Reduced non-Gaussianity by 30-second rapid update in convective-scale numerical weather prediction

Juan Ruiz[1,4], Guo-Yuan Lien[2], Keiichi Kondo[3], Shigenori Otsuka[4,5,6], and Takemasa Miyoshi[4,5,6,7,8]

[1]Centro de Investigaciones del Mar y la Atmósfera (CIMA-UBA/CONICET); Departamento de Ciencias de la Atmósfera y de los Océanos, FCEN, Universidad de Buenos Aires; Unidad Mixta Internacional-Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (UMI-IFAECI/CNRS-CONICET-UBA), Buenos Aires, Argentina
[2]Central Weather Bureau, Taipei, Taiwan
[3]Meteorological Research Institute, Tsukuba, Japan
[4]RIKEN Center for Computational Science, Kobe, Japan
[5]RIKEN Cluster for Pioneering Research, Kobe, Japan
[6]RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, Kobe, Japan
[7]University of Maryland, College Park, Maryland, USA
[8]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

Correspondence: Takemasa Miyoshi (takemasa.miyoshi@riken.jp), Juan Ruiz (jruiz@cima.fcen.uba.ar)

Abstract. Non-Gaussian forecast error is a challenge for ensemble-based data assimilation (DA), particularly for more nonlinear convective dynamics. In this study, we investigate the degree of non-Gaussianity of forecast error distributions at 1-km resolution using a 1000-member ensemble Kalman filter, and how it is affected by the DA update frequency and observation number. Regional numerical weather prediction experiments are performed with the SCALE (Scalable Computing for Advanced Library and Environment) model and the LETKF (Local Ensemble Transform Kalman Filter) assimilating every-30-second phased array radar observations. The results show that non-Gaussianity develops rapidly within convective clouds and is sensitive to the DA frequency and the number of assimilated observations. The non-Gaussianity is reduced by up to 40% when the assimilation window is shortened from 5 minutes to 30 seconds, particularly for vertical velocity and radar reflectivity.

1 Introduction

The Kalman filter (KF) is the minimum variance linear unbiased estimator of the state of a dynamical system. The Ensemble Kalman Filter (EnKF, Evensen, 2009; Houtekamer and Zhang, 2016) is a Monte Carlo extension to the KF suitable for nonlinear systems with a large number of variables, so that it became a viable choice for data assimilation (DA) in numerical weather prediction (NWP) and other geoscience applications. The EnKF is optimal in the sense of maximum likelihood estimation when the error distributions are Gaussian (Evensen, 2009), but it becomes sub-optimal when the observational and forecast error distributions depart from the Gaussian (Lei et al., 2010). Miyoshi et al. (2014); Miyoshi et al. (2015) and Kondo and Miyoshi (2018) investigated non-Gaussianity in forecast error distributions using a 10,240 member EnKF with global atmospheric models. They showed that large non-Gaussianity measured by the Kullback-Leibler divergence is found frequently in the tropics mainly due to abundance of deep moist convection and also in other active areas with a real-world NWP model at

20 relatively low 112-km resolution. In those experiments, temperature and moisture show generally more non-Gaussian distributions than winds.

Recently the horizontal resolution of operational NWP systems reached the order of 1 km, fine enough to resolve convective phenomena explicitly. Obtaining appropriate initial conditions at such high resolution is a challenge (Sun et al., 2014). The EnKF has been successfully applied to mesoscale data assimilation of radars and satellites (e.g., Stensrud et al., 2013). However,

25 previous studies (Jacques and Zawadzki, 2014; Kawabata and Ueno, 2020) revealed that the underlying assumptions such as linear error dynamics and Gaussian error distributions are much more challenging in mesoscale than in synoptic and larger scales.

Miyoshi et al. (2016a, b) developed a so-called Big Data Assimilation (BDA) system assimilating observations every 30 seconds at 100-m resolution, taking advantage of new-generation technologies like the phased array weather radar (PAWR)

30 which provides observations at unprecedented high temporal and spatial resolution. With the BDA configuration under an idealized Observing System Simulation Experiment (OSSE) framework, Maejima and Miyoshi (2020) showed that every-1-minute DA cycles resulted in better analyses than every-15-minute cycles. However, the impact of the DA frequency upon the forecast error distribution has not been investigated in real-case convective scale NWP.

This study investigates how the DA frequencies affect non-Gaussianity using a 1000-member, 1-km-mesh EnKF. 1000

35 ensemble members would be useful to detect non-Gaussian forecast error distributions as suggested by Kondo and Miyoshi (2019). Necker et al. (2020a, b) performed similar experiments and investigated the covariance structure and the effect of sampling noise at the mesoscale in a heavy rain event over Germany. Although the previous research employed data assimilation with only conventional observations at 3-hourly DA frequency, this study is fundamentally different in the convection-resolving rapid DA cycles with PAWR data as frequently as every 30 seconds. The high frequency data allows us to investigate the sources

40 of non-Gaussian distributions at the kilometer scale in the presence of rapidly-evolving deep moist convection. The paper is organized as follows: Section 2 describes methodological aspects. Results are presented in Section 3, and Section 4 provides concluding remarks.

