Proactive Quality Control: Observing System Experiments Using the NCEP Global Forecast System

TSE-CHUN CHEN AND EUGENIA KALNAY

*Department of Atmospheric and Oceanic Science, and Cooperative Institute for Satellite Earth System Studies,
University of Maryland, College Park, College Park, Maryland*

ABSTRACT

Proactive quality control (PQC) is a fully flow dependent QC based on ensemble forecast sensitivity to observations (EFSO). Past studies showed in several independent cases that GFS forecasts can be improved by rejecting observations identified as detrimental by EFSO. However, the impact of cycling PQC in sequential data assimilation has, so far, only been examined using the simple Lorenz '96 model. Using a low-resolution spectral GFS model that assimilates PrepBUFR (no radiance) observations with the local ensemble transform Kalman filter (LETKF), this study aims to become a bridge between a simple model and the implementation into complex operational systems. We demonstrate the major benefit of cycling PQC in a sequential data assimilation framework through the accumulation of improvements from previous PQC updates. Such accumulated PQC improvement is much larger than the ''current'' PQC improvement that would be obtained at each analysis cycle using ''future'' observations. As a result, it is unnecessary to use future information, and hence this allows the operational implementation of cycling PQC. The results show that the analyses and forecasts are improved the most by rejecting all the observations identified as detrimental by EFSO, but that major improvements also come from rejecting just the most detrimental 10% observations. The forecast improvements brought by PQC are observed throughout the 10 days of integration and provide more than a 12-h forecast lead-time gain. An important finding is that PQC not only reduces substantially the root-mean-squared forecast errors but also the forecast biases. We also show a case of ''skill dropout,'' where the control forecast misses a developing baroclinic instability, whereas the accumulated PQC corrections result in a good prediction.

1. Introduction

To allow efficient estimation of the impact of individual observations in data assimilation (DA), a family of forecast sensitivity to observation (FSO) techniques has been developed in the literature. These techniques construct a mapping between the observation innovations and the resulting short-term forecast error changes due to DA using various approaches. The adjoint-based FSO was first formulated by Langland and Baker (2004), using the adjoint model to propagate back (to observation time) future forecast error changes, and attributing the error changes to each individual observation by minimizing a cost function as in variational DA. The first

ensemble equivalent of the adjoint FSO was developed by Ancell and Hakim (2007) and Torn and Hakim (2008). Their formulation requires a first-order approximation of the response sensitivity to the change in model states, and the observational impact is calculated one observation at a time similar to the serial update in Whitaker and Hamill (2002) to avoid inverting an $O \times O$ matrix, where O is the number of observations. A new formulation was then introduced in Kalnay et al. (2012) which requires no such approximation and calculates the impact of all observations at once at the expense of including an extended forecast, which is available from the long forecasts performed in operation. This ensemble FSO (EFSO) approach directly maps, using the readily available ensemble forecasts from the DA system, the forecast error changes between two consecutive forecasts by DA to each individual observation. Recently, a hybrid approach (HFSO; Buehner et al. 2018) was developed that projects the forecast sensitivity to analysis using the ensemble forecasts as in EFSO, but

---

ⓐ Denotes content that is immediately available upon publication as open access.

Corresponding author: Tse-Chun Chen, tcchen@umd.edu; ekalnay@umd.edu

computes the observational impact through minimization of a cost function as in the adjoint FSO. This approach was designed to be consistent with the ensemble-variational (EnVar) DA systems.

Several operational centers and research groups have implemented at least one of the approaches of FSO to compare the impact of different observing systems on modern DA systems (e.g., Zhu and Gelaro 2008; Cardinali 2009; Gelaro et al. 2010; Lorenc and Marriott 2014; Ota et al. 2013; Sommer and Weissmann 2014). Other studies have explored the applications of FSO impacts. It was shown in Lien et al. (2018) that the long-term-averaged EFSO impact provides detailed information for optimizing data selection and the design of quality control procedures. Kotsuki et al. (2017) found that using EFSO impact as an ordering method in a serial ensemble square root filter for the Lorenz (1996) model significantly improved the analysis accuracy.

Proactive quality control (PQC; Ota et al. 2013; Hotta et al. 2017) was proposed to allow the rejection of detrimental observations based on their EFSO impact, in order to improve the quality of both the analyses and the forecasts. Using the spectral Global Forecast System (GFS) from the National Centers for Environmental Prediction (NCEP), Ota et al. (2013) and Hotta et al. (2017) demonstrated with many independent cases, using 24 and 6 h, respectively, as the EFSO verifying lead time, that denying the detrimental observations identified by the regional EFSO impact significantly reduced the resulting forecast error. Chen et al. (2017), using the same cases as in Hotta et al. (2017), found that rejecting detrimental observations based on global EFSO impact rather than regional impact provided significantly more improvement. Chen and Kalnay (2019), hereafter CK19, further explored cycling PQC in an idealized environment using the Lorenz (1996) model. The idealized experiments clearly showed that among all the deficiencies tested in the DA system, the cycling PQC improvement responded only to those present in the observations, suggesting that PQC effectively removes the impact from the detrimental observations.

In this study, we use a low-resolution spectral GFS model coupled with the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007) and assimilate real observations from the NCEP PrepBUFR dataset, as in Lien et al. (2016, 2018). This study is a bridge between the idealized Lorenz (1996) model used by CK19 and the implementation of EFSO/PQC in a complex operational system. This system with intermediate complexity allows the efficient exploration of the properties of PQC in a realistic model assimilating real observations. Section 2 briefly reviews the EFSO formulation

and the PQC algorithm. The experimental design is covered in section 3. The results are shown in section 4, followed by the summary and discussion in section 5.

2. EFSO formulation and PQC algorithm

Kalnay et al. (2012) derived a simple ensemble forecast sensitivity to observations (EFSO) that estimates for each observation whether its assimilation is beneficial or detrimental to the subsequent forecast [i.e., whether the assimilation of each observation decreases or increases the forecast error, measured, e.g., by the total energy of the error [see Eq. (8)]. If the EFSO value [Eq. (8)] is negative, it indicates that the error has been reduced, and the observation is beneficial. If the EFSO value is positive, the observation increases the error, and the observation is detrimental.

The EFSO formulation introduced here follows Kalnay et al. (2012). In the context of sequential DA with a window size of $\delta t = t_n - t_{n-1}$, suppose we estimate the impact of observations assimilated at time t_{n-1} on forecast error changes at current time t_n. We have two consecutive forecasts $\mathbf{x}^f_{t_n|t_{n-2}}$ and $\mathbf{x}^f_{t_n|t_{n-1}}$ initiated from time t_{n-2} and t_{n-1}, both valid at the verification time t_n. In practice, the truth for calculating the forecast error is unobtainable, and a verifying reference (e.g., the subsequent analysis used in this study) is required to estimate the forecast error response to the data assimilation. Then the associated forecast differences from the subsequent analysis (as verifying reference) are

$$\mathbf{e}_{t_n|t_{n-2}} = \mathbf{x}^f_{t_n|t_{n-2}} - \mathbf{x}^a_{t_n}, \qquad (1)$$

$$\mathbf{e}_{t_n|t_{n-1}} = \mathbf{x}^f_{t_n|t_{n-1}} - \mathbf{x}^a_{t_n}, \qquad (2)$$

where $\mathbf{x}^a_{t_n}$, used as verifying reference, is the posterior estimate from the prior $\mathbf{x}^f_{t_n|t_{n-1}}$ and the observations $\mathbf{y}_{t_n}$ at time t_n from the same experiment. In this setup, the later analysis from the same run was used as the verification reference, and the forecast errors were estimated using the forecast difference terms $\mathbf{e}_{t_n|t_{n-2}}$ and $\mathbf{e}_{t_n|t_{n-1}}$. While acknowledging that they are not true errors but only an estimation of such, we still refer to them as forecast errors hereafter for simplicity. Other verifying references for the EFSO calculation could be later observations (e.g., Sommer and Weissmann 2016; Cardinali 2018; Necker et al. 2018) or the analyses from independent DA systems (e.g., Kotsuki et al. 2019).

The change between the two forecast errors is introduced by the assimilation of the observations at time t_{n-1} using the analysis update equation:

$$\mathbf{x}^a_{t_{n-1}} = \mathbf{x}^f_{t_{n-1}|t_{n-2}} + \mathbf{K}\delta\mathbf{y}_{t_{n-1}}, \qquad (3)$$

$$\delta\mathbf{y}_{t_{n-1}} = \mathbf{y}_{t_{n-1}} - \mathbf{H}\mathbf{x}^f_{t_{n-1}|t_{n-2}},$$

where $\mathbf{K}$, $\mathbf{H}$, $\delta\mathbf{y}_{t_{n-1}}$, and $\mathbf{y}_{t_{n-1}}$ are, respectively, the gain matrix, the observation operator, the observation innovation, and a vector of the observations. The gain matrix $\mathbf{K}$ can be approximated with the ensemble of analysis perturbations using

$$\mathbf{K} = \frac{1}{N-1}\mathbf{X}^a\mathbf{X}^{a\mathrm{T}}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}, \tag{4}$$

where N, $\mathbf{X}^a$, and $\mathbf{R}$ represent the ensemble size, the analysis perturbations, and the observation error covariance matrix.

The change between the squared forecast errors introduced by DA at time t_{n-1} is

$$\Delta(e^2) = \mathbf{e}^\mathrm{T}_{t_n|t_{n-1}}\mathbf{C}\mathbf{e}_{t_n|t_{n-1}} - \mathbf{e}^\mathrm{T}_{t_n|t_{n-2}}\mathbf{C}\mathbf{e}_{t_n|t_{n-2}}$$

$$= \left(\mathbf{e}_{t_n|t_{n-1}} - \mathbf{e}_{t_n|t_{n-2}}\right)^\mathrm{T}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right)$$

$$= \left(\mathbf{x}^f_{t_n|t_{n-1}} - \mathbf{x}^f_{t_n|t_{n-2}}\right)^\mathrm{T}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right), \tag{5}$$

where $\mathbf{C}$ is the chosen error norm matrix. Substituting the Eq. (3) and applying the $\delta t = t_n - t_{n-1}$ forecast operator $\mathbf{M}$ that propagates the errors from time t_{n-1} to t_n into Eq. (5) gives

$$\Delta(e^2) \approx \left[\mathbf{M}\left(\mathbf{x}^a_{t_{n-1}} - \mathbf{x}^f_{t_{n-1}|t_{n-2}}\right)\right]^\mathrm{T}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right)$$

$$= \left[\mathbf{M}\mathbf{K}\delta\mathbf{y}_{t_{n-1}}\right]^\mathrm{T}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right)$$

$$= \delta\mathbf{y}^\mathrm{T}_{t_{n-1}}\mathbf{K}^\mathrm{T}\mathbf{M}^\mathrm{T}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right). \tag{6}$$

Note that the forecast operator $\mathbf{M}$ and its corresponding adjoint $\mathbf{M}^\mathrm{T}$ are merely for symbolic representation in Eq. (6). None of them are required in the actual calculation of EFSO.

We further substitute Eq. (4) and the approximations of $\mathbf{H}\mathbf{X}^a_{t_{n-1}} \approx \mathbf{Y}^a_{t_{n-1}}$ and $\mathbf{M}\mathbf{X}^a_{t_{n-1}} \approx \mathbf{X}^f_{t_n|t_{n-1}}$ into Eq. (6) and get

$$\Delta(e^2) \approx \frac{1}{N-1}\delta\mathbf{y}^\mathrm{T}_{t_{n-1}}\mathbf{R}^{-1}\mathbf{Y}^a_{t_{n-1}}\mathbf{X}^{f\mathrm{T}}_{t_n|t_{n-1}}\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right), \tag{7}$$

where $\mathbf{Y}^a_{t_{n-1}}$ is the analysis perturbation in observation space, and $\mathbf{X}^f_{t_n|t_{n-1}}$ is the forecast perturbation initiated at time t_{n-1} and valid at time t_n.

Due to the limited size of the ensemble, a localization for error covariance is required to suppress the spurious long-distance correlations. Applying such localization to Eq. (7) gives the EFSO impact formulation:

$$\Delta e^2 \approx \frac{1}{N-1}\delta\mathbf{y}^\mathrm{T}_{t_{n-1}}\mathbf{R}^{-1}\left[\rho \circ \left(\mathbf{Y}^a_{t_{n-1}}\mathbf{X}^{f\mathrm{T}}_{t_n|t_{n-1}}\right)\right]\mathbf{C}\left(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}}\right), \tag{8}$$

where ρ is a localization matrix and $\circ$ represents an element-wise multiplication (Schur product).

