On the properties of ensemble forecast sensitivity to observations

Shunji Kotsuki1,2 | Kenta Kurosawa1 | Takemasa Miyoshi1,2,3,4,5

1 Data Assimilation Research Team, RIKEN Center for Computational Science, Kobe, Japan
2 RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, Kobe, Japan
3 Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
4 Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland,
5 Prediction Science Laboratory, RIKEN Cluster for Pioneering Research, Kobe, Japan

Correspondence
Shunji Kotsuki and Takemasa Miyoshi, RIKEN Center for Computational Science (R-CSCS), 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. Email: shunji.kotsuki@riken.jp; takemasa.miyoshi@riken.jp

Funding information
Japan Aerospace Exploration Agency, Precipitation Measuring Mission RA8; Japan Society for the Promotion of Science, KAKENHI grants JP15K18128 and JP18H01549; Ministry of Education, Culture, Sports, Science and Technology, post K computer of the FLAGSHIP2020 Project (Thema, the Initiative for Excellent Young Researchers; Japan Aerospace Exploration Agency

Evaluating impacts of observations on the skill of numerical weather prediction (NWP) is important. The Ensemble Forecast Sensitivity to Observation (EFSO) provides an efficient approach to diagnosing observation impacts, quantifying how much each observation improves or degrades a subsequent forecast with a given verification reference. This study investigates the sensitivity of EFSO impact estimates to the choice of the verification reference, using a global NWP system consisting of the Non-hydrostatic Icosahedral Atmospheric Model (NICAM) and the Local Ensemble Transform Kalman Filter (LETKF). The EFSO evaluates observation impacts with the moist total energy norm and with recently proposed observation-based verification metrics. The results show that each type of observation mainly contributes to the improvement of forecast departures of the observed variable maybe due to the limitation of localization in the EFSO. The EFSO overestimates the fraction of beneficial observations when verified with subsequent analyses, especially for shorter lead times such as 6 h. We may avoid this overestimation to some extent by verifying with observations, analyses from other data assimilation (DA) systems, or analyses of an independent run with the same DA system. In addition, this study demonstrates two important issues possibly leading to overestimating observation impacts. First, observation impacts would be overestimated if we apply relaxation-to-prior methods to the initial conditions of the ensemble forecasts in the EFSO; therefore, the ensemble forecasts in the EFSO should be independent of the ensemble forecasts in the DA cycle. Second, deterministic baseline forecasts of the EFSO, which represent the forecast without DA, should be initialized by the ensemble mean of the first guess at the analysis time, not by the previous analysis.

KEYWORDS
data assimilation, ensemble forecast sensitivity to observations, ensemble Kalman filter, NICAM-LETKF, verification reference

1 | INTRODUCTION

Data assimilation (DA) plays an important role in numerical weather prediction (NWP) by providing accurate initial conditions through the combination of available observations and forecasts. Present NWP systems assimilate a large number of observation data from various platforms such as radiosondes, aircraft and satellites. It is important to evaluate the impacts of the assimilated observations on forecast skills. Impact estimates can be used to improve treatments
of observations such as quality control, data thinning, and bias correction. Data denial experiments, with and without some of the observation data (Kelly et al., 2007), are a basic approach to estimate observation impacts. However, data denial experiments have become more complex in modern DA systems because of the presence of many types of observation. In addition, data denial experiments require computationally expensive long-term DA cycles to gain statistical significance (Geer, 2016).

The Forecast Sensitivity to Observation (FSO) method is a computationally inexpensive method to estimate observation impacts. The FSO quantifies how much each observation improves or degrades a subsequent forecast. Langland and Baker (2004) proposed an adjoint FSO formulation for variational DA systems. The adjoint FSO has been implemented with operational systems (Zhu and Gelaro, 2008; Cardinali, 2009; Lorenc and Marriott, 2014). The Langland and Baker (2004) approach is a powerful tool for diagnosing observation impacts, but it requires adjoint operators of the forecast model. Liu and Kalnay (2008) and Li (2004) approach is a powerful tool for diagnosing variational DA systems. The adjoint FSO has been implemented with operational systems (Zhu and Gelaro, 2008; Cardinali, 2009; Lorenc and Marriott, 2014). The Langland and Baker (2004) approach is a powerful tool for diagnosing observation impacts, but it requires adjoint operators of the forecast model. Liu and Kalnay (2008) and Li et al. (2010) proposed an Ensemble FSO (EFSO) for the Local Ensemble Transform Kalman Filter (LETKF: Hunt et al., 2007) without using the adjoint operators. Kalnay et al. (2012, hereafter K12) simplified the Liu and Kalnay (2008) formulation to render it applicable to any kinds of ensemble Kalman filter (EnKF: Evensen, 1994). Ota et al. (2013, hereafter O13) implemented the EFSO with the National Centers for Environmental Prediction (NCEP)'s operational global EnKF system. Kotsuki et al. (2019) used the EFSO to estimate impacts of assimilated observations to study predictability of record-breaking heavy rainfall in Japan in July 2018. Hotta et al. (2017, hereafter H17) proposed a proactive quality control method (PQC) that detects detrimental observations with the EFSO. Lien et al. (2018) proposed an efficient approach to assimilating a new observing system using the EFSO. Recently, Buehner et al. (2018) proposed a new hybrid FSO, in which DA uses the adjoint model and FSO employs ensemble forecasts, for a hybrid ensemble-variational data assimilation system.

The FSO methods estimate observation impacts by measuring errors of two forecasts with and without DA. Most FSO studies for NWP systems have measured errors with dry and moist energy norms relative to reference atmospheric fields. Subsequent analysis fields of DA systems (hereafter, subsequent analyses) are usually used as the reference. It is also possible to use analysis fields from other DA systems as the reference (Lien et al., 2018). Recently, Todling (2013) proposed using observation departures (i.e. observations minus forecasts) for the verification metrics. Sommer and Weissmann (2016, hereafter SW16) and Necker et al. (2018, hereafter N18) tested the observation-based metrics for the Deutscher Wetterdienst (DWD) operational regional ensemble DA system. Similarly, Cardinali (2018, hereafter C18) introduced an observation-based metric for the adjoint FSO of the European Centre for Medium-range Weather Forecasting (ECMWF)'s global system. SW16, N18 and C18 reported that commonly used dry and moist energy norms have the following disadvantages compared with the observation-based metrics: (a) the analysis error would be correlated with the short-term forecast error, (b) the analysis is potentially affected by a model bias, and (c) the energy norms, which are mainly dominated by the errors in the troposphere, evaluate improvements in stratospheric fields insufficiently.

This study aims to investigate the following three specific issues of the FSO using a global atmospheric DA system (Terasaki et al., 2015), which comprises the Nonhydrostatic Icosahedral Atmospheric Model (NICAM: Tomita and Satoh, 2004; Satoh et al., 2008; 2014) and the Local Ensemble Transform Kalman Filter (LETKF: Hunt et al., 2007).

1. EFSO impact estimates' sensitivity to the choice of the verification reference. We estimate observation impacts relative to four references: subsequent analyses, analyses of an independent run of the NICAM-LETKF, the ERA-Interim reanalysis (Dee et al., 2011), and observation departures.
2. The role of relaxation-to-prior methods applied to the initial conditions of ensemble forecasts in the EFSO, a previously undiscussed issue. Relaxation-to-prior methods are commonly used as covariance inflation that inflates the analysis ensemble perturbations after DA. It is unclear whether we need to apply relaxation-to-prior methods to the ensemble forecasts in the EFSO (cf. section 2.4).
3. The role of baseline forecasts that represent forecasts without DA. Gasperoni and Wang (2015, hereafter GW15) conducted deterministic forecasts from the previous analysis time to the verification time. However, GW15's approach may overestimate observation impacts as discussed later. This study proposes an alternative baseline forecast from the ensemble mean of the first guess (cf. section 2.5).