2 Methodology

We use observations from the PAWR at Osaka University, Suita, Japan (Yoshikawa et al., 2013, Fig. 1a, red cross). This PAWR

45 provides a unique dataset suitable for this study with various assimilation frequencies up to every 10 seconds at the fastest. This study follows the case study of Miyoshi et al. (2016a) focusing on the period between 0400 and 0600 UTC July 13, 2013, when heavy rains produced flash floods in Kyoto. Individual convective cells moved from west to east within a quasi-stationary intense rainband (see Fig. 1b for a snapshot at 0530 UTC). For this period, full volume scans of the PAWR are available every 30 seconds with 98 elevation angles, azimuthal resolution of $1.2°$, and range resolution of 100 meters up to a maximum range

50 of 60 km (Fig. 1a, red circle). Unambiguous Doppler velocities are available in the range -50 to $50 \ ms^{-1}$. PAWR reflectivity data is quality-controlled following Ruiz et al. (2015). A simple quality control algorithm has also been applied to the Doppler velocity field to remove outliers.

In this study, the regional NWP model known as the Scalable Computing for Advanced Library and Environment model (SCALE, Nishizawa et al., 2015) is used, coupled with the local ensemble transform Kalman filter (LETKF, Hunt et al., 2007). Lien et al. (2017) and Honda et al. (2018) describe the SCALE-LETKF system in detail. The model configuration follows Lien et al. (2017) with a single-moment bulk microphysics scheme (Tomita, 2008), a level-2.5 boundary layer turbulence scheme (Nakanishi and Niino, 2004), the Model Simulation Radiation Transfer radiation scheme (Sekiguchi and Nakajima, 2008), and soil processes represented by a Beljaars-type soil model (Beljaars and Holtslag, 1991).

The SCALE-LETKF system is implemented over a single domain with horizontal resolution of 1 km, 50 vertical sigma levels, and a size of 180 km by 180 km (Fig. 1a). A 1000-member ensemble is used to assimilate the observations. Kondo and Miyoshi (2019) showed significant sampling error contaminations in non-Gaussian measures when the ensemble size is smaller than 1000. The initial conditions for the first cycle and the boundary conditions are taken from the National Centers for Environmental Prediction Global Data Assimilation System final analysis (FNL). Using FNL as the boundary conditions may be overly optimistic for the forecasting purpose, but this is not relevant to the goal of this study which focuses on non-Gaussian distributions and the impact of DA frequency. The boundary-condition ensemble is perturbed by adding balanced large scale random perturbations following (Necker et al., 2020a). These perturbations are generated by taking differences of the Climate Forecast System Reanalysis Saha et al. (2010) fields corresponding to randomly selected dates in the same season at the same time of the day. The perturbations are scaled by a multiplicative factor of 0.1 so that the amplitude of the perturbations is roughly equivalent to 10% of the climatological variability. All variables including soil variables are perturbed.

In the SCALE-LETKF system, radar data can be assimilated using different localization scales for different variables. Based on preliminary experiments with the SCALE-LETKF using smaller ensemble sizes and every-30-second PAWR data, it was found that vertical localization scale of $2km$ (with a $7.3km$ cut-off, similarly hereafter) produced good results. For horizontal localization, better results were obtained using $4-km$ localization to assimilate observations with reflectivities $> 10dBZ$. Observations of reflectivity values $\leq 10dBZ$ are assimilated with a fixed value of $10dBZ$ to avoid large observation-minus-forecast departures associated with clear air reflectivities (Aksoy et al., 2009). Also, a shorter horizontal localization scale of $2km$ is used to reduce the impact of no-rain observations at the edge of clouds. Doppler velocity observations are assimilated with horizontal and vertical localization scales of $10km$ and $3km$, respectively. A relaxation to prior ensemble spread (RTPS, Whitaker and Hamill, 2012) with a relaxation parameter of 0.9 is applied. This helps consider the inhomogeneous distribution of observations as in Lien et al. (2017).

Reflectivity and Doppler velocity observations are superobbed to horizontal resolution of $1km$ and vertical resolution of $500m$ to match the model resolution. The observational error standard deviations for these super-observations are set at $5.0dBZ$ and $3.0ms^{-1}$ for reflectivity and Doppler velocities, respectively. The radar data are assimilated up to a maximum height of $11km$.