We can then obtain the impact of each observation by decomposing the sum of the inner product of the innovation vector $\delta\mathbf{y}_{t_{n-1}}$ and the error sensitivity vector $[\partial\Delta(e^2)/\partial\delta\mathbf{y}_{t_{n-1}}] = (1/N-1)\mathbf{R}^{-1}[\rho \circ (\mathbf{Y}^a_{t_{n-1}}\mathbf{X}^{f\mathrm{T}}_{t_n|t_{n-1}})]\mathbf{C}(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}})$ into elements that correspond to each observation, so that $(1/N-1)\delta y_{t_{n-1},l}[\mathbf{R}^{-1}[\rho \circ (\mathbf{Y}^a_{t_{n-1}}\mathbf{X}^{f\mathrm{T}}_{t_n|t_{n-1}})]\mathbf{C}(\mathbf{e}_{t_n|t_{n-1}} + \mathbf{e}_{t_n|t_{n-2}})]_l$ is the estimated impact of the lth observation. This estimated impact is represented by the short-term forecast error changes due to the assimilation of the corresponding observation. We define the observations with positive EFSO impact, which increase the forecast error, as *detrimental* observations. Conversely, the observations with negative EFSO impact, which decrease the forecast error, are defined as *beneficial* observations.

PQC utilizes the EFSO impact as a QC criterion for the observations in each DA cycle. At each cycle, PQC first identifies the detrimental observations from all the assimilated observations in the cycle that was used to generate the analysis, and such analysis is then updated by repeating the data assimilation again without using the detrimental observations. Figure 1, adapted from Hotta et al. (2017) and CK19, shows the flowchart for the PQC algorithm. BKGD and ANAL stand for the pre-PQC background and the pre-PQC analysis in the sequential data assimilation. PQC_A represents the PQC-updated analysis, and PQC_B is the subsequent forecast from PQC_A. PQC_B then serves as the background for generating ANAL. In the algorithm, a temporary analysis REF_A is created to serve as a verifying reference in the EFSO calculation. The PQC algorithm up to the current time t_n can be summarized in the following steps:

1) Starting from the pre-PQC analysis at t_{n-1}, run one regular DA cycle to get both the pre-PQC background and the reference analysis at t_n ($\text{ANAL}_{t_{n-1}} \to \text{BKGD}_{t_n} \to \text{REF\_A}_{t_n}$).
2) Obtain the forecast valid at time t_n that is initiated from the PQC-updated background at t_{n-1} ($\text{PQC\_B}_{t_{n-1}} \to \text{FCST}_{t_n}$).
3) Compute the δt-EFSO impact (i.e., the observation impact on the forecast valid δt later) and identify the detrimental observations at time t_{n-1}. Note that the EFSO impact for the observations at time t_{n-1} is computed at time t_n when REF_A_{t_n} becomes available ($\text{FCST}_{t_n} + \text{BKGD}_{t_n} + \text{REF\_A}_{t_n} \to \text{EFSO}_{t_{n-1}}$).
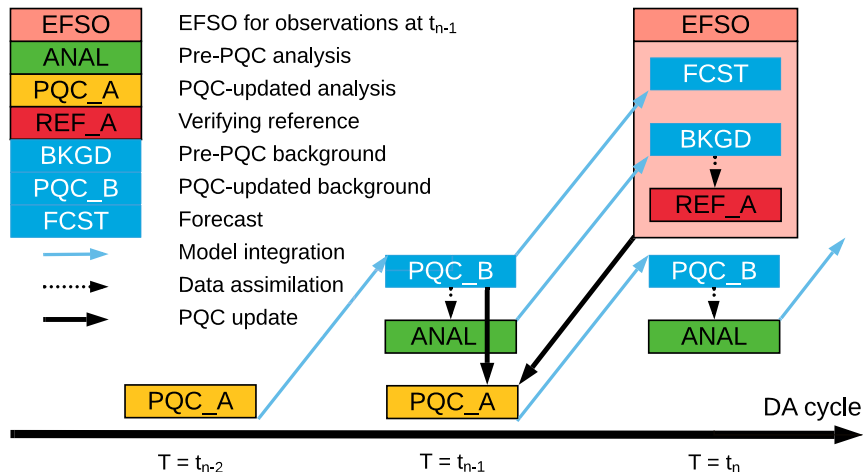4) Reject the detrimental observations based on the δt-EFSO impact and assimilate the beneficial

FIG. 1. Flowchart of cycling Proactive QC algorithm [adapted from Hotta et al. (2017) and CK19]. BKGD and ANAL stand for pre-PQC background and analysis in sequential data assimilation. PQC_A represents the PQC-updated analysis, and PQC_B is the subsequent forecast from PQC_A. PQC_B serves as the background for generating ANAL. In the algorithm, a temporary analysis REF_A is created to serve as the verifying reference for EFSO computation. We verify the performance of the proposed PQC scheme by comparing the ANAL and the forecasts initiated from ANAL in the PQC experiments with the control analysis and forecast unless otherwise stated.

observations using $PQC\_B_{t_{n-1}}$ as the background to obtain the PQC-updated analysis valid at t_{n-1}, which contains no identified detrimental impact. Note that this PQC-update for the analysis valid at t_{n-1} takes place at time t_n when the observations and the reference analysis valid at t_n become available ($EFSO_{t_{n-1}} + PQC\_B_{t_{n-1}} \rightarrow PQC\_A_{t_{n-1}}$).

5) The PQC-updated background valid at t_n is initiated from the PQC-updated analysis at t_{n-1} and serves as the background for the production of the next pre-PQC analysis at t_n ($PQC\_A_{t_{n-1}} \rightarrow PQC\_B_{t_n} \rightarrow ANAL_{t_n}$).

We note that PQC is using the EFSO impact in this study, but it could also use instead the adjoint or the hybrid formulations of FSO impact.

We further define the "current" PQC-update at current time t_n as the production of $PQC\_A_{t_n}$ from the pre-PQC analysis $ANAL_{t_n}$, which in turn benefits from all the past PQC-updates ("accumulated" PQC-update). The pre-PQC analysis $ANAL_{t_n}$ contains only the information up to the current time t_n, while the current PQC-update (production of $PQC\_A_{t_n}$) requires the "future" information ($REF\_A_{t_{n+1}}$ and hence the observations required to make it) from the next cycle at time t_{n+1}. It should be emphasized that we are promoting the application of the accumulated PQC-update into operational use since the current PQC-update is not feasible in operational applications as it requires future information. However, as we will show in the results, the accumulated improvement is much larger than the

current improvement, supporting the PQC application in operations.

3. Experimental design

a. DA system configuration

In this paper, we aim to explore the impact of cycling PQC using an intermediate complexity system to act as a bridge between the promising results obtained in CK19 using the simple Lorenz (1996) model and a full operational system. For this purpose, we use the GFS-LETKF system developed by Lien et al. (2016), which reduces the complexity of the DA system and uses a lower resolution of the spectral GFS model (compared to the operational systems) in order to expedite the execution of the experiments while still using a realistic configuration of a forecasting model, data assimilation system, and observations.

The underlying philosophy behind the design of the GFS-LETKF system is to have a simple configuration of a DA system coupled with the realistic spectral GFS model to allow fast experiments exploring innovative data assimilation techniques. The DA system consists of a generic and simple 4D-LETKF core code developed by Takemasa Miyoshi, and an interface to the spectral GFS model (available at https://github.com/takemasa-miyoshi/letkf). It preserves the flexibility to switch the GFS resolution from T62 to T574, and the choice of observation operators using the built-in conventional data

TABLE 1. Description, observed variables, and total counts of each observation type.

Observation types	Descriptions	Variables	Counts
ADPSFC	Surface land (synoptic, METAR) reports	Ps	326 724
ADPUPA	Upper-air (raob, PIBAL, RECCO, DROPS) reports	Q, T, U, V	4 092 312
AIRCAR	MDCRS ACARS aircraft reports	T, U, V	4 087 880
AIRCFT	AIREP/PIREP, AMDAR, MDCRS/ACARS, TAMDAR aircraft reports	T, U, V	3 300 461
PROFLR	Wind profiler and acoustic sounder (SODAR) wind reports	U, V	275 700
QKSWND	QuikSCAT scatterometer data (reprocessed)	U, V	980 575
SATWND	Satellite-derived wind reports	U, V	5 649 557
SFCSHP	Surface marine (ship, buoy, C-MAN platform) reports	Ps, Q, T, U, V	243 199
SPSSMI	SSM/I retrieval products (reprocessed wind speed, tpw)	U, V	835 222
SYNDAT	Synthetic tropical cyclone bogus reports	U, V	334
VADWND	Vertical azimuth display (VAD) winds from WSR-88D radars	U, V	240 261

operator (simple spatial interpolation for PrepBUFR data) or the Gridpoint Statistical Interpolation (GSI) system from NCEP to ingest advanced observations such as the satellite radiances. For computational efficiency, we chose to perform the experiments with a T62 resolution and used the built-in simple observation operator. We used 32 ensemble members for this low-resolution configuration as in Lien et al. (2016).

We assimilated at a 6-h interval only the conventional observations from the PrepBUFR dataset provided by NCEP (see Table 1 for the assimilated observation types). In our 4D-LETKF setup, the observations within the 6-h DA window were binned into 7 time slots centered on the hour from the start (-3 h) to end ($+3$ h) time in order to account for the background evolution within the window. The error statistics of the observations are also extracted from the same dataset. Adjusting to the low-resolution of our system, the observations were superobbed/thinned (see Table 2) to at most only one observation per 3D model grid point for each data type and variable in one assimilation window, which reduced the data density to one-third of the original.

Since only the conventional observations from PrepBUFR data were assimilated, the ozone concentration and the sea surface temperature were obtained from the NCEP Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) to prevent the system from long-term drift, although this should not be a significant issue for our experimental length (one month). Note that both the control and the PQC experimental runs were ingesting the CFSR data for these nonmeteorological prognostic variables so that it is a fair comparison. Additionally, the CFSR dataset, which has a much higher resolution (T382) and assimilates many more observations (including the satellite radiances) also served as the verifying truth for measuring the improvement obtained from the PQC-updates. The performance of both the control and the PQC-updated analyses and forecasts were measured with both root-mean-squared error (RMSE) and BIAS against the CFSR to show the geographical distribution of the impact. We stress here that the CFSR was not used in the EFSO calculation, and the verifying reference $\text{REF}\_A_{t_n}$ was obtained by assimilating the observations at t_n to the

TABLE 2. Superobing and thinning of each observation type. A description of the observation types can be found in Table 1.

Observation types	Methods
ADPSFC	Leave only one observation closest to the base time
	Avg observations with minimum errors within the 3D grid
ADPUPA	Avg all observations in the vertical grid
AIRCAR	Avg observations with minimum errors within the 3D grid
AIRCFT	Avg observations with minimum errors within the 3D grid
PROFLR	Avg all observations in the vertical grid
	Leave only one observation closest to the base time
QKSWND	Avg observations with minimum errors within the 3D grid
SATWND	Avg observations with minimum errors within the 3D grid
SFCSHP	Avg observations with minimum errors within the 3D grid
	Leave only one observation at the base time and do not use any observations off the base time
SPSSMI	Avg observations with minimum errors within the 3D grid
SYNDAT	Avg observations with minimum errors within the 3D grid
VADWND	Avg all observations in the vertical grid
	Leave only one observation closest to the base time

pre-PQC background $BKGD_{t_n}$ from the same run. The adaptive multiplicative inflation (Miyoshi 2011, prior variance of the inflation parameter: $v_i^b = 0.04^2$) and the relaxation to prior perturbation (RTPP; Zhang et al. 2004, relaxation parameter: $\alpha = 0.5$) were added to account for model error. A fixed horizontal length scale of 500 km and a vertical length scale of 0.4 scale height were chosen for the R localization (see Greybush et al. 2011), accounting for the insufficient ensemble size. The above setup was extensively used in Lien et al. (2016, 2018).

Note that the Proactive QC scheme is proposed to be complementary to standard QC checks, and not their replacement. Several QC checks were implemented into our GFS-LETKF system. First, the observations with observational error larger than the prescribed gross error thresholds (10^7) were not assimilated. Second, some observation types were excluded from assimilation mainly following the PrepBUFR report types used by Global GFS and GDAS GSI analyses (refer to https://www.emc.ncep.noaa.gov/mmb/data_processing/prepbufr.doc/table_2.htm). Third, we also filtered out the observations with bad quality markers from the PrepBUFR dataset. Fourth, the observations with innovation 5 times larger than its error provided in the PrepBUFR dataset were also excluded. These standard QC checks were applied both in the control and the PQC experiments. Hence the experiments show the PQC improvement in addition to the standard QC checks.