This article is organized as follows. Section 2 describes the NICAM-LETKF system and the EFSO formulation. Section 3 describes the experimental design. The results are presented and discussed in section 4. Finally, a summary is presented in section 5.

2  |  METHODOLOGY

2.1  |  NICAM-LETKF system

As the NWP component, we use the NICAM with the sixth grid division level, corresponding to the horizontal resolution of 112 km, and 38 vertical layers up to about 40 km. The Arakawa and Schubert scheme (Arakawa and Schubert, 1974) and Berry's parametrization (Berry, 1967) were used
for cumulus parametrization and large-scale condensation schemes, respectively.

The LETKF is used as the DA component. The LETKF updates the ensemble mean \( \bar{x} \) and perturbations \( X \equiv \{ x^{(1)} - \bar{x}, x^{(2)} - \bar{x}, \ldots, x^{(m)} - \bar{x} \} \) with a transform matrix, where \( m \), \( x^{(i)} \) and \( \bar{x} \) are the ensemble size, the \( i \)th ensemble state vector and the ensemble mean state vector, respectively. The analysis equations of the LETKF are given by:

\[
\bar{x}^a = \bar{x}^f + X \tilde{P}^a (Y^f)^T R^{-1} d, \tag{1}
\]

\[
X^a = X^f [(m - 1) \tilde{P}^a]^{1/2}, \tag{2}
\]

where \( H \), \( d \), \( Y \), \( R \) and \( \tilde{P}^a \) denote the linear observation operator, observation departure (\( d \equiv y^o - H \bar{x}^f \)), ensemble perturbation matrix in the observation space (\( Y \equiv HY \)), observation error covariance matrix, and analysis error covariance matrix in the ensemble space, respectively. The superscripts \( o \), \( f \) and \( a \) denote the observation, forecast (prior) and analysis (posterior), respectively. Here, the analysis error covariance matrix in the ensemble space \( \tilde{P}^a \) is defined by:

\[
\tilde{P}^a = [(m - 1) I + (Y^f)^T R^{-1} Y^f]^{-1}. \tag{3}
\]

We apply the covariance localization to the observation error covariance so that the errors of distant observations have less impact (Hunt et al., 2007; Miyoshi and Yamane, 2007). We can operate the analysis updates not only at model grid points, but also at observation points:

\[
\bar{y}^f = H \bar{x}^f + Y^f \tilde{P}^a (Y^f)^T R^{-1} d, \tag{4}
\]

\[
Y^a = Y^f [(m - 1) \tilde{P}^a]^{1/2}. \tag{5}
\]

Namely, we can obtain the ensemble mean analysis state \( \bar{y}^f \) and perturbation \( Y^a \) in the observation space within the LETKF. In realistic problems, there would be nonlinear observation operators such as satellite simulators, for which the following approximations are needed:

\[
H \bar{x} ≈ H (\bar{x} \cdot 1 + X), \tag{6}
\]

\[
Y ≈ HY \approx H (\bar{x} \cdot 1 + X^f) - H (\bar{x} \cdot 1 + X^f) \cdot 1, \tag{7}
\]

where \( H \) is the nonlinear observation operator, and \( 1 \) denotes a row vector with all \( m \) elements being equal to 1.

The EnKF needs variance inflation to mitigate the under-dispersive ensemble (Houtekamer and Mitchell, 1998). Two relaxation-to-prior (RTP) methods were implemented with the NICAM-LETKF (Kotsuki et al., 2017c). The first RTP method is the RTP spread (RTPS: Whitaker and Hamill, 2012), which relaxes the reduction of the ensemble spread of the \( i \)th variable of the state vector \( x \) as follows:

\[
X_{RLX}^{α(i)} = \left( \frac{\alpha_{RTPS} \sigma_{RLX}^i + (1 - \alpha_{RTPS}) \sigma_{i}^a}{\sigma_{i}^a} \right) X^a, \tag{8}
\]

\[
σ_{i}^a = \sqrt{\frac{1}{m-1} X^i (X^i)^T}, \tag{9}
\]

where \( α \) and \( α_{RTPS} \) denote the observation spread, and relaxation parameter of the RTPS, respectively. The other RTP method used is the RTP perturbation (RTPP: Zhang et al., 2004), which mixes the prior and posterior ensemble perturbations as follows:

\[
X_{RLX}^{α} = α_{RTPP} X^a + (1 - α_{RTPP}) X^f, \tag{10}
\]

where \( α_{RTPP} \) denotes the relaxation parameter of the RTPP. We investigate the sensitivity of these RTP methods to estimated observation impacts.

2.2 Observation data

Table 1 summarizes the observations assimilated in this study. The NICAM-LETKF assimilates the satellite radiances of the Advanced Microwave Sounding Unit-A (AMSU-A) and Global Satellite Mapping of Precipitation (GSMaP) precipitation in addition to conventional observations (CONV) from the operational system of the NCEP (a.k.a. NCEP PREPBUFR). Due to the relatively low model top height (approximately 40 km), the present NICAM-LETKF assimilates only three channels (6–8) of the AMSU-A radiances and excludes channels that are sensitive to the stratosphere and mesosphere (Terasaki and Miyoshi, 2017). The GSMaP is global surface precipitation data estimated by passive microwave radiometers (Kubota et al., 2007; Ushio et al., 2009). This study assimilates a near-real-time version of the GSMaP data with the Gaussian transformation (GT) method (Lien et al., 2013; 2016a; 2016b; Kotsuki et al., 2017b).

2.3 EFSO

The EFSO estimates the forecast error differences at forecast time \( t \) derived from assimilated observations at analysis time 0. Assume that the NICAM-LETKF computes an analysis every 6 h. The EFSO compares forecasts from two analysis times, −6 and 0 h, to the verification time \( t \left( \bar{x}_{0|t−6}, \text{and} \bar{x}_{0|t} \right) \) with any references close to the truth (Figure 1a). The baseline forecast \( \bar{x}_{0|t−6} \) represents the forecast without DA, and is discussed further in section 2.5.

The forecast errors are defined by:

\[
e_{t|t−6} ≡ \bar{x}_{t|t−6} - \bar{x}_{t|0}^{ref}, \tag{8}
\]

\[
e_{t|0} ≡ \bar{x}_{t|0} - \bar{x}_{t|0}^{ref}, \tag{9}
\]

where \( e \) and \( \bar{x}^{ref} \) are the forecast error and reference state, respectively. Langland and Baker (2004) defined a quadratic
### Table 1: List of observations assimilated in this study

| Type               | Name         | Description                                      | Variables |
|--------------------|--------------|--------------------------------------------------|-----------|
| Conventional (CONV) | ADPUPA       | Radiosondes                                      | $u, v, T, q$ |
| AIRCFT             |              | Aircrafts                                        | $u, v, T$ |
| SATWND             |              | Atmospheric motion vectors from geostationary satellites | $u, v$ |
| PROFLR             |              | Wind profilers                                   | $u, v$ |
| VADWND             |              | Wind velocity azimuth display from the next generation weather radar (NEXRAD) | $u, v$ |
| ADPSFC             |              | Surface synoptic reports                         | $p_s$     |
| SFCSHP             |              | Surface marine reports                           | $u, v, T, q, p_s$ |
| ASCATW             |              | ASCAT scatterometer over the ocean                | $u, v$     |
| GSMaP              |              | GSMaP Satellite-derived surface precipitation (w/ Gaussian transformation) | $GT’ed Precip.$ |
| AMSU-A             |              | Satellite microwave sounder radiances             | $Tb$      |