A spin-up DA experiment with every-5-minute PAWR reflectivity and Doppler velocity data is performed for an hour from 0400 UTC, July 13, 2013. Only a single PAWR volume scan closest to the analysis time is assimilated per analysis. The 1000-member analysis ensemble at 0500 UTC is used as the initial conditions for the DA experiments.
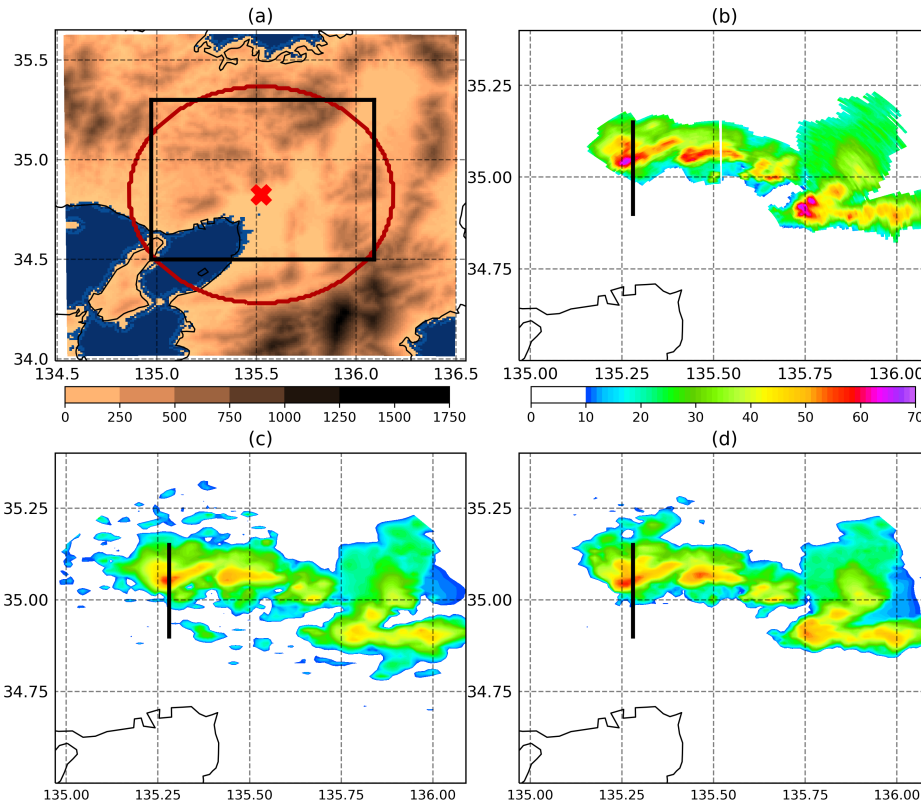
Figure 1. (a) Terrain height of the 1-km-mesh SCALE-LETKF domain (shades, m). The red circle indicates the 60-km radar range centered at the radar site (red cross) in Osaka University, Suita, Japan. The black box indicates the area shown in (b-d). (b) column-maximum PAWR observation (dBZ) at 0530 UTC, half an hour after the initialization of the data assimilation cycle, (c) 5MIN and (d) 30SEC experiments analysis ensemble-mean column-maximum radar reflectivity (dBZ) at 0530 UTC. Black lines indicate the locations of the cross-sections displayed in Fig. 2

Experiments are performed with different DA update frequency to study the impacts of DA frequency and observation number on the forecast error distributions. All experiments share the configuration described above, but the only differences are the DA frequency and the amount of the data assimilated. First, four experiments with 5, 2, 1, and 0.5 minutes DA frequencies
90   are performed, hereafter referred to as 5MIN, 2MIN, 1MIN, and 30SEC, respectively. Here, only a single volume scan closest to the analysis time is used per analysis. Namely, more frequent updates assimilate more data.

Next, to separate the impact of DA frequency and the amount of data assimilated, two additional experiments are performed using a 5-minute and 1-minute DA frequency, with all radar volumes every 30 seconds assimilated by a 4-dimensional EnKF approach Hunt et al. (2004). These experiments are referred to as 5MIN-4D and 1MIN-4D, respectively, assimilating the same
95   amount of data as 30SEC but using longer assimilation windows.

To measure the degree of non-Gaussianity of the error distributions we compute the Kullback-Leibler divergence (hereafter KLD, Kullback and Leibler, 1951) which is defined as follows:

$$KLD(P\|Q) = \int\limits_{-\infty}^{\infty} p(x)\ln\frac{p(x)}{q(x)}dx, \tag{1}$$

where $P(x)$ and $Q(x)$ are two probability density functions (PDFs). In our case $P(x)$ is the ensemble-based sample distri-
100   bution of x, and $Q(x)$ is a Gaussian distribution whose mean and standard deviation are given by the ensemble-based sample estimates. The KLD is 0 if P and Q are the same. Therefore, a low KLD value corresponds to the sample distribution close to a Gaussian. To compute the KLD for different variables and at different grid points, we approximate $P(x)$ with the sample histogram populated from the 1000-member ensemble using 32 equally-sized bins covering the range where $P(x)$ is greater than 0.