The experimental period spans from 0000 UTC 1 January to 0000 UTC 6 February 2008, with the first 5 days used as the DA spinup period. The period is chosen simply based on the availability of an existing database used by our group, and the choice of the experimental period should be irrelevant to our conclusions.

b. EFSO and PQC setup

To suppress the spurious error correlations associated with the insufficient ensemble size, localization is required in calculating EFSO, as in ensemble DA. We follow the same localization advection method applied in Ota et al. (2013) and Hotta et al. (2017), which advects the center of the initial localization function using the equally weighted average of the analysis and 6-h forecast horizontal winds at each grid to keep track of the flow-following correlation structure. We note that both studies applied a factor of 0.75 to the advection of localization function, but such a factor was not applied in this study. We assumed that the 6-h lead time is short enough so that it is acceptable to ignore the sensitivity of the EFSO/PQC performance to the advection factor. This assumption, however, is worth exploring in the future. Also, we adopted the moist total energy error

norm (MTE; Ehrendorfer 2007) for **C** in Eq. (8) to account for all the variables of meteorological interests and combine their different units naturally in terms of energy. Rather than calculating the regional EFSO impact for just the forecast skill dropout regions as in Ota et al. (2013) and Hotta et al. (2017), we calculated the EFSO impact for the entire global domain for all variables. Yet, the long-range impact of an observation was still limited by the applied localization, hence only the subdomain within the localization radius had an actual impact in the total error metric. We note that EFSO impact is norm dependent and the forecast changes by PQC are subject to the choice of such error metric.

For the selection of the EFSO verifying lead time, Ota et al. (2013) performed experiments with 24-h lead time, but Hotta et al. (2017) found that the 6-h lead time is long enough to perform PQC, in the sense that the EFSO impacts using the two lead times identified essentially the same detrimental observations. Using a lead time as short as 6 h is favorable for the operational use of PQC for two reasons. First, the lead time matches the DA window, and hence the 6-h forecast perturbation $\mathbf{X}_{t_n|t_{n-1}}^f$ required by the EFSO calculation is available from the EnKF system. Second, the subsequent analysis used as verifying reference is only one cycle later than the PQC-updated analysis. Using lead times longer than 6 h will increase the computational costs on extending the ensemble forecast length from the original length for EnKF and the additional intermediate DA cycles between the PQC-updated analysis and the analysis used as verifying reference. Therefore, we used the 6-h lead time in the experiments in this study. We note that even shorter lead times might be appropriate for a regional system with shorter update cycles.

This study aims to demonstrate that cycling PQC performs well not only in Lorenz (1996) model as shown in CK19 but also in a realistic model using real observations. Additionally, we would like to address the following two important questions: *1) How important is the benefit from the accumulation of the successive PQC improvement on the analyses (i.e., the accumulated impact of PQC when it is cycled)? More precisely, is the accumulated PQC-update large enough so that we could skip the current PQC-update, which requires waiting 6 h for future observations? 2) How sensitive is cycling PQC to the number of rejected observations?* To answer these questions, four experiments were conducted and summarized in Table 3. The "CNTL" is the control run that did not reject any detrimental observations identified by EFSO (no PQC). The "PQC" experiments reject the observations following the threshold method in CK19 which rejects any observations having EFSO impact larger than the threshold for the specified rejecting percentage. The threshold for each rejecting

TABLE 3. A list of experiments performed with the GFS-LETKF system with the PQC configurations. The thresholds for the rejecting percentages were obtained by their corresponding percentile values of EFSO impact statistics from the CNTL. Any observations having EFSO impact larger than such threshold are rejected in each experiment.

Expt name	Cycling PQC	Threshold (J kg^{-1})	Rejecting percentage	Note
CNTL	No	—	—	Control
PQC-10	Yes	1.67×10^{-5}	10%	Full cycling PQC
PQC-30	Yes	1.41×10^{-6}	30%	Full cycling PQC
PQC-50	Yes	0	50%	Full cycling PQC

percentage was obtained by the corresponding percentile value of the EFSO impact statistics from the CNTL. For example, the most detrimental 10% of the observations were those having an impact larger than the 90th percentile value from the CNTL, since positive EFSO values represent detrimental impacts. In this case, the 90th percentile value is the threshold for the rejecting percentage of 10%.

A schematic of the PQC algorithm (adapted from CK19) in sequential data assimilation is shown in Fig. 1. The schematic shows that the reference analysis required to perform EFSO would be using information from the next cycle. For instance, calculating EFSO for observations at time t_{n-1} requires the reference analysis at time t_n. And the reference analysis at time t_{n+1} would be required if we were to perform a current PQC-update and produce the PQC-updated analysis PQC_A at t_n. By contrast, the pre-PQC analysis ANAL contains only information up to time t_n. To make a fair assessment on the PQC performance so that the results are applicable to the operational NWP, we did not examine the accuracy of the PQC-updated analysis (PQC_A), which was obtained from the current PQC-update that requires information from the next cycle t_{n+1} and hence would not be possible in operations. Instead, *the pre-PQC analysis (ANAL) and the subsequent 10-day forecast, which contain only information up to time t_n of the PQC experiments, were used to compare against the CNTL analysis and forecast*. In addition to absolute RMSE, we also used relative RMSE reduction = $[(\text{RMSE}_{\text{PQC}} - \text{RMSE}_{\text{CNTL}})/\text{RMSE}_{\text{CNTL}}]$ to show the percentage of the control RMSE reduced by PQC.

To examine the benefit from the successively improved analyses and the accumulated improvement in the backgrounds, we differentiated the improvement from the current PQC-update in addition to that from the accumulated PQC-update by comparing the forecasts initiated from the pre-PQC analyses (ANAL) and the PQC-updated analyses (PQC_A). The difference between the forecasts from ANAL and those from PQC_A provides an estimate of the improvement made only by the current PQC update. The "PQC-10," "PQC-30," and "PQC-50" experiments compare the PQC performance

with 6-h lead time for different observation rejecting percentages of 10%, 30%, and 50%, respectively. Here, the 50% simply stands for all the detrimental observations since the averaged detrimental percentage of observations is about 50% as widely reported from most applications (e.g., Gelaro et al. 2010).

This study aims to show a proof of concept that PQC works not only for a simple model but also for the realistic GFS model that assimilates real observations. To simplify the investigation and implementation of PQC into the system, the straightforward data denial method, equivalent to the PQC_H method in CK19, was adopted for PQC-update. The PQC_H method is not optimal for PQC-update as shown in CK19 due to its higher computational requirements, and the unintentional inflation of the ensemble (compared to the pre-PQC analysis) that leads to an inconsistency between the increased spread and the more accurate PQC-updated analysis. However, in this study, such unintentional change in spread by the data denial was alleviated by the RTPP (Zhang et al. 2004) and the adaptive inflation (Miyoshi 2011) applied in our DA system. Since the posterior perturbations were relaxed toward the prior perturbations, and the posterior spreads of both the pre-PQC and the PQC-updated analyses (ANAL and PQC_A) were adjusted based on the same innovation statistics derived by Desroziers et al. (2005), we did not experience the same degradation when rejecting too many (e.g., 50%) observations using the data denial (PQC_H) method as we did in CK19, where no relaxation method was applied. We showed in CK19 that the gain-reusing (PQC_K) method is more accurate since it reuses the original gain matrix that produced the pre-PQC analysis (ANAL) and it preserves the ensemble spread of the PQC-updated analysis (PQC_A) to be the same as the pre-PQC analysis (ANAL). Thus, the PQC_K method is a cheaper and more accurate update method than PQC_H, which we used in this study, but, on the other hand, implementing PQC_K in this study would require a considerable additional effort in modifying the DA system. We, therefore, used PQC_H, but note that PQC_K should be further explored and compared in future studies.

4. Results

a. EFSO impact

Figure 2 shows the geographical distribution of the observations assimilated during the experimental period over the world, and Fig. 3 the corresponding distribution of each observation type. Note that, for clarity, only one of every 400 observations are shown in Fig. 2. For both figures, the detrimental observations (red) appear mixed with the beneficial observations (blue), regardless of the observation type. The observations with larger impacts were generally distributed in regions with fewer radiosondes like the oceans in the Southern Hemisphere. It is noticeable that the individual observational impact in the United States and Europe was generally much smaller than that anywhere else in the world due to the high density of observations in these regions. And the impact magnitudes were typically larger over the mid- to high-latitude oceans and lands, which may result from the combined effect of data scarcity and the prevailing baroclinic instability. There seemed to be more clustering of detrimental ADPSFC observations in Central America, Central Africa, South Asia, and Western Australia. The SFCSHP observations provided a very large beneficial impact over the oceans in the Southern Hemisphere. Each of the SYNDAT observations may provide a significant contribution to the forecasting of tropical cyclones given their generally large impact magnitude. These characteristics of the impact geographical distribution deserve a more focused analysis.

The EFSO total impact and the impact per observation of each observing system type are summarized in Figs. 4a and 4b, respectively. Table 1 provides a brief description of the observation types (based on the NCEP documents at https://www.nco.ncep.noaa.gov/sib/decoders/BUFRLIB/toc/prepbufr/prepbufr_bftab/). Consistent with the general findings from past studies (e.g., Zhu and Gelaro 2008; Cardinali 2009; Gelaro et al. 2010; Lorenc and Marriott 2014; Ota et al. 2013; Sommer and Weissmann 2014), the most beneficial observation types were the satellite feature-tracking winds, upper-air soundings, and aircraft measurements due to their advantages in quality, number, and/or location. Surprisingly, the surface pressure measurements from synoptic reports and aviation routine weather reports (METARs) had a net slightly detrimental impact in this experiment. Further investigation shows that the majority of the detrimental impact of this observation type originated from its update in the moisture field. The moist total energy norm we adopted for EFSO computation includes the kinetic energy (u and v winds), potential energy (temperature and surface pressure), and moisture
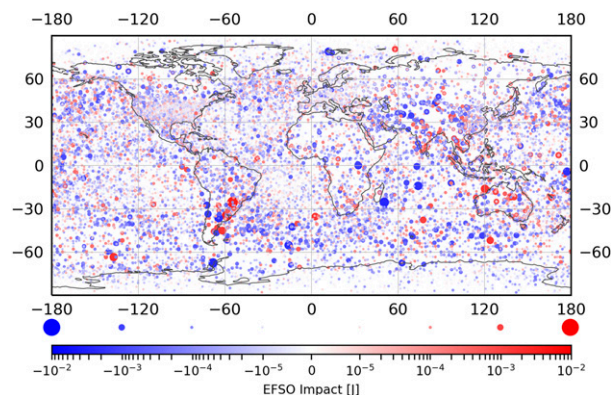

FIG. 2. Geographical distribution of one of every 400 observations throughout the experimental period and their EFSO impact (J). Red (blue) represents detrimental (beneficial) observations. The magnitude of EFSO impact is shown by the size of each dot.

norm (specific humidity). We found that the surface pressure data was very beneficial in both the kinetic and potential energy norms, but it was more detrimental in the moisture norm (not shown), indicating that the moisture update from surface pressure data degraded the analysis and forecast. The synthetic tropical cyclone bogus reports (SYNDAT) rank as the top beneficial type on the impact per observation, indicating that they had a key role in improving at least some of the forecasts of tropical cyclones.