**Figure 1** (a) Schematic image of the Ensemble Forecast Sensitivity to Observations (EFSo), and (b) workflow of the EFSo in the NICAM-LETKF system. The original NICAM-LETKF (blue-dashed box) was extended with newly developed workflows and algorithms for the EFSo (green-dashed box). Variables with purple shading are inputs for the EFSo.
error norm $\Delta e^2$ as follows:

$$
(\Delta e^2)_{\text{MTE}}^{\text{true}} \equiv e_{i0,\text{MTE}}^2 - e_{i-6,\text{MTE}}^2 \\
= \frac{1}{2}(e_{i0} - e_{i-6})^T C(e_{i0} + e_{i-6}),
$$

where $C$ is the positive definite matrix that defines the forecast error norm, and the superscript 'true' denotes the true error difference. The moist total energy (MTE) norm is given by:

$$
\begin{align*}
\frac{e^2_{\text{MTE}}}{2} &= \frac{1}{2}e^T C e = \frac{1}{2}(KE + PE + ME), \\
\end{align*}
$$

$$
PE = \frac{1}{S} \int_S \left[ \int_0^1 C_p T'' d\sigma_v + \frac{R_d T_t}{\rho^2_l} T'' d\sigma_s \right] dS, \tag{12}
$$

$$
KE = \frac{1}{S} \int_S \int_0^1 (u'^2 + v'^2) d\sigma_v d\sigma_s, \tag{13}
$$

$$
ME = \frac{1}{S} \int_S \int_0^1 \frac{L^2}{C_p T_t} q'^2 d\sigma_v d\sigma_s, \tag{14}
$$

where $S$ represents the target region, $\sigma_v$ is the vertical sigma coordinate, and $T''$, $p_t$, $u'$, $v'$, and $q'$ denote temperature, surface pressure, zonal and meridional winds, temperature and specific humidity of the forecast error $e$, respectively. $R_d$, $C_p$, and $L$ are the gas constant of dry air, specific heat at constant pressure, and latent heat of condensation per unit mass, respectively. $\rho$ and $T_t$ are the reference surface pressure and temperature, set to $10^5$ (Pa) and $280$ (K), respectively. PE, KE and ME denote potential, kinetic, and moist energy, respectively (cf. O13).

K12 extended the adjoint-based FSO for ensemble DA. The impact of the $l$th observation on the forecast at the $j$th model grid point is given by:

$$
(\Delta e^2)_{\text{MTE}}^{\text{true}} \approx (\Delta e^2)_{\text{FSO}}^{\text{MTE}} = \sum_j \sum_l (\Delta e^2)_{\text{MTE},j,l}^{\text{EFSO}}, \tag{15}
$$

$$
(\Delta e^2)_{\text{MTE},j,l}^{\text{EFSO}} = \frac{1}{2m} \rho_{l,j} \left[ (\mathbf{R}_0^{-1} \mathbf{Y}_0^f)^T (\mathbf{C}_0 + \mathbf{d}_{i0} + \mathbf{d}_{i-6}) \right]_l, \tag{16}
$$

where $\rho_{l,j}$ is the localization function at the $j$th model grid point for the $l$th observation. The superscript EFSO denotes the forecast error difference estimated by the EFSO. See appendix A for a detailed derivation.

This study defines an observation-based verification norm as follows:

$$
(\Delta e^2)_{\text{NOD}}^{\text{true}} = \frac{1}{Np} \sum_l \sum_j (\Delta e^2)_{\text{NOD},j,l}^{\text{FSO}},
$$

where $Np$ is the number of observations. Hereafter, this metric is referred to as the normalized observation departure (NOD). As with Equations 15 and 16, the impact of $l$th observation at time 0 on the forecast departure of $n$th observation at time $t$ can be approximated by:

$$
(\Delta e^2)_{\text{NOD}}^{\text{true}} \approx (\Delta e^2)_{\text{NOD}}^{\text{EFSO}} = \frac{1}{Np} \sum_l \sum_j (\Delta e^2)_{\text{NOD},n,j}^{\text{EFSO}}, \tag{18}
$$

$$
(\Delta e^2)_{\text{NOD},n,j}^{\text{EFSO}} = \frac{1}{Np} \sum_l \sum_j \left[ (\mathbf{R}_0^{-1} \mathbf{Y}_0^f)^T (\mathbf{C}_0 + \mathbf{d}_{i0} + \mathbf{d}_{i-6}) \right]_l, \tag{19}
$$

where $\rho_{l,j}$ is the localization function at the $j$th and $n$th observations. Here, the observation error covariance $\mathbf{R}_0$ is assumed to be diagonal for verifying observations. Following SW16, this study employs the observations used in subsequent data assimilation cycles. We use the same localization function $\rho$ for Equations 16 and 19 as defined later in section 3.

### 2.4 EFSO with RTP methods

Here, we discuss the role of RTP methods applied to the initial conditions of ensemble forecasts in the EFSO. In the LETKF, the Kalman gain $\mathbf{K}$ is approximated by

$$
\mathbf{K} \approx \left( \frac{1}{m} \rho \mathbf{X}_0 \mathbf{Y}_0^T \mathbf{R}_0^{-1} \right)
$$

where $\rho$ and $\rho$ are the Schur product and localization matrix (cf. appendix A). The forecast perturbation matrix $\mathbf{X}_0^{f,i}$ is given by $\mathbf{M}\mathbf{X}_0^{i}$, where $\mathbf{M}$ is the tangent linear forecast model. The fully nonlinear model $\mathbf{M}$ is commonly used to approximate $\mathbf{M}$ in the EFSO. Following the original formulation, the Kalman gain $\mathbf{K}$ used in the EFSO should be consistent with the $\mathbf{K}$ in the LETKF (Equation A2). Therefore, we hypothesize that the EFSO needs no relaxation to $\mathbf{X}_0$ in contrast to DA cycles that need the relaxation ($\mathbf{X}_0^{a} \rightarrow \mathbf{X}_0^{a,RLX}$) to avoid under-dispersive ensembles. This study experimentally examines changes in the estimated observation impacts $(\Delta e^2)_{\text{EFSO}}$ with and without the relaxation.

This issue is important in terms of computational cost. If two impact estimates with and without the relaxation are significantly different, it is necessary to perform computationally expensive independent ensemble forecasts only for the EFSO. Care should be taken too when the analysis perturbation is altered such as by relaxation and additive inflation methods. Multiplicative inflation methods for the prior perturbation/covariance in DA cycles are free from this issue. Multiplicative inflation should not be applied to the forecast perturbation in the EFSO (Equation 16).

### 2.5 Interpretation of forecasts in the EFSO

Finally, we bring to light a previously undiscovered issue: interpretation of forecasts $\mathbf{x}_{0,i0}$ and $\mathbf{x}_{0,i-6}$ in the EFSO. Since
K12 did not define these forecasts explicitly, previous studies interpreted these forecasts differently. For example, H17 used the mean of ensemble forecasts ($\overline{x}_t = M_t(x_0)$) and $\overline{x}_{t-6} = M_{t-6}(x_{6})$. Alternatively, GW15 used deterministic forecasts started from ensemble means ($\overline{x}_{t-6} = M_{t-6}(x_{6})$) and $\overline{x}_0 = M_0(x_0)$). Ensemble square root filters including the LETKF update the ensemble-mean first guess; therefore, it would be more straightforward to represent $\overline{x}_0$ and $\overline{x}_{t-6}$ by deterministic forecasts from the ensemble-mean first guess and analysis. The deterministic forecast is also superior to the ensemble forecast in terms of the computational cost.

The difference of the two forecasts of GW15 ($M_{t-6}(x_a)$ and $M_{t}(x_0)$) are caused by different assimilated observations and by the difference between deterministic and ensemble forecasts from the previous ($t-6$) to target ($t$) analysis times ($M_{t-6}(x_a)$ or $M_{t-6}(x_0)$; see also Figure 8). To exclude the second factor, this study performs deterministic forecasts from the ensemble-mean first guess as the baseline forecast to represent the forecast without DA ($\overline{x}_{t-6} = M_{t-6}(M_{t-6}(x_0))$). Another forecast with DA is initialized by the ensemble-mean analysis ($\overline{x}_0 = M_0(x_0)$). These two forecasts $\overline{x}_{t-6} = M_{t-6}(M_{t-6}(x_0))$ and $\overline{x}_0 = M_0(x_0)$ have the same forecasting period from the initial time to the verification time. Section 4.6 discusses sensitivities of EFSO impact estimates to the baseline forecasts.