105   **3   Results**

All experiments show that the analyzed reflectivity fields are in good agreement with the observation. However, some differences can be found between the experiments that assimilate different amounts of data and with different assimilation windows. For example, Figures 1c and d show that 30SEC captures the strong reflectivity areas (>45 dBZ, orange and red shadings) better than 5MIN. 5MIN shows noisy patterns of spurious convective cells surrounding the main convective rainband.
110   First, the impact of data assimilation frequency is explored by the 5MIN, 2MIN, 1MIN, and 30SEC experiments. Here, more observations are assimilated with more frequent data assimilation. Figure 2 top row (a-d) shows that the reflectivity (Z) patterns (shades) are similar among all experiments, but vertical velocity (W, contours) are different. Figure 2e shows strong non-Gaussianity in the first-guess ensemble in W and temperature (T). KLD for W and T are consistently reduced with more frequent DA (Fig. 2e-h, shades and blue contours), although the reduction is smaller for T. Overall, KLD is reduced more from
115   5MIN to 2MIN than from 1MIN to 30SEC. The forecast error distributions for W and Z at the location of maximum KLD for W show some discrepancies from the Gaussian distribution (Figs. 2i-p). The ensemble spread for W is reduced significantly from 5MIN to 2MIN (Fig. 2e-h, red contours). 5MIN shows strong non-Gaussianity for W at the southern edge and the highest peak of the convective cell (Fig. 2e), which is probably related to the development of a new updraft in the southern edge and the top of the strong updraft, respectively. Weaker low-level maxima south of the convective line are associated with shallow
120   convective clouds that are not effectively corrected by radar observations. Kondo and Miyoshi (2019) found that in synoptic scales, the ensemble spread maxima are co-located with the KLD maxima. At convective scales for W, the ensemble spread maxima (Fig. 2e-h, red contours) are slightly out of phase with respect to the KLD maxima (shades). The KLD maxima for T (blue contours) are approximately collocated as those for W (shades). These KLD maxima can be associated with non-Gaussianity in W through vertical advection of scalar quantities such as T and moisture. Another KLD maximum for T is
125   found near the surface south of the convective cell, probably associated with the gust front.
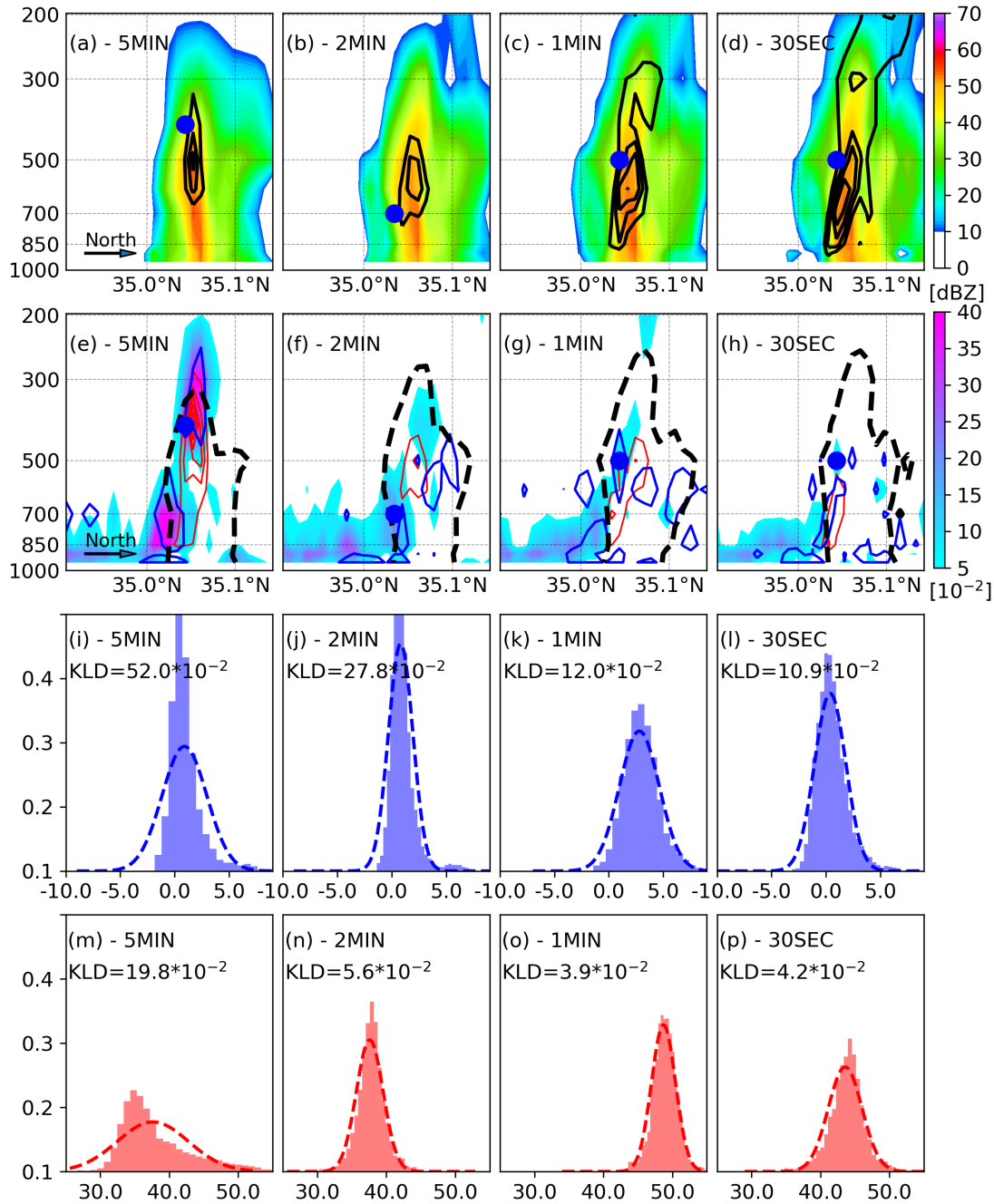
Figure 2. (a-h) South-North vertical cross-section along the black line indicated in Fig. 1b-d at 0530 UTC for (a-d) first-guess ensemble-mean reflectivity (Z, shades, dBZ) and vertical velocity (W, contours every 2.5 ms^{-1}), (e-h) vertical velocity KLD (shades, 10^{-2}) and variance (ensemble spread, red contours every $3ms^{-1}$), and temperature KLD over 0.04 (blue contours). Blacked dashed contours indicate reflectivity over 30 dBZ. (i-p) Sample histograms for (i-l) vertical velocity and (m-p) reflectivity at the location of the maximum KLD for vertical velocity (blue circles in a-h). Thick dashed curves indicate fitted Gaussian functions, and KLD non-Gaussian measures are also indicated. Each column corresponds to 5MIN, 2MIN, 1MIN, and 30SEC from left to right, respectively.
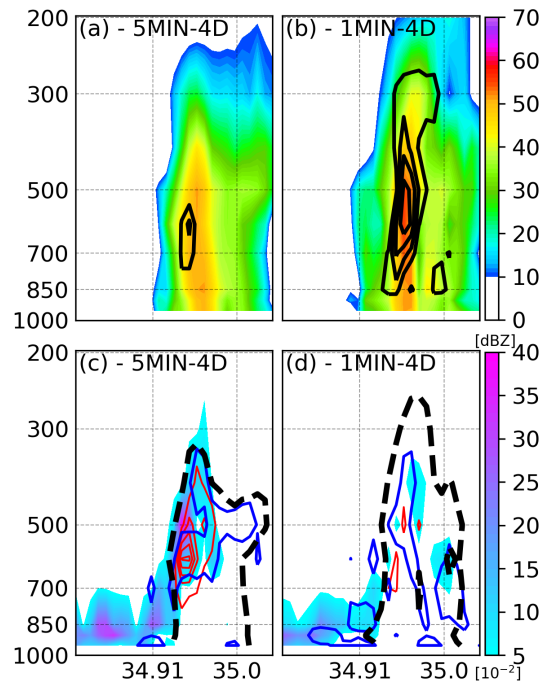
Figure 3. Similar to Fig. 2a-h but for (a,c) 5MIN-4D and (b,d) 1MIN-4D.

