPa, and the strongest amplitude is near 950–925 hPa. The tapering function in the SPHT scheme reduces perturbation in the lowermost and uppermost layers to secure stability. We did not perturb zonal wind as did in temperature and specific humidity, but ensemble spread of zonal wind was indirectly increased. Similarly, the stratosphere (e.g., from 10 to 1 hPa) showed an increased ensemble spread due to propagation from the perturbations in other layers.

We investigated the time series of globally averaged ensemble spread and ensemble mean error to assess the SPHT scheme (Fig. 12.3). Compared to CTRL, STOC increases the ensemble spread and decreases ensemble mean error for temperature,

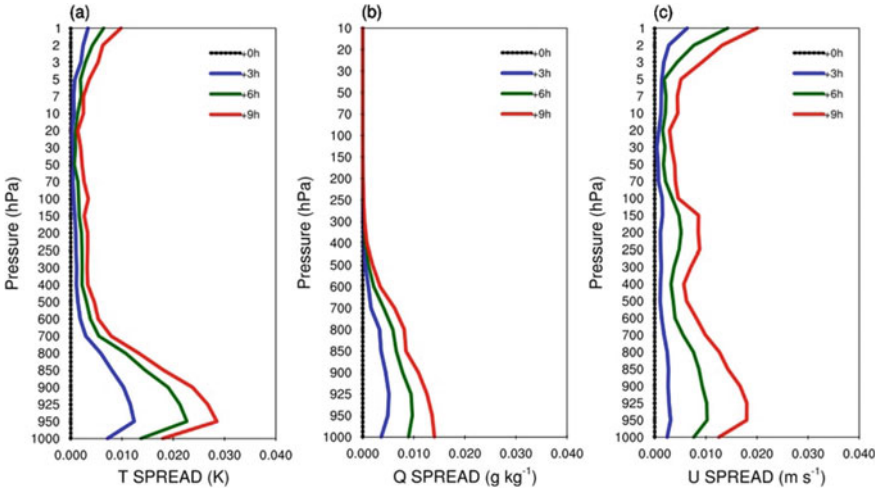


Fig. 12.2 Differences of the ensemble spread between STOC and CTRL (i.e., STOC-CTRL) in the globally averaged vertical profiles of **a** temperature (T ; in K), **b** specific humidity (Q ; in g kg^{-1}), and **c** zonal wind (U ; in m s^{-1}) for forecast times of +3 h (blue), +6 h (green), and +9 h (red) from the initial time (+0h; black dots) at 1200 UTC 22 June 2018. Variations in the ensemble spread reflect the effect of the new SPHT scheme. ©2020 Authors. Distributed under CC BY 4.0 License

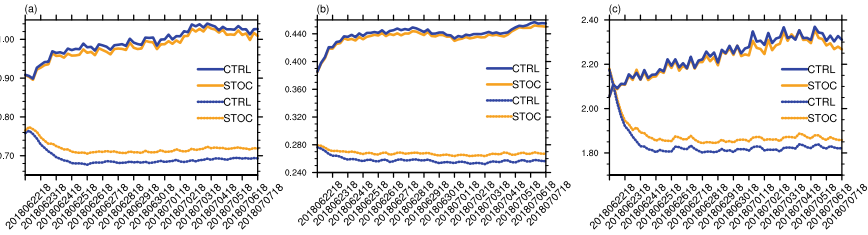


Fig. 12.3 Time series of the globally averaged ensemble spread (dotted line) and the ensemble mean error (solid line) in the prior for CTRL (blue line) and STOC (orange line). **a** is temperature (in K), **b** is specific humidity (in g kg^{-1}), and **c** is zonal wind (in m s^{-1}). The x -axis is the analysis time with 6 h assimilation window

specific humidity, and zonal wind. As a result, the prior CI method using the SPHT scheme successfully increased the ensemble spread by 3.7%, 3.9%, and 2.3% for temperature, specific humidity, and zonal wind, respectively, and decreased the ensemble mean error by 1.1%, 0.9%, and 0.6%, respectively. To summarize, the notable improvements occurred in temperature, and they were effective in the tropics below 700 hPa. In detail, the SPPT scheme mainly works in the overall ensemble spread except for the southern hemisphere, and the SPDT scheme weakly compensates for the unresolved ensemble spread in the southern hemisphere (not shown). But, the underestimated ensemble BEC still remained in the near-surface atmosphere (not shown).

12.5 Discussion and Conclusions

Model uncertainties can be addressed by the stochastic perturbation schemes. Before taking this approach, it is recommended to examine where the underestimated ensembles occurred within the current status, because it helps to determine which stochastic perturbation system is suitable to compensate for the underestimating the ensemble spread. Our current experiment showed underestimation in ensemble spread for temperature and specific humidity in most of the troposphere, especially below 700 hPa. To solve this problem, we implemented the stochastic perturbation hybrid tendencies (SPHT) scheme that perturbs both dynamical and physical tendencies as the prior covariance inflation method. It simultaneously explains uncertainties that come from the computational representations of underlying partial differential equations and the imperfect physical parameterizations. As a result, most underestimated ensemble spread in the troposphere was alleviated.

However, the near-surface uncertainties over the land are still unresolved due to uncertain interaction between the land and atmosphere. In particular, land surface models (LSMs) contain heterogeneous land cover and soil texture, which is hard to capture in coarser model resolution. Since the parameters and parameterization schemes in LSMs contribute to the model uncertainties (Liu et al. 2023), the parameters or soil variable (e.g., soil temperature or soil moisture) can be perturbed to solve the near-surface uncertainty (MacLeod et al. 2016; Draper 2021). Because the atmosphere and the land surface have different scales and the ensemble dispersion is sensitive to perturbation, the random forcing tuning parameters should be carefully determined to generate adequate perturbations in LSMs (Bouttier et al. 2012; Lupo et al. 2020). Therefore, as a further study, the stochastic perturbation scheme in LSM can be used to improve the model uncertainty for near-surface variables and atmospheric variables below PBL through the heat flux changes.