### 3 EXPERIMENTAL SETTINGS

We first describe the experimental settings of the DA system. The ensemble size is set to be 40. For the covariance localization $p$ (Equations 16 and 19), the Gaussian function is used with standard deviations $LS_h = 400$ km in the horizontal and $LS_v = 0.4$ natural log pressure in the vertical. The localization function is replaced by zero beyond $2\sqrt{10/3} \cdot LS_{h,v}$, which corresponds to the influence-radius of Gaussian function (Gaspari and Cohn, 1999; Hamill et al., 2001). The relaxation parameters $\delta_{\text{RTPS}}$ and $\delta_{\text{RTPP}}$ are set to be 0.95 and 0.80 (Kotsuki et al., 2017c). The LETKF assimilates temporally distributed observations over a 6 h time window centred on the analysis time (a.k.a. 4D-LETKF: Fertig et al., 2007; Miyoshi et al., 2010). Ensemble forecasts are performed for 9 h, and observation data in the previous 6 h period are assimilated (cf. fig. 3a of Terasaki and Miyoshi, 2017). Temporal localization (Miyoshi and Yamane, 2007) is not applied in this study, enabling us to assume the same transform matrix during the 6 h time window. Namely, this assumption enables the LETKF to compute analysis perturbations in the observation space ($Y_a$) distributed over the 6 h time window. Note that we conducted preliminary experiments with and without the temporal localization, and concluded that analysis accuracy was insensitive to the temporal localization (not shown).

Next, we detail the experimental settings of the EFSO. This study sets 6 and 12 h for the forecast verification times (FTs). H17 compared the EFSO with 6 and 24 h forecasts and pointed out that the EFSO was insensitive to these lead times. The localization scale in the EFSO is the same as that in the LETKF. The advection of the localization centre (O13) was not important in our preliminary experiments (cf. Table B1 of appendix B); therefore, the advection scheme is not used hereafter. We use ADPUPA, SATWND, and AMSU-A radiances for the observation-based verification metric. The NOD metric (Equations 17–19) is sensitive to the normalizing matrix $R$ and the number of verifying future observations ($N_p$). Therefore, we perform the EFSOs separately for ADPUPA, SATWND, or AMSU-A radiances as the verification references. For the AMSU-A-based verification, we apply the same bias correction coefficients used in the DA cycle (i.e. for 6 h forecasts) to 12 h forecasts. After applying the bias correction, mean departures (K) for 6 and 12 h forecasts are $-0.003$ and $0.010$ for channel 6, $-0.030$ and $-0.029$ for channel 7, and $-0.084$ and $-0.085$ for channel 8, respectively. This suggests that the bias correction coefficients for 6 h forecasts can be applicable to 12 h forecasts in the NICAM-LETKF.

Table 2 summarizes the experimental settings. The control experiment (CTRL) assimilates CONV, GSMaP and

| Experiment | Observations | Covariance inflation ($X^a$) | Period          |
|------------|--------------|-----------------------------|-----------------|
| CTRL       | X            | X                           | RTPS            | June–July 2014 |
| TWIN-EXP*  | X            | X                           | RTPS            | -              |
| DNY-EXP    | X            | X                           | RTPS            | July 2014      |
| RLX-EXP1   | X            | X                           | RTPS            | RTPS           | July 2014      |
| RLX-EXP2   | X            | X                           | RTPS            | -              |
| RLX-EXP3   | X            | X                           | RTPP            | RTPP           |

*Réal-EXP uses an independent initial ensemble from CTRL, whereas the experimental setting of TWIN-EXP is the same as CTRL.*
AMSU-A observations. We also conduct a twin experiment (TWIN-EXP) with the same experimental setting as the CTRL, but from different initial ensemble data. The initial ensembles for CTRL and TWIN-EXP were obtained from the 1st–40th and 41st–80th members of a long-term 128-member NICAM-LETKF experiment (Terasaki et al., 2019). The TWIN-EXP is used as an independent verification reference. A data denial experiment is conducted with excluding AMSU-A radiances (DNY-EXP) in DA. Experiments with different RTP methods are compared: RTPS for CTRL and RLX-EXP1, and RTPPP for RLX-EXP2 and RLX-EXP3. RLX-EXP1 and RLX-EXP3 apply the RTP methods to the initial conditions of ensemble forecasts in the EFSO, whereas CTRL and RLX-EXP2 do not as described in section 2.4 (cf. Figure 1b).

The data assimilation experiments of CTRL and TWIN-EXP are performed for 2 months from 0000 UTC 1 June to 1800 UTC 31 July 2014. DNY-EXP and RLX-EXP1–3 are started from 0000 UTC 1 July succeeding after a 1-month DA cycle of CTRL. The results obtained over the last 4 weeks (4–31 July) are used for discussion.

4 | RESULTS AND DISCUSSION

4.1 | Estimated impacts verified with atmospheric states

Figure 2 shows the estimates of the average 6 h forecast error differences (Δ\(e_{\text{MTE}}^2\)) in CTRL verified with subsequent analyses of CTRL and ERA-Interim reanalysis. We show the results for KE, PE and ME separately to physically understand how each type of observation improves the atmospheric fields. For example, AMSU-A mainly reduces KE, suggesting that AMSU-A radiances improve wind fields despite that AMSU-A is sensitive to atmospheric temperature profiles. The GSMaP mainly improves moisture fields. Among the three energy norms of KE, PE and ME, the largest improvement is observed for KE.

In terms of the overall impact, the ADPUPA makes the largest contribution to the forecast error reduction, followed by SATWND and then AMSU-A or ASCATW. Estimated impacts of the AIRCFT and AMSU-A are smaller than those by O13 and H17. The impacts of AMSU-A are smaller probably because the present NICAM-LETKF assimilates only three channels of AMSU-A (Terasaki and Miyoshi, 2017), whereas operational centres assimilate 10 channels. The impacts of AIRCFT would also be smaller because public PREPBUFR contains a limited number of aircraft observations (Dr. Akira Yamazaki, personal communications). The estimated MTE forecast error differences (Δ\(e_{\text{MTE}}^2\)) have good agreements with the true differences (Δ\(e_{\text{MTE}}^2\)) in general (Table B1 in appendix B), suggesting accurate impact estimates by the EFSO.

Between the two verification references (subsequent analyses and ERA-Interim reanalysis), relatively large differences in estimated impacts are found for ADPUPA, SATWND and ASCATW, where verification with their own analysis shows generally smaller amplitudes (Figure 2c). This may be because the quadratic error norm was used to measure the forecast errors. NICAM forecasts would be closer to subsequent analyses than to the ERA-Interim reanalysis; therefore, the total error reduction relative to the ERA-Interim reanalysis was larger than that relative to subsequent analyses. On the other hand, forecast error reductions in ME are smaller when we verify with the ERA-Interim reanalysis (Figure 2b). This may be due to a dry bias of NICAM-LETKF in the lower troposphere relative to ERA-Interim (Kotsuki et al., 2018). Due to the bias in humidity, improved humidity fields relative to subsequent analyses do not guarantee improvements relative to the ERA-Interim reanalysis.

4.2 | Estimated impacts verified with observations

Figure 3 shows the estimates of the average 12 h forecast error differences (Δ\(e_{\text{NOD}}^2\)) in CTRL, verified with ADPUPA, SATWND, and AMSU-A radiances. The EFSO diagnoses that each type of observation generally improves the departures of the observed variable. For example, assimilation of the AMSU-A radiances significantly reduces the forecast departure of the AMSU-A radiances, but not necessarily for other observations (Figure 3c). This feature would come from a limitation of EFSO formulation rather than reflecting the true observation impacts.