4D-EnKF experiments allow us to investigate the impact of changing the assimilation frequency while keeping the observation number unchanged. 5MIN-4D shows almost the same ensemble spread for W (Fig. 3c, red contours). KLD for W (Fig. 3c, shades) is lower, indicating that the observation number contributes to reducing non-Gaussianity. This is not the case for T for which KLD is similar or larger (Fig. 3c, blue contours). 1MIN-4D is close to 1MIN and 30SEC in terms of non-Gaussianity and the shape and strength of the convective cell (Figs. 3b and d).

We further investigate the non-Gaussianity by averaging the KLD vertically and temporally (Fig. 4). In 5MIN, the central and eastern side of the convective area shows relatively low KLD values because the impact of radar DA is generally bigger in the convective areas (Fig. 4a).

KLD consistently decreases with increasing DA frequency (Figs. 4b-d). KLD is reduced by up to 40% in 30SEC with respect to the 5MIN. KLD is reduced more in the convective area, where more observations are assimilated. Increasing the DA frequency and the observation number produce a more substantial impact over the western part of the convective line where KLD maxima are found associated with convective cells entering the radar range from the West.

KLD in 1MIN-4D is as low as that in 30SEC and lower than that in 1MIN. This result suggests that both observation number and DA frequency contribute to reducing non-Gaussianity, at least for high DA frequencies. KLD in 5MIN-4D is lower than that in 5MIN, so that a larger observation number helps to reduce non-Gaussianity. However, KLD in 5MIN-4D is larger than that in 30-SEC or 1MIN-4D, so that DA frequency is equally important. Moreover, the impact of DA frequency can be larger
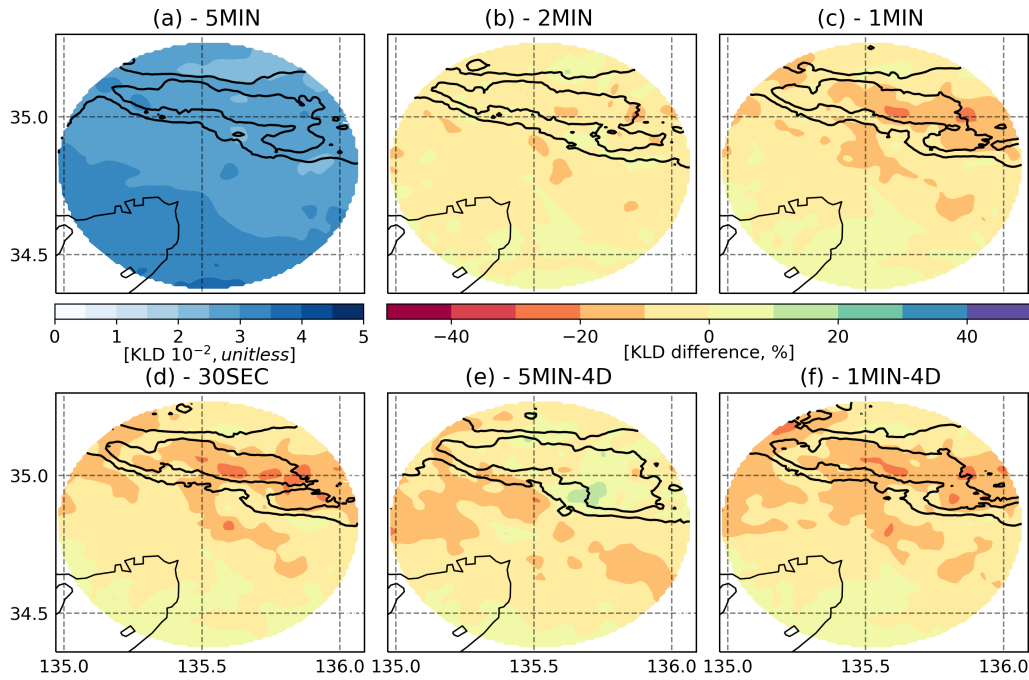
Figure 4. (a) Column-averaged KLD for zonal wind for 5MIN, averaged for the experiment period from 0500 to 0600 UTC. Relative KLD difference (%) from 5MIN for (b) 2MIN, (c) 1MIN, (d) 30SEC, (e) 5MIN-4D, and (f) 1MIN-4D. Warm colors correspond to smaller KLD values.

in the case of variables like T and moisture. As already found in the vertical cross-sections (Fig. 3), for those variables, KLD in 5MIN and 5MIN-4D is almost the same, while KLD is clearly reduced for 1MIN, 1MIN-4D, and 30SEC (not shown).