Acknowledgements This work is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A6A1-A08025520) and by the NRF grant funded by the Korean government (MSIT) (NRF-2021R1A2C1-095535).

References

- Anderson JL, Anderson SL (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon Weather Rev* 127:2741–2758
- Berner J, Shutts GJ, Leutbecher M et al (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J Atmos Sci* 66:603–626
- Bouttier F, Vié B, Nuissier O et al (2012) Impact of stochastic physics in a convection-permitting ensemble. *Mon Weather Rev* 140:3706–3721
- Bowler NE, Clayton AM, Jardak M et al (2017) Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Q J R Meteorol Soc* 143:1280–1302
- Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q J R Meteorol Soc* 125:2887–2908
- Draper CS (2021) Accounting for land model uncertainty in numerical weather prediction ensemble systems: toward ensemble-based coupled land-atmosphere data assimilation. *J Hydrometeorol* 22:2089–2104
- Duc L, Saito K, Hotta D (2020) Analysis and design of covariance inflation methods using inflation functions. Part 1: theoretical framework. *Q J R Meteorol Soc* 146:3638–3660
- Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. *Q J R Meteorol Soc* 125:723–757
- Hong SY, Kwon YC, Kim TH et al (2018) The Korean Integrated Model (KIM) system for global weather forecasting. *Asia Pac J Atmos Sci* 54:267–292
- Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. *Mon Weather Rev* 126:796–811
- Hunt BR, Kostelich EJ, Szunyogh I (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* 230:112–126
- Isaksen L, Bonavita M, Buizza R et al (2010) Ensemble of data assimilations at ECMWF. ECMWF Tech Memo, p 45
- Kang JH, Chun HW, Lee S et al (2018) Development of an observation processing package for data assimilation in KIAPS. *Asia Pac J Atmos Sci* 54:303–318
- Kim JEE, Koo MS, Yoo C et al (2021) Seasonal performance of a nonhydrostatic global atmospheric model on a cubed-sphere grid. *Earth Space Sci*. <https://doi.org/10.1029/2021EA001643>
- Kleist DT, Ide K (2015) An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: system description and 3D-hybrid results. *Mon Weather Rev* 143:433–451
- Koo MS, Hong SY (2014) Stochastic representation of dynamic model tendency: formulation and preliminary results. *Asia Pac J Atmos Sci* 50:497–506
- Kotsuki S, Ota Y, Miyoshi T (2017) Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere. *Q J R Meteorol Soc* 143:2001–2015
- Leutbecher M, Lock SJ, Ollinaho P et al (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Q J R Meteorol Soc* 143:2315–2339
- Lim S, Koo MS, Kwon IH et al (2020) Model error representation using the stochastically perturbed hybrid physical-dynamical tendencies in ensemble data assimilation system. *Appl Sci*. <https://doi.org/10.3390/app10249010>
- Liu J, Fertig EJ, Li H et al (2008) Comparison between local ensemble transform Kalman filter and PSAS in the NASA finite volume GCM-perfect model experiments. *Nonlinear Process Geophys* 15:645–659
- Liu J, Yang ZL, Jia B et al (2023) Elucidating dominant factors affecting land surface hydrological simulations of the community land model over China. *Adv Atmos Sci* 40:235–250
- Luo X, Hoteit I (2013) Covariance inflation in the ensemble Kalman filter: a residual nudging perspective and some implications. *Mon Weather Rev* 141:3360–3368

- Lupo KM, Torn RD, Yang SC (2020) Evaluation of stochastic perturbed parameterization tendencies on convective-permitting ensemble forecasts of heavy rainfall events in New York and Taiwan. *Weather Forecast* 35:5–24
- MacLeod DA, Cloke HL, Pappenberger F et al (2016) Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Q J R Meteorol Soc* 142:79–90
- Mitchell HL, Houtekamer PL (2000) An adaptive ensemble Kalman filter. *Mon Weather Rev* 128:416–433
- Miyoshi T (2011) The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon Weather Rev* 139:1519–1535
- Palmer TN, Buizza R, Doblas-Reyes F et al (2009) Stochastic parametrization and model uncertainty. *ECMWF Tech Memo* 598:1–42
- Shin S, Kang JS, Jo Y (2016) The local ensemble transform Kalman filter (LETKF) with a global NWP model on the cubed sphere. *Pure Appl Geophys* 173:2555–2570
- Shin S, Kang JH, Chun HW et al (2018) Real data assimilation using the local ensemble transform Kalman filter (LETKF) system for a global non-hydrostatic NWP model on the cubed-sphere. *Asia Pac J Atmos Sci* 54:351–360
- Shutts G (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q J R Meteorol Soc* 131:3079–3102
- Thépaut JN (2003) Satellite data assimilation in numerical weather prediction: an overview. In: *Proceedings of the ECMWF seminar on recent developments in data assimilation for atmosphere and ocean*. ECMWF, Reading, UK, 8–12 Sept 2003, pp 75–96
- Whitaker JS, Hamill TM (2012) Evaluating methods to account for system errors in ensemble data assimilation. *Mon Weather Rev* 140:3078–3089
- Ying Y, Zhang F (2015) An adaptive covariance relaxation method for ensemble data assimilation. *Q J R Meteorol Soc* 141:2898–2906
- Zhang F, Snyder C, Sun J (2004) Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon Weather Rev* 132:1238–1253