When we assimilate an observation, its impact can propagate away. However, the EFSO limits the estimated impact to the observation’s immediate vicinity because of localization (Equation 19). For example, assume that we evaluate the impact of an assimilated observation A using an observation B as the verification reference. If the localization function between observations A and B is zero (\(\rho = 0\) in Equation 19), the EFSO automatically diagnoses that the observation A has no impact onto the observation departure for B because of the localization. This artefactual mechanism would be particularly relevant to the AMSU-A radiances that observe temperature profiles in the upper atmosphere. In fact, observations near the surface (ADPSFC, SFCSHP and ASCATW) have negligible impact onto the EFSO verified with AMSU-A radiances (Figure 3c).

It is important to evaluate if assimilating observations improves or degrades the forecast departure of other observed quantities. The EFSO with observation-based metrics is a useful tool for this purpose although its interpretation has to be done carefully acknowledging the limitations described above. In the NICAM-LETKF system, no observation degraded departures of other observed quantities,
excepting the slight detrimental impacts of PROFLR onto AMSU-A radiances (Figure 3c).

4.3 Impact of AMSU-A radiances

We investigate the impacts on the upper troposphere and stratosphere from AMSU-A radiances which have a large coverage across the entire globe. Figure 4 shows the spatial patterns of the estimated 6 h forecast MTE differences (J/kg) above 350 hPa relative to the ERA-Interim reanalysis in CTRL. Figure 4a,b show the estimated forecast MTE differences at each model grid point from AMSU-A and the other observations (\( \sum_{l \in \text{AMSU-A}} (\Delta e_{MTE}^{\text{EFSO,}}_{j,l} + \Delta e_{MTE}^{\text{EFSO,}}_{j,l}) \)) and (\( \sum_{l \in \text{OtherObs.}} (\Delta e_{MTE}^{\text{EFSO,}}_{j,l}) \)). Estimated impacts of AMSU-A radiances over land are smaller than those over the ocean (Figure 4a) maybe because other observations (e.g. radiosondes) have dominant impacts over land. In general, assimilating AMSU-A radiances has beneficial impacts across the globe, especially in the Southern Hemisphere (SH). However, assimilating AMSU-A radiances would degrade forecasts in the Tropics (TR; Figure 4a). The detrimental impacts in the TR may be related to a temperature bias between the NICAM-LETKF analysis and the ERA-Interim reanalysis. DA systems perform bias corrections for satellite radiances so that the time-mean observation-minus-first-guess innovation becomes statistically zero (e.g. Bormann and Bauer, 2010). The NICAM-LETKF analysis has a warmer temperature bias in the stratosphere over the TR relative to the ERA-Interim reanalysis (Kotsuki et al., 2018). Assimilating the bias-corrected AMSU-A radiances would degrade the analyses and subsequent forecasts relative to the ERA-Interim reanalysis.

We compare the observation impacts estimated by the EFSO with the data denial experiment for AMSU-A radiances (DNY-EXP). The EFSO impact estimates are not exactly equivalent to the impact estimates with the data denial experiments. A data denial experiment estimates observation data impact by comparing two DA cycles with and without the data. This provides a direct estimate of the data impact, while the EFSO provides indirect impact estimates. The EFSO is...
Figure 3 shows the estimated 12 h forecast error differences ($\Delta e_2^{\text{EFSO}}$) verified with (a) ADPUPA, (b) SATWND and (c) AMSU-A radiances, showing the contribution from each type of observation over 4 weeks from 4 to 31 July 2014. Numbers of panels show the estimated impacts for (a) ADPUPA, (b) SATWND and (c) AMSU-A radiances, respectively. EFSO can be used to detect detrimental observations in general. However, the opposite results are found in some areas such as the North Atlantic and northwest Pacific Ocean.

### 4.4 Fraction of beneficial observations

Here we investigate the fraction of beneficial observations (FBO), or the percentage of positively contributing observations, in CTRL verifying with different references. We define the FBO by $N_{\text{beneficial}} / (N_{\text{beneficial}} + N_{\text{detrimental}})$ where $N_{\text{beneficial}}$ and $N_{\text{detrimental}}$ are the numbers of assimilated beneficial and detrimental observations. Observations having no impact, that is, $\sum_j (\Delta e_2^{\text{EFSO}})_{j,l} = 0$, are excluded. The estimated MTE impacts of individual observations assimilated at 0000 UTC on 11 July 2014 are shown in Figure 6 ($\sum_j (\Delta e_2^{\text{MTE}})_{j,l}^{\text{EFSO}}$).

When we verify relative to the ERA-Interim reanalysis (Figure 6a,b), both beneficial and detrimental observations are seen over the globe. However, we mainly see beneficial observations when we verify relative to the subsequent analyses, especially for the shorter 6 h forecast (Figure 6c,d). Verifying relative to subsequent analyses would overestimate the FBO. However, this is not the case when we verify relative to the analyses of TWIN-EXP (Figure 6e,f).

Table 3 summarizes the time-mean FBO. The FBO is 51.9–55.4% when we verify relative to the analyses of TWIN-EXP, ERA-Interim reanalysis, and AMSU-A radiances. Previous studies also found that the FBO was just over 50% (Lorenc and Marriott, 2014; SW16; N18). With a toy chaotic model (Lorenz, 1996; Lorenz and Emanuel, 1998), a similar FBO was also reported (Kotsuki et al., 2017a). However, the FBO is remarkably higher when subsequent analyses are used as the reference (56.1–58.9%). Interestingly, such a higher FBO is not observed when we verify relative to the analyses of TWIN-EXP (54.0–54.4%).

The EFSO regards the verification reference as truth. However, the errors in subsequent analyses may be positively correlated with the propagated analysis increments for short-term forecasts. When they are positively correlated, assimilated observations may tend to be evaluated as beneficial. K12 and H17 reported that the FBO increased as the lead time shortened. This would be more significant when we use the subsequent analyses as the reference. The propagated analysis increments may have larger correlation relative to the errors in subsequent analysis than to the errors in TWIN. Unfortunately, it is almost impossible to investigate such correlations since we never know the true state in practice. Therefore, it is difficult to investigate whether the FBO is overestimated in the presence of positive error correlation between the propagated analysis increments and the errors in the subsequent analyses. We need to investigate this issue further with observing system simulation experiments where we know the true state, but these additional experiments are beyond the scope of this study. This study experimentally...
90°N (a) MTE vs ERA; AMSU–A MTE vs ERA; All Obs. – AMSU–A

FCST MTE Reduction Estimated by EFSO (J/kg); Stratosphere (~350hPa)

**FIGURE 4** Spatial patterns of the estimated 6 h forecast MTE reduction (J/kg) above 350 hPa relative to the ERA-Interim reanalysis in CTRL, averaged over 4 weeks from 4 to 31 July 2014. (a,b) Estimated impact from AMSU-A radiances ($\sum_{\ell \in \{\text{AMSU-A}\}} (\Delta e_{2,\text{MTE}}^{\text{EFSO},j})$) and the other observations ($\sum_{\ell \in \{\text{OtherObs.}\}} (\Delta e_{2,\text{MTE}}^{\text{EFSO},j})$), respectively.

|  |  |
|---|---|
| << Improved >> | From 2014070400 To 2014073118 |
| Degraded >> | |

**FIGURE 5** Global pattern of difference of time-mean 6 h forecast MTE (J/kg) above 350 hPa relative to the ERA-Interim reanalysis between CTRL and DNYEXP (i.e. with and without AMSU-A radiances; $e_{\ell,\text{MTE}}^{\text{CTRL}}$ minus $e_{\ell,\text{MTE}}^{\text{DNY-EXP}}$), averaged over 4 weeks from 4 to 31 July 2014. Green and brown colours represent improvement and degradation due to assimilation of AMSU-A radiances, respectively.

revealed that overestimated FBO can be mitigated by verifying relative to future observations, analyses of different DA systems, or analyses of an independent DA cycle with the same system.