We also investigate the vertical distribution of non-Gaussianity by the spatially averaged vertical profile of KLD at "rain"
145 grid points, defined by the ensemble-mean column-maximum reflectivity $> 30 dBZ$, and "no-rain" grid points, defined by the ensemble-mean column-maximum reflectivity $< 0 dBZ$. At the rain grid points (Figures 5a-d) KLD for temperature and vertical velocity is maximum at mid-levels coinciding with the maximum in latent heat release within convective clouds and with the maximum ensemble spread for these two variables (not shown). KLD for temperature, vertical velocity, and specific humidity maximizes at lower heights over the no-rain area since, as stated before, at such locations non-Gaussianity is mainly
150 associated with shallow convection. For instance, for the vertical velocity, the ensemble spread in the shallow convection is usually low, but the KLD can be larger. An upper-level maximum in KLD is found for the meridional wind (Figs. 5d,h), also coinciding with the maximum ensemble spread (not shown). Convective outflows are stronger at the top of convective clouds and can be one of the mechanisms contributing to increase non-Gaussianity at these levels over the rain area. Overall, KLD in 30SEC is lower than that in 5MIN with reductions of more than 40%. The reduction of KLD in the no-rain area is smaller
155 because the radar DA is inherently less effective in these areas (Figs. 5e-h).
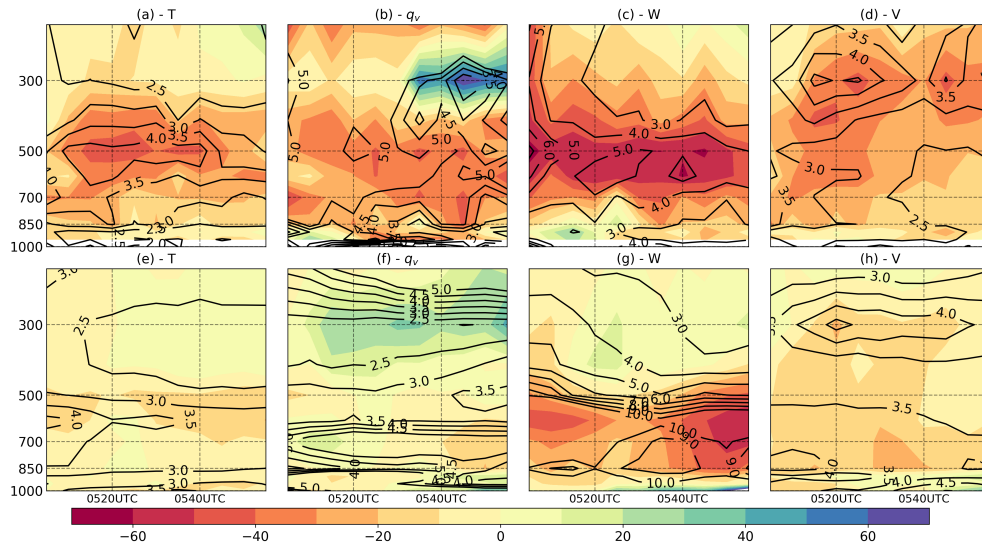
Figure 5. Time-vertical cross-section of KLD for 5MIN (contours, 10^{-2}) and the relative difference for 30SEC (shaded), averaged over the (a-d) "rain" ($>$30dBZ) and (e-h) "no-rain" ($<$0dBZ) grid points for (a,e) temperature, (b,f) specific humidity, (c,g) vertical velocity and (d,h) meridional wind.

4   Conclusions

1000-member 1-km-resolution ensemble DA experiments were performed using real phased array radar observations and a mesoscale NWP model to investigate the impact of DA frequency and observation number on the non-Gaussian error distributions. We found that a DA frequency of 5 minutes, although it was already much faster than the typical DA frequency, resulted in strong non-Gaussianity possibly affecting the performance of the EnKF. Non-Gaussianity is stronger for vertical velocity as has been found by Kawabata and Ueno (2020). Non-Gaussianity is also larger at mid-levels within convective cells, near the level of larger latent-heat release and vertical accelerations associated with convective instability. At convective scales, some of the local maxima in KLD can be related directly to advection by mesoscale circulations associated with strong convective cells, but other processes not specifically presented in this study may also possibly contribute to the generation of non-Gaussianity, such as those not directly associated with clouds, like differential heating circulations or gravity waves. We found that increasing the analysis update frequency and observation number from 5 minutes to 30 seconds has a huge impact upon non-Gaussianity in the error distributions for all model variables but particularly for vertical velocity and reflectivity which are the ones showing larger KLD from Gaussianity at these scales. Increasing the assimilation frequency to 30 seconds and assimilating more observations can reduce KLD by up to 40%. Moreover, 4D-EnKF experiments revealed that for frequent DA of every 1 minute, the observation number explained most of the reduction in non-Gaussianity; in contrast, for a longer window of 5 minutes, even the experiments using all 30-second-frequency observations presents significant departures from the Gaussian. While convective clouds are particularly favorable for nonlinear error growth, non-Gaussianity is not necessarily larger within convective clouds. This is mainly due to the convective-scale radar DA is usually most effective within precipitating clouds. This is the

first attempt to investigate the impact of assimilation frequency and observation number on non-Gaussianity using an EnKF
employing a large 1000-member ensemble and every-30-second observations from a PAWR. In this first set of experiments, we
evaluate the impact on the non-Gaussianity of the ensemble-based sample distribution. Future experiments will be performed
to investigate the overall quality of the analysis obtained with different assimilation windows and number of observations and
also the impact of assimilation window upon the structure of the error covariance matrix.