Operationally, each observation is received at different times throughout the analysis window, so we explored how this time difference relates to their impact. Our 4D-LETKF system binned the asynchronous observations into seven time slots from the start (-3 h) to the end ($+3$ h) time of the 6-h window, so that each slot was on the hour, and the fourth slot centered at the analysis time. Figure 4c shows the averaged impact per observation for each time slot. It is clear from this figure that the observations in the first two time slots were significantly less beneficial when compared to the five later time slots, indicating that the later observations provided more useful information. On the other hand, the impacts of the last five time slots were very similar.

b. Global distribution of PQC error reduction

Figure 5a shows the global view of the monthly averaged analysis and 24-h forecast error reduction by the accumulation of PQC-updates (i.e., pre-PQC analysis/ANAL and subsequent forecast without using any future information) for several representative variables: u wind and v wind at 500 hPa, temperature at 700 hPa, and specific humidity at 850 hPa (using experiment "PQC-30" as an example). It is clear that PQC
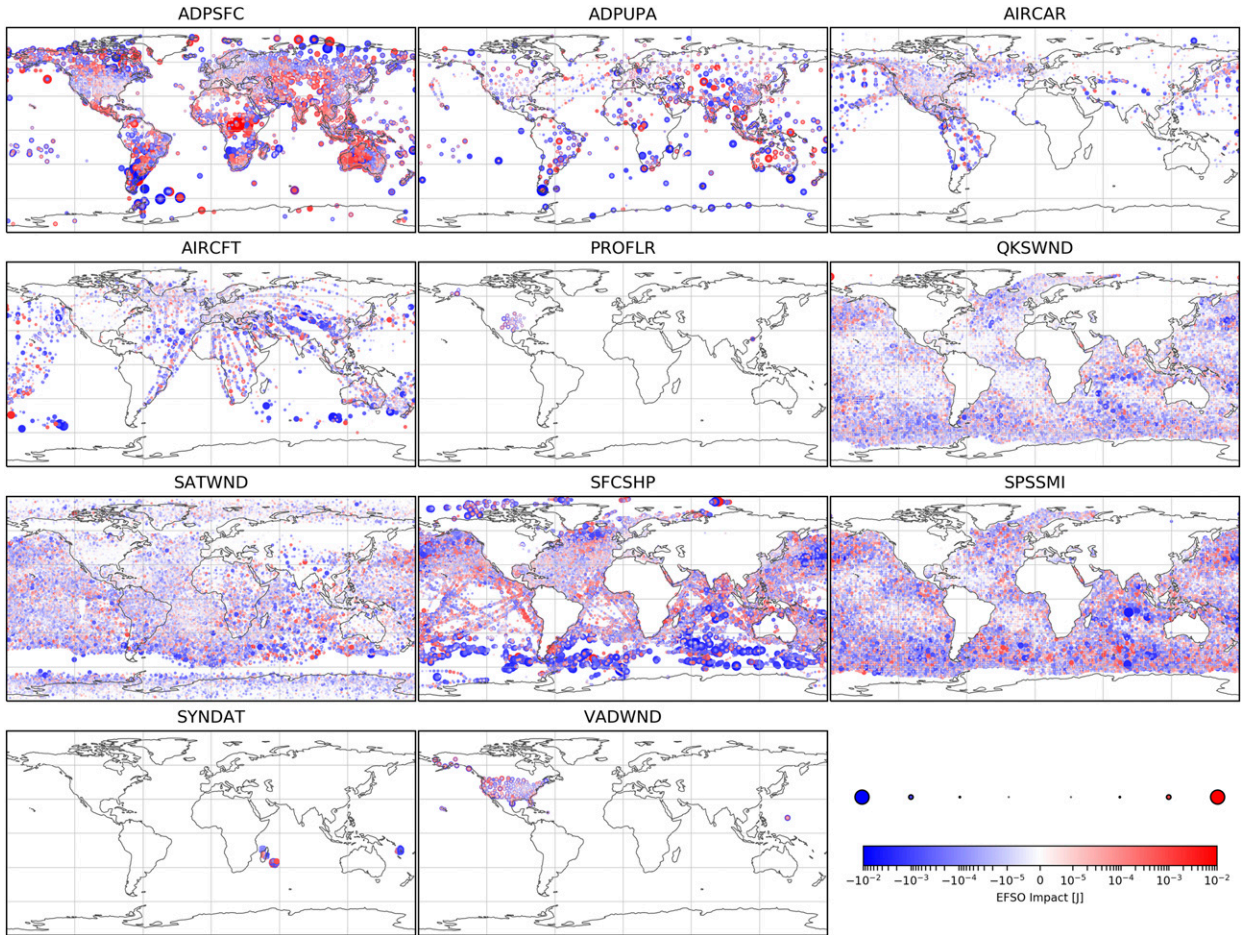
FIG. 3. As in Fig. 2, but for different observation types. All observations are shown for ADPSFC, PROFLR, SFCSHP, SYNDAT, and VADWND. One of every 100 observations are shown for ADPUPA, AIRCAR, AIRCFT, and SATWND. For QKSWND and SPSSMI, one of every 20 observations are shown. A description of observation types can be found in Table 1.

consistently reduces the analysis error almost everywhere on the globe for all the variables. For wind and temperature, the reduction is greatest over the oceans in the Southern Hemisphere, while the improvement in specific humidity mainly takes place in lower latitudes over the ocean. The distribution of the error reduction is probably associated with the dominant dynamical characteristics and observation density. Figure 5a indicates that the mid- to high-latitude oceans in the Southern Hemisphere may have more "room for PQC improvement" due to the lack of conventional observations, as well as the prevailing baroclinicity over the region. The high resemblance between the error reduction from the accumulated PQC-update in 24-h forecasts and that in analyses shows that the PQC improvements were quite persistent during the first 24 h of the forecasts.

The right side of Fig. 5b shows the same map view, but for the error reduction produced only by the current PQC-updates (i.e., the difference between the PQC-updated analysis/PQC_A and the pre-PQC analysis/ANAL) that would use the 6-h future observations. The magnitude of error reduction by the current PQC-updates was considerably smaller than that obtained by the accumulated past PQC-updates. Note the values are doubled to show some features with the same color bar as in Fig. 5a. The current PQC-updates, which we did not include because they would require using future observations, improved very slightly the wind and temperature analyses on the Eurasian continent and northern Pacific, and the humidity analysis in the Maritime Continent, Australia, Central America, and the southern part of South America. In North America, the wind and temperature analyses showed slight degradation by the current PQC-updates. However, such slight degradation became an improvement after 24 h of model forecast. The 24-h wind and temperature forecasts improved the most in the midlatitudes in both the Northern and Southern Hemispheres, and the humidity forecasts showed more improvement
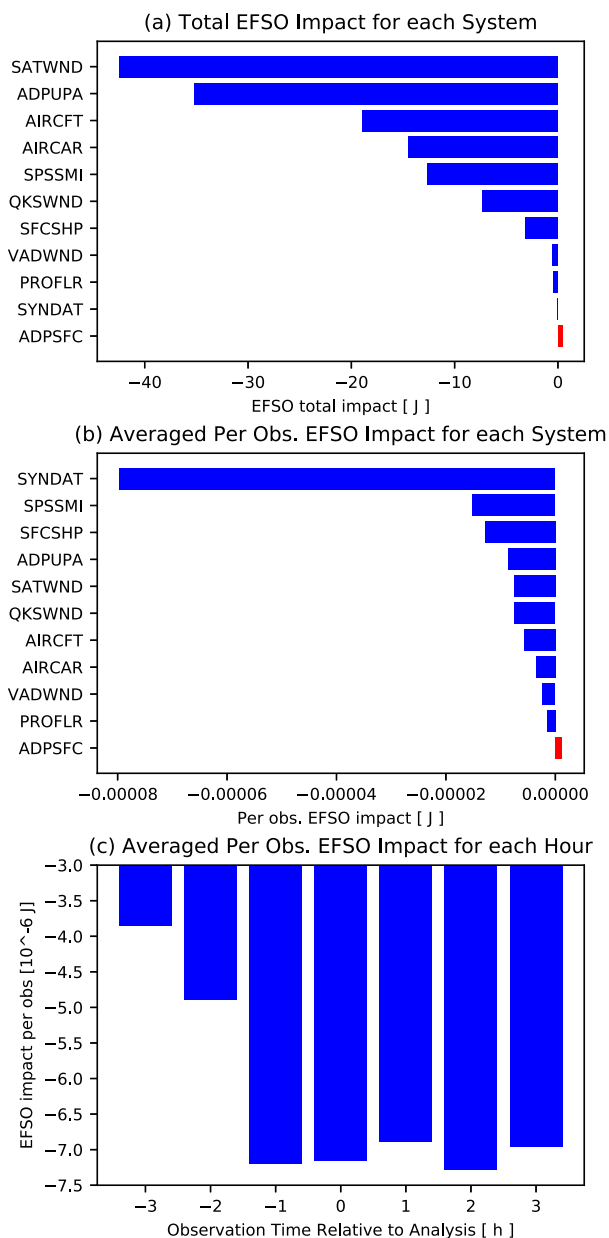
FIG. 4. (a) EFSO total impact (J) and (b) averaged EFSO impact per observation (J) for each PrepBUFR observation type in descending order. (c) Averaged EFSO impact per observation (J) for seven 4D-LETKF time slots that accounts for asynchronous observation time. The detrimental impact is shown in red. (See Table 1 for the descriptions of the observation types.)

over the lower latitudes, which is consistent with the error reduction pattern by the accumulated PQC-updates.

The vertical profiles of the analysis and 24-h forecast RMSE are shown in Fig. 6 for u wind, temperature, and specific humidity. The black lines represent the CNTL experiment. PQC-50 (magenta) significantly reduced the analysis and forecast errors at all pressure levels. For

u wind and temperature, the control RMSE was largest in the Southern Hemisphere at the upper levels, while the majority of the error in specific humidity came from lower- to midlevels in the tropics. The PQC error reduction at each level was roughly proportional to the error magnitude for u wind and specific humidity. For temperature, the error reduction was almost uniformly distributed throughout each level.

c. Comparison of accumulated and "current" PQC improvements

To assess the benefit of cycling PQC as opposed to noncycling PQC, and demonstrate that cycling PQC improves the quality of analyses and forecasts even without using any future information, we compared the error reduction of forecasts initialized from the pre-PQC analysis (ANAL) and the PQC-updated analysis (PQC_A). The former contains only the improvement from previous PQC-updates, and the latter includes the latest current PQC-update using the future analysis as verifying reference for evaluating EFSO impact for current observations. We refer to the PQC-update in ANAL as the accumulated update, PQC_A as the total update, and the difference between the two as the current update. The current error reduction is defined as the analysis improvement obtained by instantaneous PQC-update in the current DA cycle, while the accumulated counterpart is defined as the improvement originating from the improved background due to previous PQC-updates. The main benefit of cycling PQC is the accumulated improvements throughout the past DA cycles since the improved subsequent forecast from the PQC-updated analysis serves as a more accurate background and thus further enhances the accuracy of the following analysis. From a fully cycling PQC experiment, we extracted the current component from total improvement by comparing the pre-PQC analysis (ANAL) and its subsequent forecast, which were only improved by the past PQC-updates (but not the current PQC), with the analysis (PQC_A) and the forecast directly updated by the latest PQC. Figure 7 compares the 10-day relative forecast error reduction of the total, accumulated, and current improvements in the PQC-50 experiment. As we can see, the primary advantage of cycling PQC came from the accumulation of past PQC-updates. The current error reduction (due to the future observations not available operationally) was negligible in the tropics and the Southern Hemisphere. The future observations would introduce improvements of only less than 5% relative to the CNTL in the NH, whereas the accumulated PQC improvement provides 13% or more error reduction, which is consistent with the results in Fig. 5. Interestingly, the benefit from the accumulated
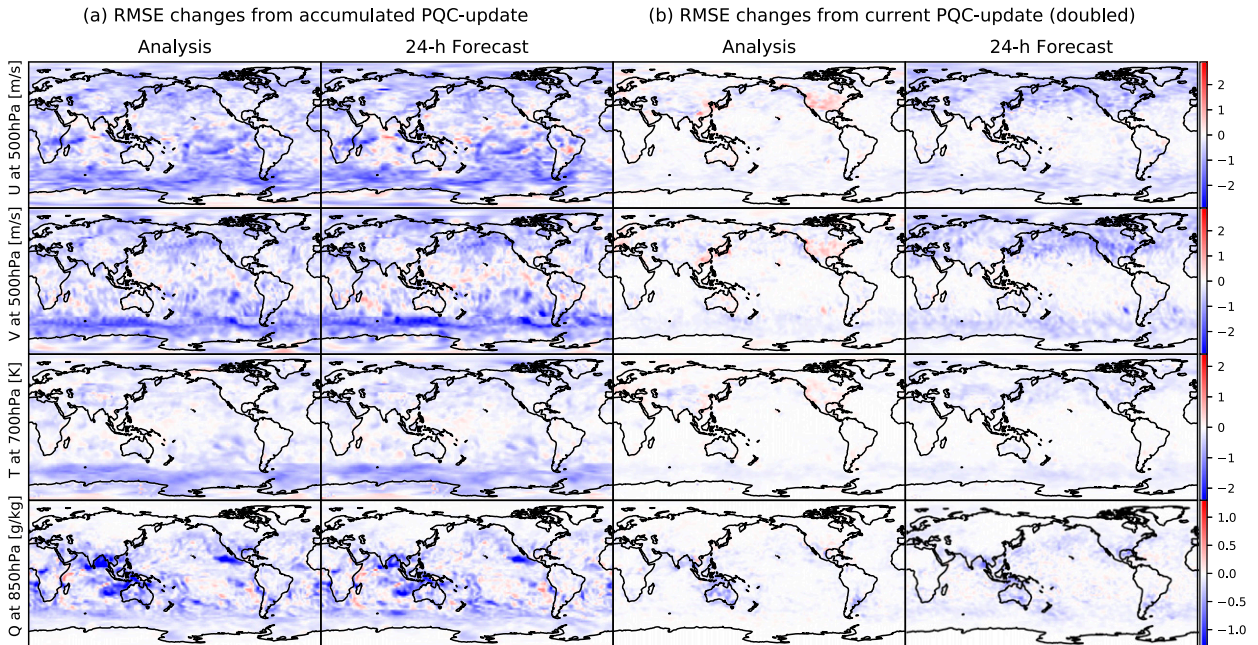
FIG. 5. (a) (left) Monthly mean analysis and (right) 24-h forecast RMSE changes from the accumulated PQC-update for the u wind and v wind (m s^{-1}) at 500 hPa, temperature (K) at 700 hPa, and specific humidity (g kg^{-1}) at 850 hPa in experiment PQC-30. (b) As in (a), but for the doubled RMSE changes from the "current" PQC-update.

impact has a more significant contribution to the total PQC impact in the tropics and Southern Hemisphere compared to that in the Northern Hemisphere, where the conventional observations are much more abundant. It is worth pointing out that the majority of the most reliable observing system, the radiosondes, is located in the Northern Hemisphere.