There are significant reductions in the FBO between FT 6 and 12 h when we verify relative to the subsequent analyses. Small reductions are also seen even when we verify relative to the analyses of TWIN-EXP or ERA-Interim reanalysis, perhaps due to the assimilation of similar observations. On the other hand, no such reduction is observed when we verify relative to the AMSU-A radiances. Observation-based metrics may be more appropriate to estimate the FBO for a shorter lead time since it can avoid an overestimation of the FBO.

4.5 | Estimated impacts with and without RTP methods

Here, we discuss the third issue, that is, whether we need to apply the RTP methods to the initial condition of the ensemble forecasts in the EFSO. Here, the ERA-Interim reanalysis is used as the verification reference.

Figure 7 compares the 6 and 12 h forecast MTE error differences: real reduction ($\Delta e_{2,\text{MTE}}^{\text{true}}$) and estimated reduction ($\Delta e_{2,\text{MTE}}^{\text{EFSO}}$). First, we discuss the RTPS (Figure 7a,b). The EFSO impact estimates in CTRL (without RTPS) show good agreements with truth at FT = 6 h, while the EFSO slightly underestimate forecast error reductions at FT = 12 h. On the other hand, the EFSO in RLX-EXP1 (with RTPS) overestimates forecast error reductions at both FTs. With the RTPP, the EFSO in RLX-EXP3 significantly overestimates forecast error reductions (Figure 7c,d).

As discussed in section 2.4, we hypothesized that the EFSO would need no relaxation to $X^0_a$ based on the original formulation. As hypothesized, observation impacts can be overestimated if we apply the RTP methods to the initial condition $X^0_a$ in the EFSO. The overestimated impacts with RTPP methods are more significant than those with RTPS.

Error growth characteristics of RTPP and RTPS may cause different degrees of overestimation of the observation impacts. The RTPS provides a less physically balanced analysis than the RTPP. Therefore, the ensemble spread of the RTPS analysis grows slower than that of the RTPP (Whitaker and Hamill, 2012; Kotsuki et al., 2017c). Owing to the slower
FIGURE 6 Estimated impact (Δ$e_{\text{MTE}}^2$) of all observations assimilated in CTRL at 0000 UTC on 11 July 2014, verified with (a,b) ERA-Interim reanalysis, (c,d) subsequent analyses of CTRL, and (e,f) analyses of TWIN, respectively. Red and blue dots denote beneficial and detrimental impacts, respectively. For each observation, the square of circle has a linear relation with the absolute impact. (a,c,e) and (b,d,f) show the estimated impact verified at FT 6 and 12 h, respectively. The black circles at the left bottom of the panels correspond to a 2.0 $\times 10^{-3}$ error difference in MTE (J/kg).

TABLE 3 The time-mean fraction of beneficial observations (%) in CTRL averaged from 4 to 31 July 2014

| Verification reference | FT (h) | CTRL | TWIN-EXP | ERA-Interim | AMSU-A |
|------------------------|--------|------|----------|-------------|--------|
| Fraction of beneficial observations (%) | 06 | 58.9 | 54.4 | 55.4 | 51.9 |
|                        | 12 | 56.1 | 54.0 | 54.3 | 52.2 |

error growth, observation impacts with the RTPS were overestimated less significantly than those with the RTPP.

4.6 Impact of the choice of baseline forecast

Finally, we investigate the sensitivity to the choice of the baseline forecast, as discussed in section 2.5. Here, we consider three deterministic forecasts at the verification time (Figure 8) initiated by the analysis at the previous analysis time ($M_{t|0}(x_0) = M_{t|0}(M_{t|6}(x^f_{t-6}))$). This study uses $M_{t|0}(x_0) - M_{t|0}(x^f_{t-6})$ for $x^f_{t|0} - x^f_{t|6}$ in EFSO, whereas GW15 used $M_{t|0}(x_0) - M_{t|6}(x^f_{t-6})$.

Figure 9a–c compares the spatial patterns of the 6 h forecast MTE error differences (Δ$e_{\text{MTE}}^2$) for $M_{t|0}(x_0) - M_{t|0}(x^f_{t-6})$ and $M_{t|0}(x_0) - M_{t|6}(x^f_{t-6})$ verified with the ERA-Interim reanalysis. The MTE error differences with $M_{t|0}(x_0) - M_{t|0}(x^f_{t-6})$ and $M_{t|0}(x_0) - M_{t|6}(x^f_{t-6})$ show similar spatial patterns (Figure 9a,b). However, the MTE error reduction with $M_{t|0}(x_0) - M_{t|6}(x^f_{t-6})$ is generally increased compared to that with $M_{t|0}(x_0) - M_{t|0}(x^f_{t-6})$.
**Figure 7** Forecast error differences in the MTE (J/kg) verified with ERA-Interim reanalysis over 4 weeks from 4 to 31 July 2014. Horizontal and vertical axes are true and estimated differences, that is, \((\Delta e_{MTE}^{\text{true}})\) and \((\Delta e_{MTE}^{\text{EOF}})\), respectively. (a,b) and (c,d) compare CTRL and RLX-EXP1 (black and blue dots), and RLX-EXP2 and RLX-EXP3 (black and blue dots), respectively. (a,c) and (b,d) show comparison at FT 6 and 12 h, respectively.

**Figure 8** Three forecasts used to investigate the sensitivity of forecast error differences to baseline deterministic forecasts. In addition, \(M_{t|0}(\overline{x}_0)\) is better than \(M_{t|0-6}(\overline{x}_{-6})\) significantly (Figure 9c), implying that the mean of the ensemble forecasts \((\overline{M}_{0|0-6}(\overline{x}_{-6}))\) is statistically better than the deterministic forecast \((M_{0|0-6}(\overline{x}_{-6}))\).

Figure 10 compares the spatial patterns of the root mean square difference (RMSD) of zonal wind \((U; \text{m/s})\) for \(\overline{M}_{0|0-6}(\overline{x}_{-6})\) and \(M_{0|0-6}(\overline{x}_{-6})\) at the target analysis time relative to the ERA-Interim reanalysis. The mean of the ensemble forecasts \((\overline{M}_{0|0-6}(\overline{x}_{-6}))\) is significantly better than the deterministic forecast \((M_{0|0-6}(\overline{x}_{-6}))\) in similar regions where \(M_{t|0}(\overline{x}_0)\) is better than \(M_{t|0-6}(\overline{x}_{-6})\) (Figures 9d and 10c).
Relatively large differences are observed in the TR where convective systems are dominant. We may need to consider this baseline deterministic forecast issue carefully, especially for the EFSO for convective events.

The above discussion suggests that the actual forecast MTE error differences \( (\Delta e_{\text{MTE}})^{\text{true}} \) are overestimated if we use \( M_{0|0}(\vec{X}) \) to represent the forecast without DA. Therefore, our deterministic forecasts would be more appropriate than GW15’s approach to representing forecasts with and without DA. It does not necessarily mean that the estimated forecast MTE error difference \( (\Delta e_{\text{MTE}})^{\text{EFSO}} \) is always overestimated because other uncertainties such as the approximation of \( (\Delta e_{\text{MTE}})^{\text{true}} \) by \( (\Delta e_{\text{MTE}})^{\text{EFSO}} \) can be sources of an underestimation.

Schraff et al. (2016) proposed an independent update of a deterministic run, which introduces analysis increments into the deterministic forecasts based on the Kalman gain for the analysis ensemble mean. The independently updated deterministic runs can be compared with \( M_{0|0}(\vec{X}) \).