Code and data availability.   The codes used for the main results of this study can be accessed at a public github repository (https://github.com/takemasa-miyoshi/letkf). Essential data to reproduce the results of this study are stored for 5 years in RIKEN R-CCS. Due to the large volume of data and limited disk space, data will be shared online upon request (takemasa.miyoshi@riken.jp). The phased array radar data can be visualized at https://pawr.nict.go.jp/index_en.html

Author contributions.   All the authors participate in the conception of the ideas, in the design of the experiments and in draft manuscript preparation. Guo-Yuan Lien wrote the SCALE-LETKF code used in this study. Juan Ruiz conducted the experiments and analyzed the results. All authors reviewed the results and approved the final version of the manuscript.

Competing interests.   The authors declare that they have no conflict of interest.

Acknowledgements.   We thank the members of RIKEN CCS Data Assimilation Research Team and JST AIP project for valuable discussions. The PAWR data was provided by the National Institute of Information and Communications Technology (NICT) science cloud system. This work was supported by CREST, JST projects 'EBD: Extreme Big Data – Convergence of Big Data and HPC for Yottabyte Processing' (grant number: JPMJCR1303), 'Innovating "Big Data Assimilation" technology for revolutionizing very-short-range severe weather prediction' (grant number: JPMJCR1312), JST AIP acceleration research 'Big Data Assimilation and AI Creating New Development in Real-time Weather Prediction' (grant number JPMJCR19U2) and Advancement of meteorological and global environmental predictions utilizing observational "Big Data" of the social and scientific priority issues (Theme 4) to be tackled by using FUGAKU computer of the FLAGSHIP2020 Project of the Ministry of Education, Culture, Sports. This research used computational resources of the K-computer at RIKEN R-CCS (Project IDs ra000015, ra001011, hp150019, hp160162, hp170178, hp180062, hp190051), Post-K project ID hp 120282 and the Joint Center for Advanced High Performance Computing (JCAHPC) Oakforest-PACS (Project IDs hp190051 and hp200026). This study was also supported by grants PICT-2033/2017 for Agencia Nacional de Promoción Científica y Tecnológica and 20020170100504 BA from the University of Buenos Aires, and by JSPS KAKENHI Grant Number JP16K17807.

References

200   Aksoy, A., Dowell, D. C., and Snyder, C.: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses, Monthly Weather Review, 137, 1805 – 1824, https://doi.org/10.1175/2008MWR2691.1, 2009.

Beljaars, A. C. M. and Holtslag, A. A. M.: Flux Parameterization over Land Surfaces for Atmospheric Models, Journal of Applied Meteorology, 30, 327–341, https://doi.org/10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2, 1991.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems Magazine, 29, 83–104,
205   https://doi.org/10.1109/MCS.2009.932223, 2009.

Honda, T., Miyoshi, T., Lien, G.-Y., Nishizawa, S., Yoshida, R., Adachi, S. A., Terasaki, K., Okamoto, K., Tomita, H., and Bessho, K.: Assimilating All-Sky Himawari-8 Satellite Infrared Radiances: A Case of Typhoon Soudelor (2015), Monthly Weather Review, 146, 213–229, https://doi.org/10.1175/MWR-D-16-0357.1, 2018.

Houtekamer, P. L. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144,
210   4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1, 2016.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, Tellus A: Dynamic Meteorology and Oceanography, 56, 273–277, https://doi.org/10.3402/tellusa.v56i4.14424, 2004.

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

215   Jacques, D. and Zawadzki, I.: The Impacts of Representing the Correlation of Errors in Radar Data Assimilation. Part I: Experiments with Simulated Background and Observation Estimates, Monthly Weather Review, 142, 3998 – 4016, https://doi.org/10.1175/MWR-D-14-00104.1, 2014.

Kawabata, T. and Ueno, G.: Non-Gaussian Probability Densities of Convection Initiation and Development Investigated Using a Particle Filter with a Storm-Scale Numerical Weather Prediction Model, Monthly Weather Review, 148, 3 – 20, https://doi.org/10.1175/MWR-D-
220   18-0367.1, 2020.

Kondo, K. and Miyoshi, T.: Non-Gaussian statistics in global atmospheric dynamics: a study with a 10 240-member ensemble Kalman filter using an intermediate AGCM, Nonlinear Processes in Geophysics Discussions, 2018, 1–42, https://doi.org/10.5194/npg-2018-52, 2018.

Kondo, K. and Miyoshi, T.: Non-Gaussian statistics in global atmospheric dynamics: a study with a 10 240-member ensemble Kalman filter using an intermediate atmospheric general circulation model, Nonlinear Processes in Geophysics, 26, 211–225,
225   https://doi.org/10.5194/npg-26-211-2019, 2019.

Kullback, S. and Leibler, R. A.: On Information and Sufficiency, Ann. Math. Statist., 22, 79–86, https://doi.org/10.1214/aoms/1177729694, 1951.