The fact that the accumulation of past improvements contributes to a major portion of the full impact of cycling PQC has two important implications. First, the PQC improvement had a long-lived impact that remains even after several DA cycles. Second, this supports the feasibility of implementing PQC in operational NWP. Operational centers need to initiate the forecast as soon as the analysis is completed to deliver timely forecast products, so we can only afford to perform PQC after the current forecast is done and released, meaning that the current update from PQC is not possible in operations. Therefore, the huge dominance of the accumulated past corrections provides encouraging evidence that even without the latest current update from PQC in the current cycle, the forecast improvement from the accumulated PQC can still be close to that from the total PQC-update.

To illustrate the global forecast error reduction from a different perspective, we now define the *gained forecast lead time* as the δt such that RMSE_PQC$_{t'+\delta t|t} \approx$ RMSE_CNTL$_{t'|t}$, where t and t' are the forecast initial

time and the CNTL forecast valid time. A schematic of the definition is shown in Fig. 8. The gained forecast lead time represents the extra forecast time for the accumulated PQC-updated forecast error to reach the same level as the CNTL forecast error. We show the gained forecast lead time from the accumulated PQC-update in the experiment PQC-50 for the u wind at 500 hPa, the temperature at 700 hPa, and the specific humidity at 850 hPa in Fig. 9. The result shows that for most of the forecast times, the gained forecast lead time was more than 12 h, and for specific humidity, it was over 18 h. Note the gray area masks the periods when adding the forecast time and the gained lead time exceed the 10-day forecast range.

d. Sensitivity of forecast improvement to the rejecting percentage of observations

One of the major focuses in Hotta et al. (2017) was on designing the data-denial strategy, due to the speculation that rejecting all the detrimental observations could lead to forecast degradation since they consist of close to 50% of the total counts. However, as shown in CK19 using Lorenz (1996) model, rejecting the most detrimental 10% observations based on global EFSO impact contributed to the majority of the forecast improvement, and the maximum improvement was obtained by removing nearly all the detrimental observations (around 30%–50%). Here we examine whether this
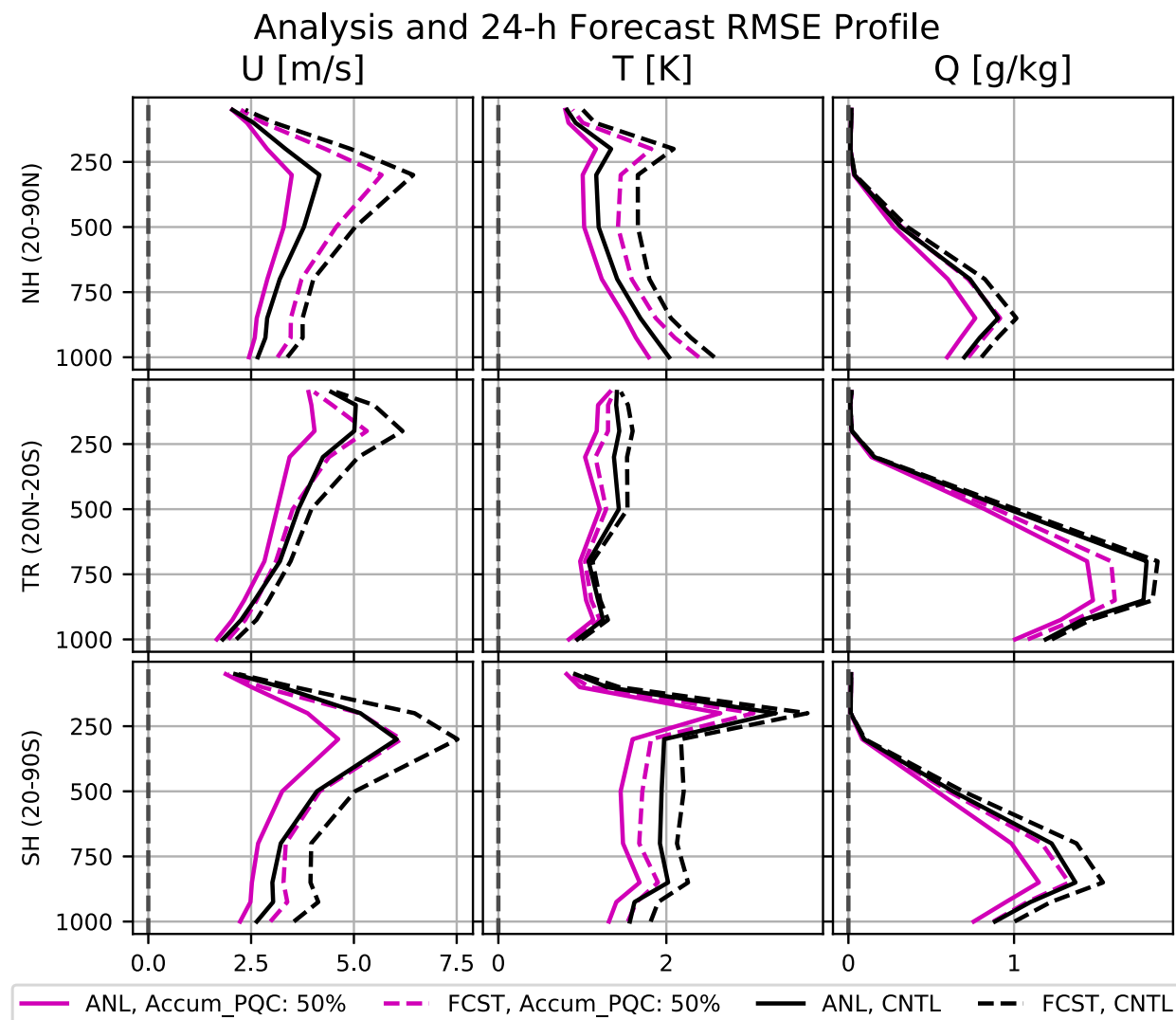
FIG. 6. Monthly mean profile of analysis (solid) and 24-h forecast (dashed) RMSE in PQC-50 (magenta) and CNTL (black) in u wind (m s^{-1}), temperature (K), and specific humidity (g kg^{-1}) at each pressure level for the Northern Hemisphere (20°–90°N), the tropics (20°N–20°S), and the Southern Hemisphere (20°–90°S). The dashed gray line indicates zero RMSE.

property is also valid for this higher-dimensional and realistic GFS model.

In Fig. 10, we show the relative error reduction of forecasts valid from 0–240 h and compare the improvements from rejecting the most detrimental 10%, 30%, and 50% of observations. For most regions and variables, the relative improvement was the largest within 12 h forecasts, ranging from 8%–25% and started decaying afterward. After 5 days, the improvement remained at about 3%–14%, and not until after 10 days did the forecast improvement vanish. Consistent with the previous results in CK19 using Lorenz (1996) model, the majority of improvement came from rejecting the most detrimental 10% of observations, and the improvement seemed to saturate at rejecting about the most

detrimental 30%–50% of observations (PQC-50 did not provide too much additional improvement over PQC-30). This result dispels the speculation that the denial of an excessive number of the detrimental observations may be harmful, by showing that the forecast was improved rather than degraded when essentially all of the detrimental observations were denied.

e. Biases in analysis and forecast after PQC update

To ensure the analysis and forecast RMSEs are not reduced at the cost of increased bias, we examine the differences in the bias in analysis and forecast between the CNTL and the PQC experiments. The bias assessment shown here was calculated against the CFSR

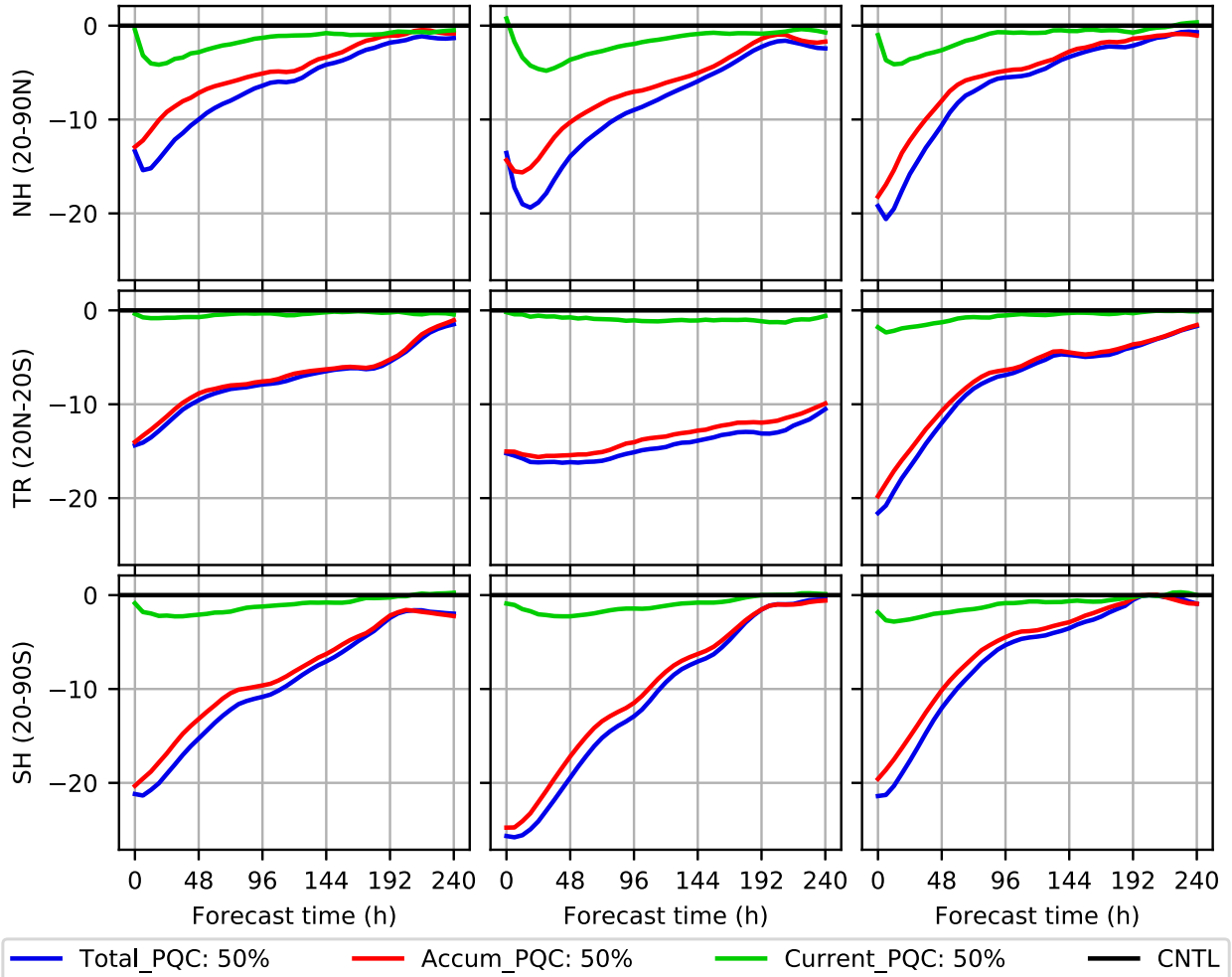## Relative RMSE changes by Accumulated and Current PQC update