### 4.7 Implications for other DA systems

The present NICAM-LETKF system assimilates fewer observations than typical operational systems. Here we
discuss what are specific to the NICAM-LETKF, and what are general properties applicable to other implementations.

Discussions on the covariance relaxation and baseline forecasts (sections 4.5 and 4.6) would be insightful to other ensemble-based systems since the experimental results generally agree with theoretical implications. Discussion on the FBO (section 4.4) would be generic properties for both ensemble- and adjoint-based FSOs. With respect to the MTE-based impact estimates (section 4.1), estimated impacts of each observation type would be different from other operational systems. Compared to operational systems (e.g. O13; LM14; H17; C18; Buehner et al., 2018), AIRCFT and AMSU-A radiances generally show smaller contributions in the NICAM-LETKF maybe because fewer observations are assimilated (Figure 2d).

As for the observation-based impact estimates (section 4.2), the limitation of EFSO should be discussed. Each type of observation mainly improves departures of the observed variable in NICAM-LETKF (Figure 3a–c). For example, AMSU-A ($-214.8 \times 10^{-3}$) occupies about 96.1% of the total forecast error reduction for the AMSU-A departures ($-223.6 \times 10^{-3}$). Other observations would also reduce AMSU-A departures in operational systems in which many more observations sensitive to the upper troposphere and stratosphere (e.g. more AIRCFT, all-sky microwave imagers, Infrared Atmospheric Sounding Interferometer, and Advanced Technology Microwave Sounder) are assimilated. In addition, adjoint-based FSO does not apply localization and would better capture propagated impacts. In contrast, localization of the EFSO limits impacts of observations to their immediate vicinity. Improving the ensemble size or localization method would help estimate propagated observation impacts better.

5 | SUMMARY

The EFSO is a powerful and computationally feasible tool to estimate observation impacts. This study compared EFSO impact estimates using multiple references. We implemented the EFSO using two error norms (MTE and NOD) with the NICAM-LETKF. The main findings are summarized as follows:

1. We visualized the dry and moist energy errors with separated KE, PE and ME errors. This separation may help understand specific contributions from each type of observation.
2. With observation-based impact estimates, each type of observation mainly contributes to the improvement of forecast departures of the observed variables. For example, the assimilation of the AMSU-A radiances significantly reduces the forecast departure of the AMSU-A radiances, but only slightly reduces the forecast departure of other observations. This feature would come from a limitation of the localization in the EFSO. Improving the ensemble size or the localization method would help estimate propagated observation impacts better. However, the EFSO with observation-based metrics can be a useful tool to evaluate if assimilating observations improves or degrades the forecast departure of other observed quantities.
3. The EFSO overestimates the fraction of beneficial observations when we verify relative to subsequent analyses, especially for shorter lead times such as FT = 6 h. We could mitigate this overestimation to some extent by verifying with observations, analyses from other DA systems, or analyses of an independent run with the same DA system.
4. Observation impacts are overestimated if we apply the RTPS or RTTP to the initial conditions of the ensemble.
forecasts in the EFSO. Therefore, the ensemble forecasts in the EFSO should be independent of the ensemble forecasts in the DA cycle, especially for RTPP.

5. Deterministic baseline forecasts should be initialized by the ensemble mean of the first guess at the analysis time, not by the previous analysis. The fraction of beneficial observations was just over 50% in the NICAM-LETKF. It would be worth investigating designing additional QC methods, such as PQC (H17), for further forecast improvements.

ACKNOWLEDGEMENTS
The authors thank the members of Data Assimilation Research Team, RIKEN Center for Computational Science (R-CCS), Y. Ota of the Japan Meteorological Agency (JMA), D. Hotta of the Meteorological Research Institute, JMA, and A. Yamazaki of the Japan Agency for Marine-Earth Science and Technology for valuable suggestions and discussions. This study was partly supported by the Japan Aerospace Exploration Agency (JAXA) Precipitation Measuring Mission (PMM), advancement of meteorological and global environmental predictions utilizing observational “Big Data” of the social and scientific priority issues (Theme 4) to be tackled by using post K computer of the FLAGSHIP2020 Project of the Ministry of Education, Culture, Sports, Science and Technology Japan (MEXT), the Initiative for Excellent Young Researchers of MEXT, and the Japan Society for the Promotion of Science (JSPS) KAKENHI grants JP15K18128 and JP18H01549. The experiments were performed using the Supercomputer for earth Observation, Rockets, and Aeronautics (SORA) at JAXA, and the K computer provided by the R-CCS (Project IDs: ra000015, ra001011, hp150289, hp160162, hp160229, and the K computer provided by the R-CCS (Project IDs: ra000015, ra001011, hp150289, hp160162, hp160229, hp170178, hp170246, hp170305, hp180194 and hp180062).

ORCID
Shunji Kotsuki https://orcid.org/0000-0002-0717-9192
Kenta Kurosawa https://orcid.org/0000-0002-2443-8758
Takemasa Miyoshi https://orcid.org/0000-0003-3160-2525

REFERENCES
Arakawa, A. and Schubert, W.H. (1974) Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. Journal of the Atmospheric Sciences, 31, 674–701. https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.
Berry, E. (1967) Cloud droplet growth by collection. Journal of the Atmospheric Sciences, 24, 688–701. https://doi.org/10.1175/1520-0469(1967)024<0688:CDBGBC>2.0.CO;2.
Bormann, N. and Bauer, P. (2010) Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. Quarterly Journal of the Royal Meteorological Society, 136, 1036–1050. https://doi.org/10.1002/qj.616.
Buehner, M., Du, P. and Bédard, J. (2018) A new approach for estimating the observation impact in ensemble–variational data assimilation. Monthly Weather Review, 146, 447–465. https://doi.org/10.1175/MWR-D-17-0252.1.
Cardinali, C. (2009) Monitoring the observation impact on the short-range forecast. Quarterly Journal of the Royal Meteorological Society, 135, 239–250. https://doi.org/10.1002/qj.366.
Cardinali, C. (2018) Forecast sensitivity observation impact with an observation-only based objective function. Quarterly Journal of the Royal Meteorological Society, 144(716), 2089–2098. https://doi.org/10.1002/qj.3305.
Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hölm, E.V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137, 553–597. https://doi.org/10.1002/qj.828.
Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143–10162. https://doi.org/10.1029/94JC00572.
Fertig, E.J., Harlim, J. and Hunt, B.R. (2007) A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: perfect model simulations with Lorenz-96. Tellus A, 59, 96–100. https://doi.org/10.1111/j.1600-0870.2006.00205.x.
Gaspari, G. and Cohn, S. (1999) Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society, 125, 723–757. https://doi.org/10.1002/qj.49712555417.
Gasperoni, N.A. and Wang, X. (2015) Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. Monthly Weather Review, 143, 1981–2000. https://doi.org/10.1175/MWR-D-14-00272.1.
Geer, A.J. (2016) Significance of changes in medium-range forecast scores. Tellus A, 68, 30229. https://doi.org/10.3402/tellusa.v68.30229.
Hamill, T.M., Whitaker, J.S. and Snyder, C. (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review, 129, 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.
Hotta, D., Chen, T.-C., Kalnay, E., Ota, Y. and Miyoshi, T. (2017) Proactive QC: a fully flow-dependent quality control scheme based on EFSO. Monthly Weather Review, 145, 3331–3354. https://doi.org/10.1175/MWR-D-16-0290.1.
Houtekamer, P.L. and Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126, 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUEAK>2.0.CO;2.
Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform
Kalman filter. *Physica D*, 230, 112–126. https://doi.org/10.1016/j.physd.2006.11.008.

Kalnay, E., Ota, Y., Miyoshi, T. and Liu, J. (2012) A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, 64, 18462. https://doi.org/10.3402/tellusa.v64i0.18462.