Lei, J., Bickel, P., and Snyder, C.: Comparison of Ensemble Kalman Filters under Non-Gaussianity, Monthly Weather Review, 138, 1293 – 1306, https://doi.org/10.1175/2009MWR3133.1, 2010.

230   Lien, G.-Y., Miyoshi, T., Nishizawa, S., Yoshida, R., Yashiro, H., Adachi, S. A., Yamaura, T., and Tomita, H.: 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, 2017.

Maejima, Y. and Miyoshi, T.: Impact of the Window Length of Four-Dimensional Local Ensemble Transform Kalman Filter: A Case of Convective Rain Event, SOLA, 16, 37–42, https://doi.org/10.2151/sola.2020-007, 2020.

235   Miyoshi, T., Kondo, K., and Imamura, T.: The 10,240-member ensemble Kalman filtering with an intermediate AGCM, Geophysical Research Letters, 41, 5264–5271, https://doi.org/10.1002/2014GL060863, 2014.

Miyoshi, T., Kondo, K., and Terasaki, K.: Big Ensemble Data Assimilation in Numerical Weather Prediction, Computer, 48, 15–21, https://doi.org/10.1109/MC.2015.332, 2015.

Miyoshi, T., Kunii, M., Ruiz, J., Lien, G.-Y., Satoh, S., Ushio, T., Bessho, K., Seko, H., Tomita, H., and Ishikawa, Y.: "Big Data
240   Assimilation" Revolutionizing Severe Weather Prediction", Bulletin of the American Meteorological Society, 97, 1347 – 1354, https://doi.org/10.1175/BAMS-D-15-00144.1, 2016a.

Miyoshi, T., Lien, G., Satoh, S., Ushio, T., Bessho, K., Tomita, H., Nishizawa, S., Yoshida, R., Adachi, S. A., Liao, J., Gerofi, B., Ishikawa, Y., Kunii, M., Ruiz, J., Maejima, Y., Otsuka, S., Otsuka, M., Okamoto, K., and Seko, H.: "Big Data Assimilation" Toward Post-Petascale Severe Weather Prediction: An Overview and Progress, Proceedings of the IEEE, 104, 2155–2179,
245   https://doi.org/10.1109/JPROC.2016.2602560, 2016b.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification, Boundary-Layer Meteorology, 112, 1–31, https://doi.org/doi.org/10.1023/B:BOUN.0000020164.04146.98, 2004.

Necker, T., Geiss, S., Weissmann, M., Ruiz, J., Miyoshi, T., and Lien, G.-Y.: A convective-scale 1,000-member ensemble simulation and potential applications, Quarterly Journal of the Royal Meteorological Society, 146, 1423–1442,
250   https://doi.org/https://doi.org/10.1002/qj.3744, 2020a.

Necker, T., Weissmann, M., Ruckstuhl, Y., Anderson, J., and Miyoshi, T.: Sampling Error Correction Evaluated Using a Convective-Scale 1000-Member Ensemble, Monthly Weather Review, 148, 1229 – 1249, https://doi.org/10.1175/MWR-D-19-0154.1, 2020b.

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

255   Ruiz, J. J., Miyoshi, T., Satoh, S., and Ushio, T.: A Quality Control Algorithm for the Osaka Phased Array Weather Radar, SOLA, 11, 48–52, https://doi.org/10.2151/sola.2015-011, 2015.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue,
260   Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Sekiguchi, M. and Nakajima, T.: A k-distribution-based radiation code and its computational optimization for an atmospheric general circulation model, Journal of Quantitative Spectroscopy and Radiative Transfer, 109, 2779 – 2793,
265   https://doi.org/https://doi.org/10.1016/j.jqsrt.2008.07.013, 2008.

Stensrud, D. J., Wicker, L. J., Xue, M., Dawson, D. T., Yussouf, N., Wheatley, D. M., Thompson, T. E., Snook, N. A., Smith, T. M., Schenkman, A. D., Potvin, C. K., Mansell, E. R., Lei, T., Kuhlman, K. M., Jung, Y., Jones, T. A., Gao, J., Coniglio, M. C., Brooks, H. E., and Brewster, K. A.: Progress and challenges with Warn-on-Forecast, Atmospheric Research, 123, 2–16, https://doi.org/https://doi.org/10.1016/j.atmosres.2012.04.004, 6th European Conference on Severe Storms 2011. Palma de Mallorca,
270   Spain, 2013.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., Xu, M., and Pinto, J.: Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges, Bulletin of the American Meteorological Society, 95, 409 – 426, https://doi.org/10.1175/BAMS-D-11-00263.1, 2014.

Tomita, H.: New Microphysical Schemes with Five and Six Categories by Diagnostic Generation of Cloud Ice, Journal of the Meteorological Society of Japan. Ser. II, 86A, 121–142, https://doi.org/10.2151/jmsj.86A.121, 2008.

Whitaker, J. S. and Hamill, T. M.: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation, Monthly Weather Review, 140, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1, 2012.

Yoshikawa, E., Ushio, T., Kawasaki, Z., Yoshida, S., Morimoto, T., Mizutani, F., and Wada, M.: MMSE Beam Forming on Fast-Scanning Phased Array Weather Radar, IEEE Transactions on Geoscience and Remote Sensing, 51, 3077–3088, https://doi.org/10.1109/TGRS.2012.2211607, 2013.