FIG. 7. Monthly mean forecast relative RMSE changes (%) initiated from PQC_A (total update; blue), ANAL (accumulated update; red), and the difference between the two ("current" update; green) in *u* wind at 500 hPa, temperature at 500 hPa, and specific humidity at 700 hPa for the Northern Hemisphere (20°–90°N), the tropics (20°N–20°S), and the Southern Hemisphere (20°–90°S) throughout 10-day forecast for experiment PQC-50.

reanalysis. The vertical profile of bias in the analyses and 24-h forecasts is shown in Fig. 11. It is clear that the bias either did not change much or actually became smaller after applying the PQC-updates. The biases of the *u* wind in NH and SH, as well as the specific humidity in NH, did not change much, while the biases in other variables were strongly reduced by applying PQC. This was also true for the bias evolution throughout 10-day forecasts as shown in Fig. 12. The bias assessment against the radiosondes was also calculated and concluded the same that the biases were either unchanged or reduced by PQC (not shown). We note that PQC greatly reduced the tropical temperature bias, as could be expected given the almost

constant RMSE reduction with increasing forecast lead time in Fig. 10.

f. Case study: 0000 UTC 28 January 2008

In this section, we show a representative case from the PQC-30 experiment where the accumulated PQC-update improved the midlatitude 5-day forecast by recovering a developing baroclinic instability that was initiated at 0000 UTC 28 January 2008. The geopotential height at 500 hPa and the sea level pressure are shown in Figs. 13 and 14, respectively. In the figures, there were no clear differences between the TRUTH (CFSR; left column), PQC-30 (middle column), and CNTL (right column) at the analysis time. After 3 forecast days, a
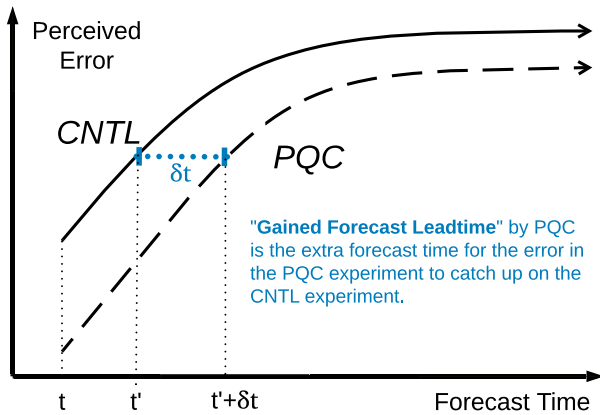
FIG. 8. Schematic of the gained forecast lead time. Solid curve shows the perceived error growth with forecast time for the control experiment; dashed curve shows that for the PQC experiment. The gained forecast lead time is defined as the δt such that RMSE_PQC$_{t'+\delta t|t} \approx$ RMSE_CNTL$_{t'|t}$, where t and t' is the forecast initial time and the CNTL forecast valid time.

baroclinic structure started to develop in TRUTH. On day 5, a deep trough had developed at 65°N, 0°, and a short-wave trough was located slightly west of 15°E. A

surface low pressure located at 65°N, 15°E is shown in the forecast on day 5 in Fig. 14. The system was in a developing stage as the surface low leading the deep trough at 500 hPa, and the features were intensifying over time. However, we do not observe these features in the CNTL (right column in Figs. 13 and 14). With slight modifications at the analysis time, the accumulated PQC-update recovers the intensifying of both the surface low pressure and the upper-level trough and significantly improves the forecast skill. The PQC improvement also intensified with the baroclinic system, showing that the improvement was associated with an actual dynamical instability. Note that no future information was used to provide this accumulated PQC-update.

5. Summary and discussion

This study provides the first demonstration of cycling PQC in a DA system with intermediate complexity using a low-resolution realistic spectral GFS model and a simple LETKF system for assimilating real observations from the PrepBUFR dataset to bridge the
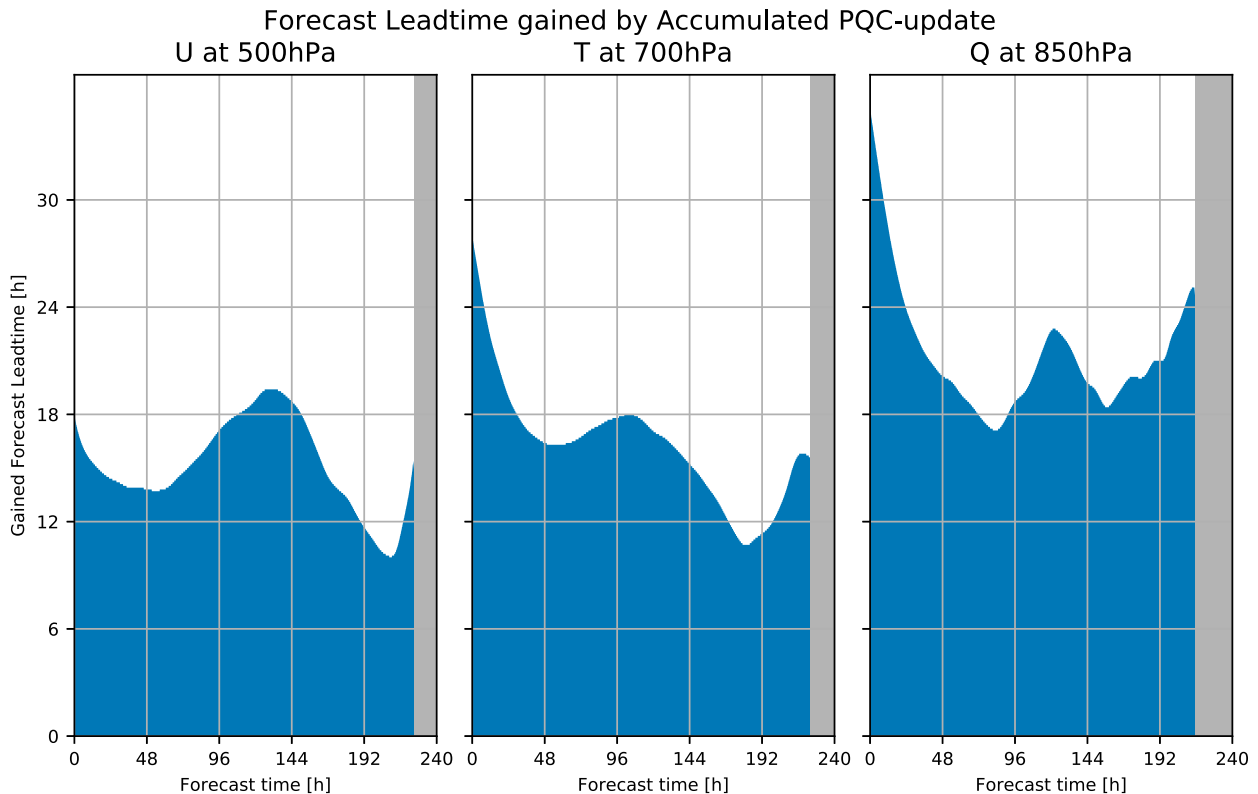

FIG. 9. Averaged gained forecast lead time (h) by accumulated PQC-update throughout 10-day forecast in u wind (m s^{-1}) at 500 hPa, temperature (K) at 700 hPa, and specific humidity (g kg^{-1}) at 850 hPa for experiment PQC-50. The gained forecast lead time shows the forecast time advantage for PQC forecast error to reach the same error level in CNTL. The gray area masks the period when adding the forecast time and the corresponding gained lead time exceeds the range of 10 days.
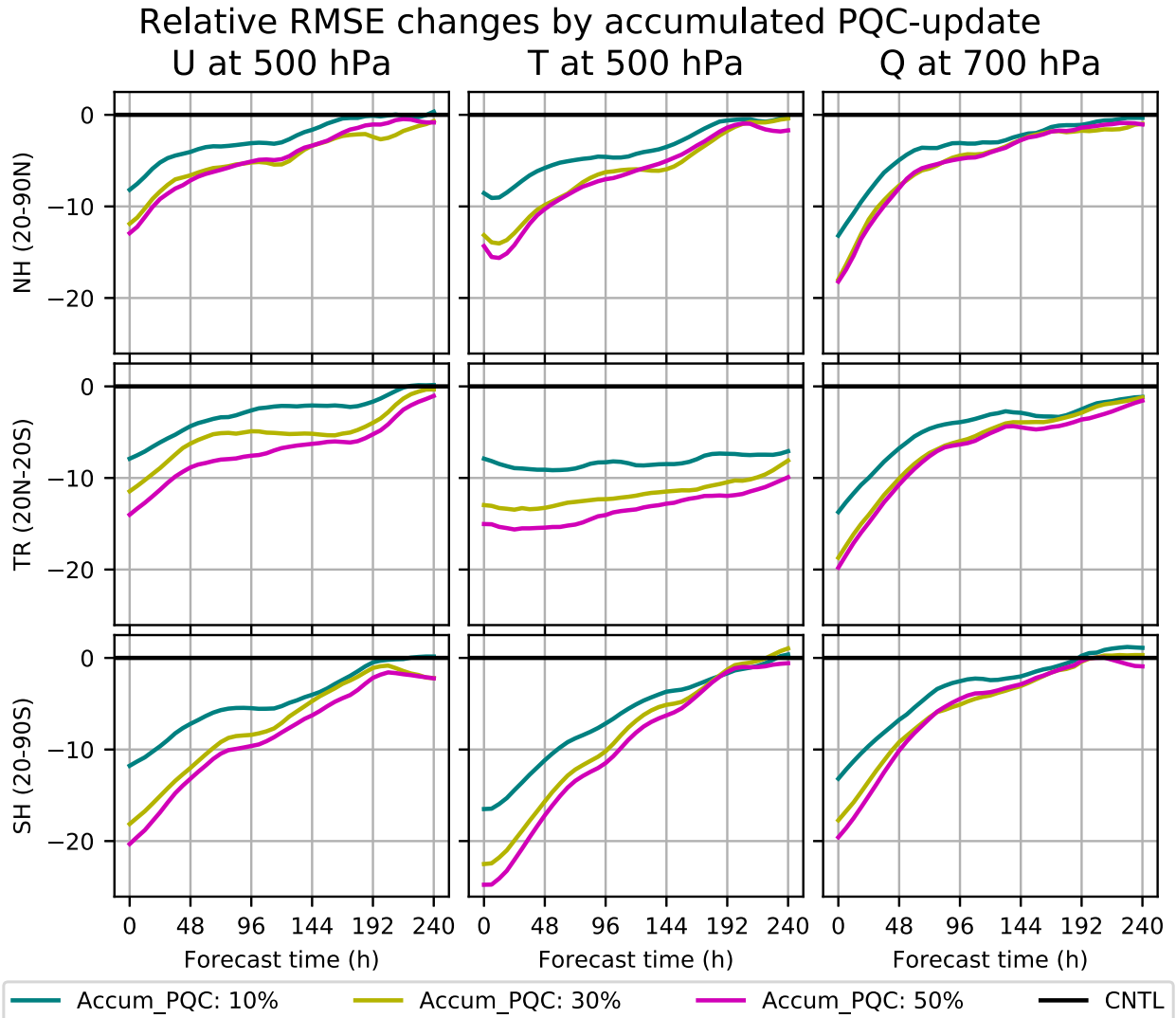
Relative RMSE changes by accumulated PQC-update

FIG. 10. Monthly mean forecast relative RMSE changes (%) in the *u* wind at 500 hPa, temperature at 500 hPa, and specific humidity at 700 hPa for the Northern Hemisphere (20°–90°N), the tropics (20°N–20°S), and the Southern Hemisphere (20°–90°S) throughout 10 days. The curves represent the improvement by rejecting overall 10% (cyan), 30% (yellow), and 50% (magenta) of observations with cycling PQC using 6-h EFSO lead time.

gap from the idealized setup in CK19 to an operational configuration.

We demonstrated throughout the globe monthly averaged improvement in the analyses and forecasts across variables including wind, temperature, and specific humidity throughout all pressure levels. The majority of the improvement, equivalent to more than 12 h of forecast lead came from the removal of the most detrimental 10% of the observations, while the PQC improvement was the largest in the experiments when rejecting the most detrimental 30%–50% of the observations, which is consistent with the results of CK19. And the forecast improvements from the accumulated PQC-updates lasted throughout the 10-day forecast. The accumulated

PQC-updates provided more than 12-h forecast lead-time gain in this study. In addition, the analysis and forecast biases were either unchanged or reduced by PQC. We also presented a case study showing how the accumulated PQC-update improved a 5-day forecast by recovering a developing baroclinic instability in the midlatitude that had been lost in the control experiment.