Kelly, G., Buizza, R. and Cardinali, C. (2007) The value of observations. *Quarterly Journal of the Royal Meteorological Society*, 133, 1803–1815. https://doi.org/10.1002/qj.150.

Kotsuki, S., Greybush, S.J. and Miyoshi, T. (2017a) Can we optimize the assimilation order in the serial ensemble Kalman filter? A study with the Lorenz-96 model. *Monthly Weather Review*, 145, 4977–4995. doi: 10.1175/MWR-D-17-0094.1.

Kotsuki, S., Ota, Y. and Miyoshi, T. (2017c) Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 143(705), 2001–2015. https://doi.org/10.1002/qj.3060.

Kotsuki, S., Terasaki, K., Lien, G.-Y. and Kalnay, E. (2017b) Assimilating the global satellite mapping of precipitation data with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM). *Journal of Geophysical Research: Atmospheres*, 122, 631–650. https://doi.org/10.1002/2016JD025355.

Kotsuki, S., Ota, Y. and Miyoshi, T. (2017e) Adaptive covariance relaxation methods for ensemble data assimilation: experiments in the real atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 143(705), 2001–2015. https://doi.org/10.1002/qj.3060.

Lien, G.-Y., Kalnay, E. and Miyoshi, T. (2016a) Stochastic properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation. *Monthly Weather Review*, 144, 663–679. https://doi.org/10.1175/MWR-D-15-0150.1.

Lien, G.-Y., Miyoshi, T. and Kalnay, E. (2016b) Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecasting system. *Monthly Weather Review*, 144, 643–661. https://doi.org/10.1175/MWR-D-15-0149.1.

Liu, J. and Kalnay, E. (2008) Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 134, 1327–1335. https://doi.org/10.1002/qj.280.

Lorenc, A.C. and Marriott, R.T. (2014) Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 140, 209–224. https://doi.org/10.1002/qj.2122.

Lorenz, E. (1996) Predictability – a problem partly solved. In: *Proc. Seminar on Predictability*. Reading, UK: ECMWF, pp. 1–18.

Lorenz, E. and Emanuel, K.A. (1998) Optimal sites for supplementary weather observations: simulation with a small model. *Journal of the Atmospheric Sciences*, 55, 399–414. https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Miyoshi, T., Sato, Y. and Kadowaki, T. (2010) Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Monthly Weather Review*, 138, 2846–2866. https://doi.org/10.1175/2010MWR3209.1.

Miyoshi, T. and Yamane, S. (2007) Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Monthly Weather Review*, 135, 3841–3861. https://doi.org/10.1175/2007MWR1873.1.

Necker, T., Weissmann, M. and Sommer, M. (2018) The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society*, 144(714), 1667–1680. https://doi.org/10.1002/qj.3390.

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

Sato, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., Noda, A.T., Yamada, Y., Goto, D., Sawada, M. and Miyoshi, T. (2014) The non-hydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, 227, 3486–3514. https://doi.org/10.1016/j.jcp.2007.02.006.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periañez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142, 1453–1472. https://doi.org/10.1002/qj.2748.

Sommer, M. and Weissmann, M. (2016) Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus A*, 68, 27885. https://doi.org/10.3402/tellusa.v68i8.27885.

Terasaki, K., Kotsuki, S. and Miyoshi, T. (2019) Multi-year analysis using the NICAM-LETKF data assimilation system. *SOLA*, 15, 41–46. https://doi.org/10.1175/SOLA-2019-009.

Terasaki, K. and Miyoshi, T. (2017) Assimilating AMSU-A radiances with the NICAM-LETKF. *Journal of the Meteorological Society of Japan*, 95, 433–446. https://doi.org/10.2151/jmsj.2017-028.

Terasaki, K., Sawada, M. and Miyoshi, T. (2015) Local ensemble transform Kalman filter experiments with the nonhydrostatic icosahedral
expanded as follows:

\[
\begin{align*}
M_{d0}Kd_0 &= M_{d0}P_0^aH^TR_0^{-1}d_0 \\
&\approx M_{d0}\left(\frac{1}{m-1}\rho^{xz\times z}\sigma X_0^a X_0^{aT}\right)HTR_0^{-1}d_0 \\
&\approx \frac{1}{m-1}M_{d0}(\rho^{zx\times q}\sigma X_0^f Y_0^{fT})R_0^{-1}d_0 \\
&\approx \frac{1}{m-1}(\rho^{zx\times q}\sigma X_0^f Y_0^{fT})R_0^{-1}d_0.
\end{align*}
\] (A2)

where the symbol “\(\sigma\)” represents the element-wise multiplication (Schur product). \(\rho\) are the localization matrices whose sizes are \(z\times z\) and \(z\times q\) where \(z\) and \(q\) are the numbers of model state variables and observations. The localization function \(\rho^{zx\times q}\) (4th line of Equation A2) can be different from \(\rho^{zx\times q}\) (3rd line of Equation A2) if we consider the advected localization function due to the forecast (e.g. Kalnay et al., 2012; O13). Practically, the fully nonlinear model \(M\) is commonly used to obtain the forecast perturbation matrix \(X_{f0}^f\) so that the \(i\)th column of \(X_{f0}^f\) is \([Mx_0^{(o)} - Mx_0^{(d)}]\) (K12). The observation-based verification metric (Equation 19) can be introduced in the same manner with an additional approximation:

\[
(d_{f0} - d_{t-6}) = H\tilde{x}_{t-6} - H\tilde{x}_0 \\
\approx H(\tilde{x}_{t-6} - \tilde{x}_0) \\
= -H(e_{t0} - e_{t-6}).
\] (A3)

APPENDIX B: VERIFICATION OF THE EFSO

This appendix validates the forecast error differences estimated by the EFSO. Table B1 compares the correlations between estimated and true MTE forecast error differences, that is, \((\Delta e_{\text{MTE}})_{\text{EFSO}}^2\) and \((\Delta e_{\text{MTE}})^2_{\text{true}}\). The MTE forecast error differences at 0000 and 1200 UTC tend to be larger than those at 0600 and 1800 UTC mainly because most radiosondes are launched only at 0000 and 1200 UTC. Computing correlation coefficients using forecast error differences at different verification times altogether would lead to overestimated correlations. To avoid this issue, here we compute correlations separately for assimilated observations at 0000 and 1200 UTC and for those at 0600 and 1800 UTC. The EFSO impact estimates show good agreements with the true error differences in general (correlation coefficient \(>0.817\)). Table B1 also suggests that the advection of the localization centre (O13) does not have a significant effect on the correlation coefficients for short lead time as noted by H17.

Table B2 compares the estimated and true NOD forecast error differences, that is, \((\Delta e_{\text{NOD}})_{\text{EFSO}}^2\) and \((\Delta e_{\text{NOD}})^2_{\text{true}}\). As described in section 3, we perform the EFSOs separately with ADPUPA, SATWND, or AMSU-A radiances
used as the verification references. For ADPUPA-based verification, we use ADPUPA observations only at 0000 and 1200 UTC. Compared to the MTE-based verification norm, correlations between the true and estimated impacts are relatively weak with the observation-based verification norms probably because we used inhomogeneous observations as the verification reference. Once an observation is assimilated, the impact propagates away spatially. The MTE-based verification norm captures the propagated impact with homogeneous model grids. In contrast, the observation-based verification norm would not detect a part of propagated impacts due to the spatially inhomogeneous distribution of observations (cf. discussion in section 4.2). In addition, the observation-based verification norm contains an additional approximation (Equation A3; $\tilde{H}_x - \tilde{H}_0 \approx H(x_0 - \tilde{x}_0)$) that is not used in the MTE-based verification norm. This approximation would be more problematic in strongly nonlinear observation operators such as radiative transfer models for satellite radiances. The nonlinear observation operator may play a role in relatively low correlation coefficients for the impact estimates verified with the AMSU-A radiances.