One of the key results of this study is identifying the dominant role of the accumulated PQC impact in comparison to the current update. It indicates that there was a long memory of PQC improvement, and hence that the current PQC-update, which requires future information, is not needed to make use of PQC. This is strong supportive evidence demonstrating the feasibility
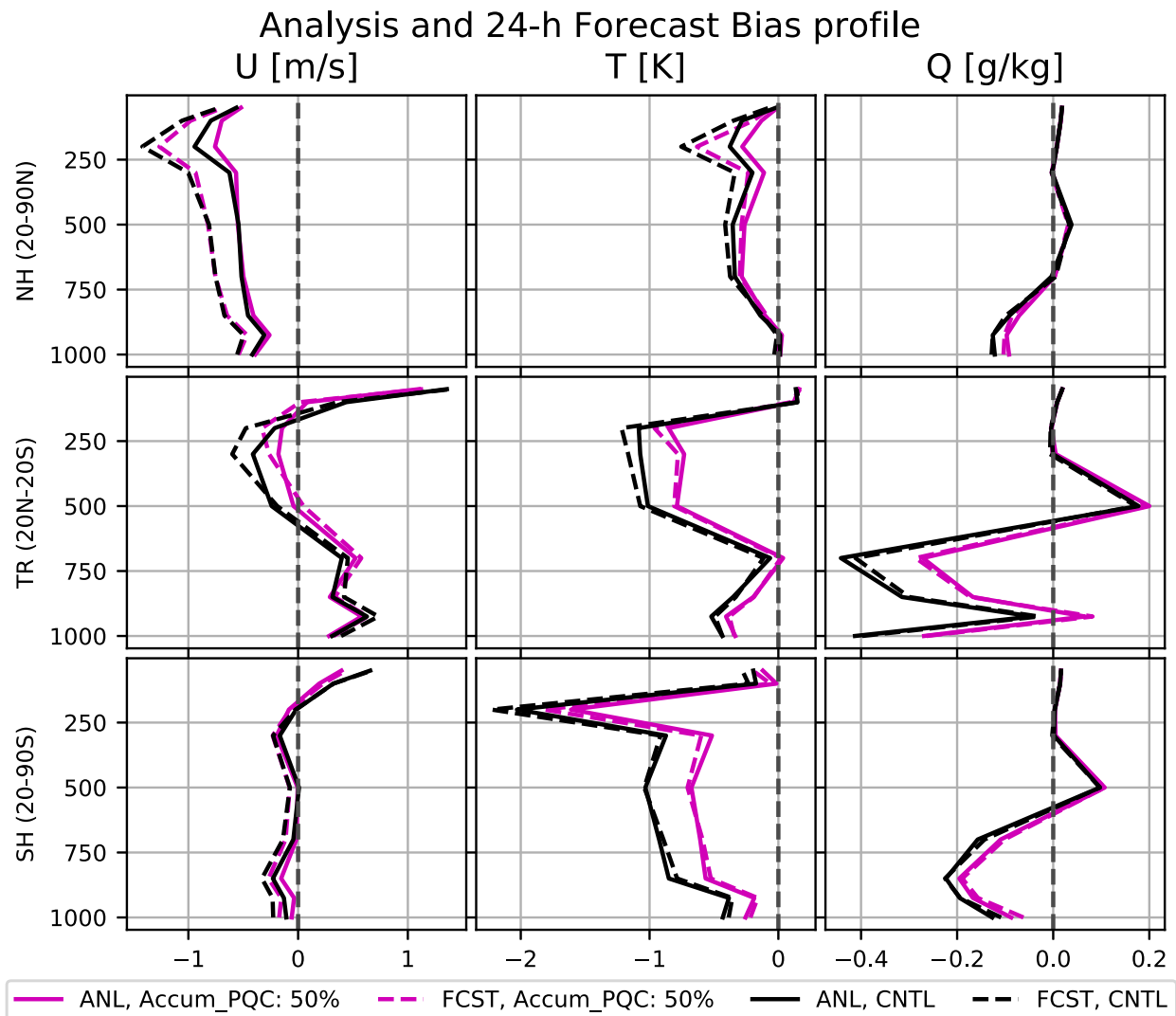
FIG. 11. Monthly mean bias profile at each pressure level of the analysis (solid) and 24-h forecast (dashed) of the accumulated PQC by rejecting 50% observations (magenta) and the CNTL (black) in u wind (m s^{-1}), temperature (K), and specific humidity (g kg^{-1}) for the Northern Hemisphere (20°–90°N), the tropics (20°N–20°S), and the Southern Hemisphere (20°–90°S). The dashed gray line indicates zero bias. The bias assessment was calculated against the CFSR reanalysis.

of PQC in operational implementation as proposed by Hotta et al. (2017).

In CK19, five PQC-update methods were proposed and compared. The direct data denial method (PQC_H) was found to be suboptimal due to the inconsistency between the improved accuracy of the PQC-updated analysis and the unintentionally inflated ensemble spread. This data denial method was nevertheless chosen in this paper because of its simple implementation, and because the implementation of RTPP and adaptive inflation alleviated this inconsistency. In future studies exploring an operational implementation, we would still recommend testing the efficient gain-reusing PQC_K approach.

We note that almost 50% of observations were detrimental on the average as reported by every operational center and research group. It is very hard to believe all these (E)FSO-identified detrimental observations were flawed, and the removal of the flawed observations is only one of the potential reasons for the PQC improvement. Another possible reason is the smoother aspect of PQC since the observations from the next cycle were used indirectly to update the analysis (PQC_A) as explained in CK19. In fact, CK19 pointed out that the removal of the most detrimental observations contributes the majority of the forecast improvement because the associated detrimental analysis increments from these observations were in the unstable growing
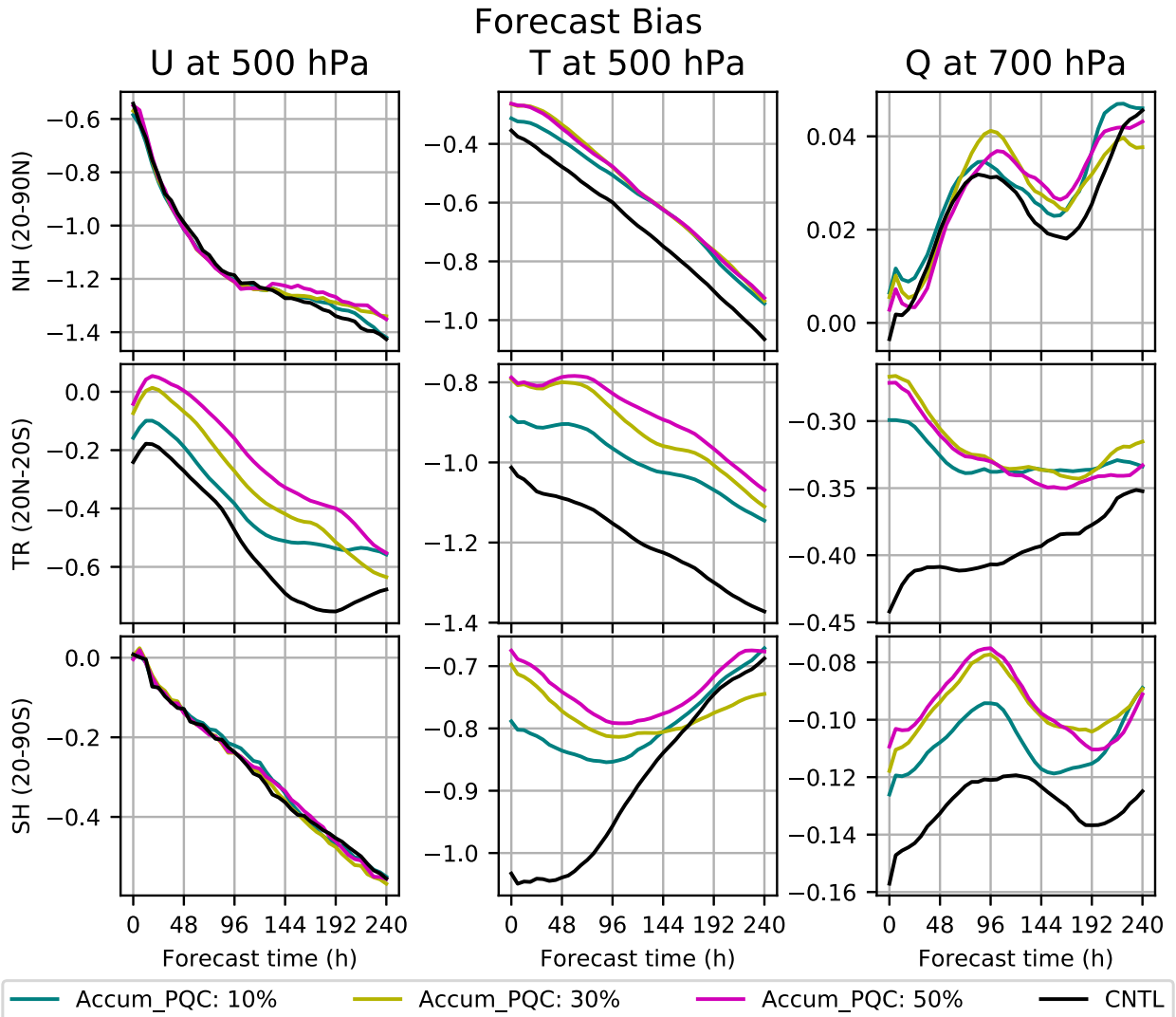
Forecast Bias

FIG. 12. Monthly mean bias of the accumulated PQC forecast throughout 10-day with rejecting overall 10% (cyan), 30% (yellow), and 50% (magenta) of observations in u wind (m s^{-1}), temperature (K), and specific humidity (g kg^{-1}) for the Northern Hemisphere (20°–90°N), the tropics (20°N–20°S), and the Southern Hemisphere (20°–90°S). The black curves represent the mean bias of the control experiment without PQC. The bias assessment was calculated against the CFSR reanalysis.

subspace (whereas those from the rest of the detrimental observations were not). While rejecting nearly 50% of the observations seems to be aggressive, we would like to point out that a portion of them did not provide effective analysis increments from the unstable growing perspective, which is also supported by the fact that PQC-50 did not provide much additional improvement compared with PQC-30. An alternative view of PQC is that we simply include the observations at t_n to perform a "buddy-check" for the observations at t_{n-1} by propagating the information through time using EFSO, and the detrimental observations can be viewed to be inconsistent with the observations from the subsequent cycle. Given the fact that EFSO/PQC is using

observations beyond the analysis time, it would be interesting to compare this approach with performing DA using a long window (e.g., 12 h) and examine the interaction between the two approaches when applied together (they are not mutually exclusive) to advance the fundamental understanding of EFSO and PQC.

In the experimental setup of this study, PQC requires performing two more analyses (i.e., REF_A and PQC_A) in addition to the standard analysis (ANAL) in each cycle. Hotta et al. (2017) outlined the implementation of PQC under the dual-track analysis framework of NCEP (see their Fig. 9). It was proposed that the PQC for GDAS final analysis at time t_n should take place right after the GFS early analysis valid at time t_{n+1} is done and

Geopotential Height at 500 hPa

(a) Contour: TRUTH
Color: Error Change by PQC

(b) Contour: PQC
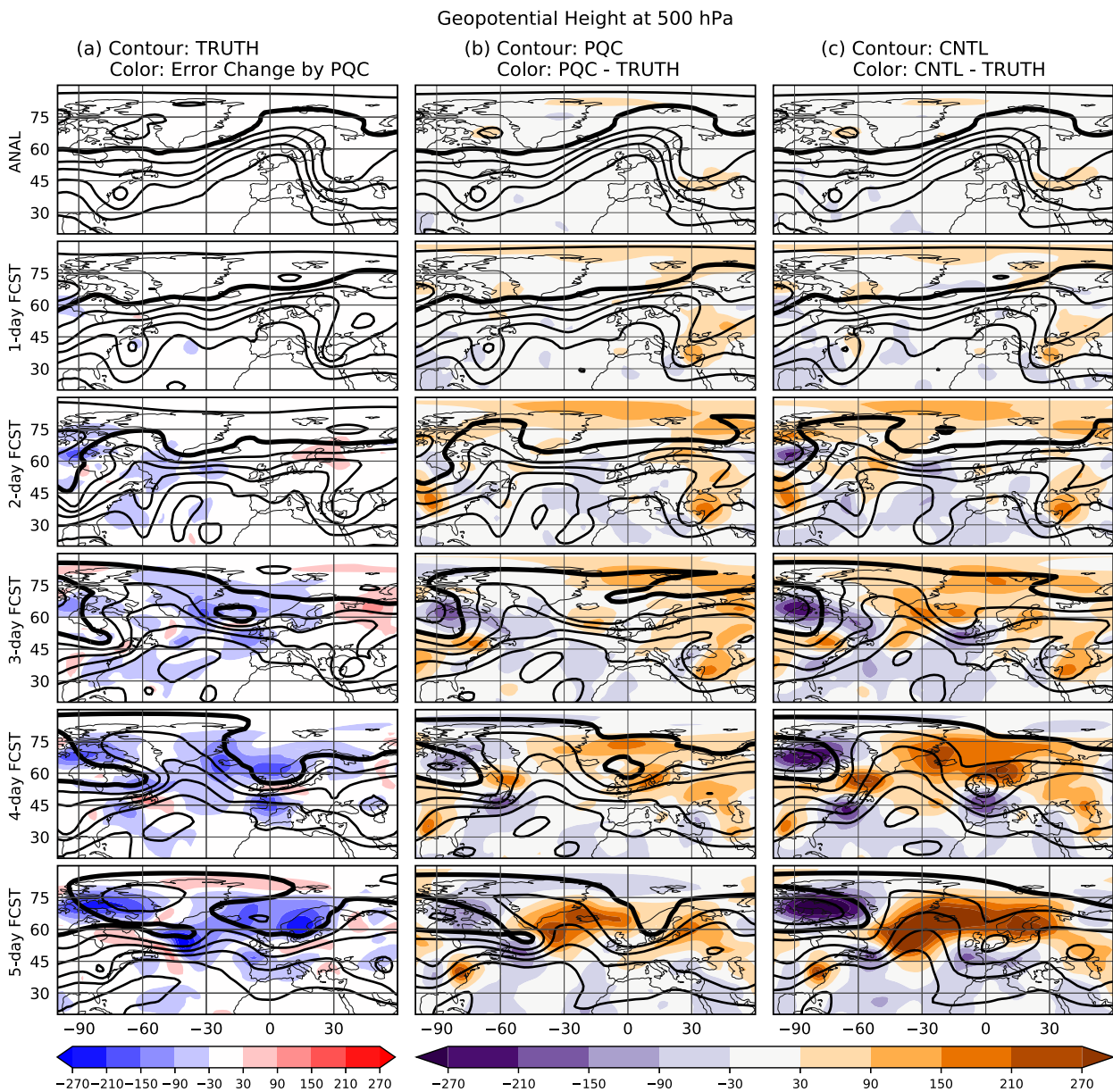Color: PQC − TRUTH

(c) Contour: CNTL
Color: CNTL − TRUTH

FIG. 13. The 5-day evolution of the geopotential height (m) forecast and forecast error changes at 500 hPa from one of the cycles (0000 UTC 28 Jan 2008) in the experimental period. (a) Geopotential height from verifying truth (CFSR) in contours and the forecast error changes from accumulated PQC-updates in colors. Blue colors represent forecast improvement; red colors represent forecast degradation. The thick lines highlight the contour of 5100 m. (b) Geopotential height modified by the accumulated PQC-updates (PQC-30) in contours and the difference between the modified fields and the verifying truth in colors. Orange colors represent an overestimation of heights; purple colors represent an underestimation of heights. (c) As in (b), but for the control geopotential height (CNTL).

use the early analysis as the EFSO verifying reference (REF_A) to reduce the computational cost of making an additional temporary analysis. This substitution would make PQC-update available to GDAS requiring only one additional analysis (PQC_A), but the trade-off would be that the PQC-update on GFS long forecasts will be delayed by two cycles (12 h), instead of the original one cycle. Also, the GFS early analysis ingests

fewer observations and is inevitably less accurate compared to the GDAS analysis though the difference is becoming smaller, and the preliminary results (Chen 2018) found that the correlation of EFSO impacts using GFS analysis and that using GDAS analysis was as high as 0.95. Such a two-cycle delay between the PQC-update and the initialization of the GFS forecast and using the less accurate early analysis as verifying
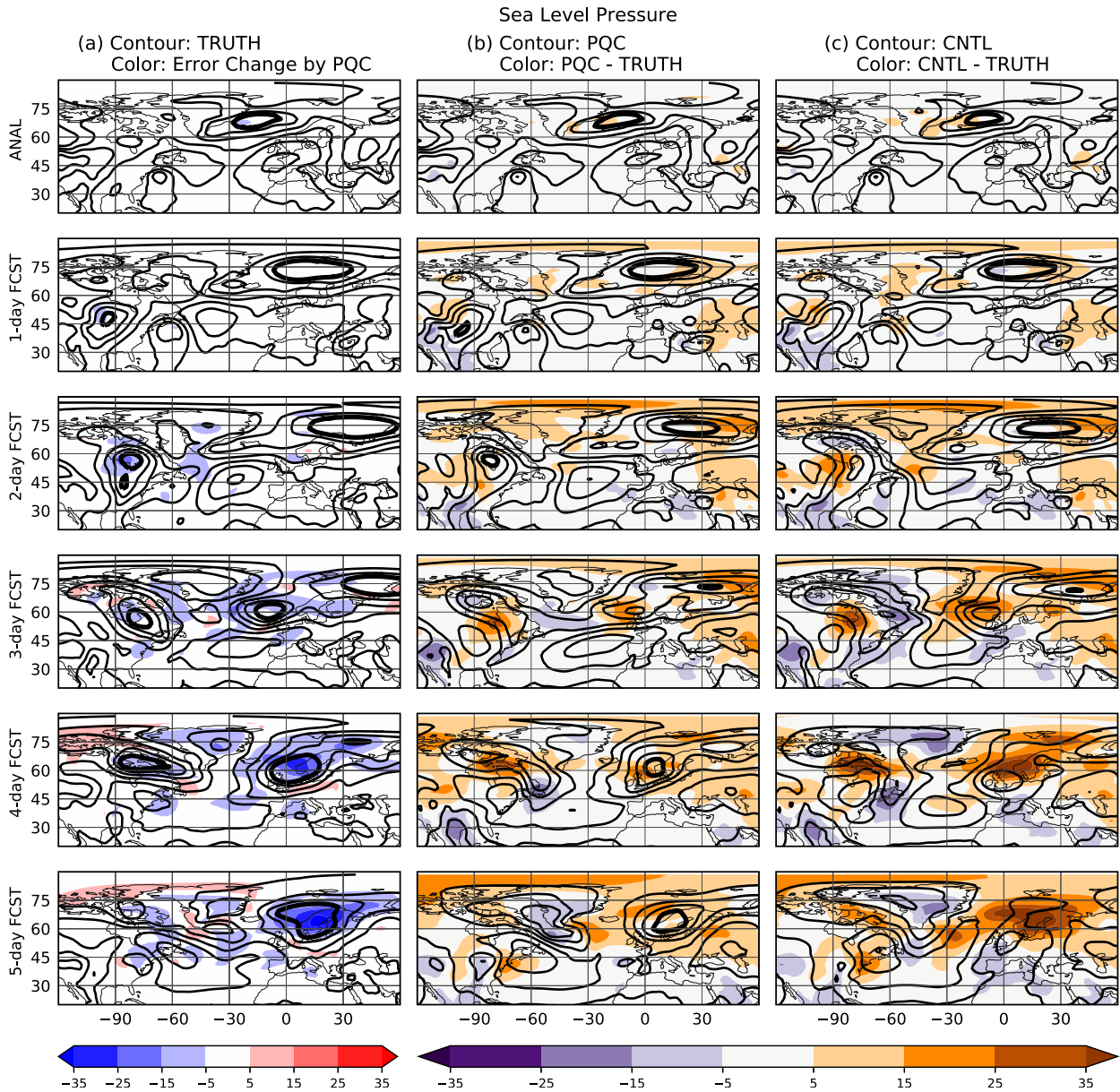
Sea Level Pressure

FIG. 14. As in Fig. 13, but for sea level pressure (hPa). The thick lines highlight the contour of 985 hPa.

reference may result in a smaller improvement when compared to that shown in this study. Ota et al. (2013) and Hotta et al. (2017) also formulated a more efficient method (PQC_K) for PQC-update to replace the direct data denial PQC_H method adopted in this study. The PQC_K method, proven to be both more accurate and cheaper in CK19 using Lorenz (1996) model, will avoid performing that additional data assimilation for PQC_A analysis.

For future directions, we point out some components worth further examination. First, we used PrepBUFR observations, and thus did not test EFSO on satellite radiance assimilation. Second, it is known that EFSO is dependent on the error norm. A past study (Hotta et al. 2017) showed slight differences between EFSO impact using dry energy norm and moist total energy norm. It is worth exploring this dependence of PQC performance. Third, CK19, using Lorenz (1996) model, showed that the PQC-updates using the original Kalman gain (PQC_K method) outperformed, with lower computational costs, the direct data-denial updates (PQC_H method) used in this study. The superiority of the cheaper and more accurate PQC_K update method should be tested in a high-dimensional global model.

Fourth, the Hybrid EnVar DA system is used in many operational centers, and the observations assimilated in the EnKF subsystem within are usually not identical to those assimilated in the main variational system. For this reason and the fact that it is more consistent with EnVar, the hybrid formulation of FSO (Buehner et al. 2018) may be preferable in such DA systems. It is thus worth comparing the PQC performance based on both HFSO and EFSO in an EnVar system.

In addition to operational NWP, PQC can also be applied to reanalysis applications, which have automatic access to the future observations and hence the current PQC-update. The future observations that extend outside of the DA window have never been used in current reanalyses that mimic operational DA systems. The future observations, by contrast, could be utilized by EFSO/PQC with the potential to significantly improve the accuracy of future reanalyses, and this advantage should be properly exploited.

To summarize, following the promising results of cycling PQC in CK19 using Lorenz (1996) model, we have shown that cycling PQC further improves the analyses and forecasts in a realistic NCEP spectral GFS model that assimilates real conventional observations.

REFERENCES

Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

Buehner, M., P. Du, and J. Bédard, 2018: A new approach for estimating the observation impact in ensemble–variational data assimilation. *Mon. Wea. Rev.*, **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. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, https://doi.org/10.1002/qj.366.

——, 2018: Forecast sensitivity observation impact with an observation-only based objective function. *Quart. J. Roy.*

Meteor. Soc., **144**, 2089–2098, https://doi.org/10.1002/qj.3305.

Chen, T.-C., 2018: Applications of ensemble forecast sensitivity to observations for improving numerical weather prediction. Ph.D. dissertation, University of Maryland, 180 pp.

——, and E. Kalnay, 2019: Proactive quality control: Observing system simulation experiments with the Lorenz '96 model. *Mon. Wea. Rev.*, **147**, 53–67, https://doi.org/10.1175/MWR-D-18-0138.1.

——, ——, and D. Hotta, 2017: Use of EFSO for online data assimilation quality monitoring and proactive quality control. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, No. 47, 9 pp., http://bluebook.meteoinfo.ru/uploads/2017/docs/01_Chen_Tse_Chun_EFSO,_Quality_Monitoring,_Proactive_QC.pdf.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Ehrendorfer, M., 2007: A review of issues in ensemble-based Kalman filtering. *Meteor. Z.*, **16**, 795–818, https://doi.org/10.1127/0941-2948/2007/0256.

Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX observation impact intercomparison experiment. *Mon. Wea. Rev.*, **138**, 4009–4025, https://doi.org/10.1175/2010MWR3393.1.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522, https://doi.org/10.1175/2010MWR3328.1.

Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017: Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Mon. Wea. Rev.*, **145**, 3331–3354, https://doi.org/10.1175/MWR-D-16-0290.1.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 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., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus*, **64A**, 18462, https://doi.org/10.3402/tellusa.v64i0.18462.

Kotsuki, S., S. J. Greybush, and T. Miyoshi, 2017: Can we optimize the assimilation order in the serial ensemble Kalman filter? A study with the Lorenz-96 model. *Mon. Wea. Rev.*, **145**, 4977–4995, https://doi.org/10.1175/MWR-D-17-0094.1.

——, K. Kurosawa, and T. Miyoshi, 2019: On the properties of ensemble forecast sensitivity to observations. *Quart. J. Roy. Meteor. Soc.*, **145**, 1897–1914, https://doi.org/10.1002/qj.3534.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, https://doi.org/10.3402/tellusa.v56i3.14413.

Lien, G.-Y., T. Miyoshi, and E. Kalnay, 2016: Assimilation of TRMM Multisatellite Precipitation Analysis with a low-resolution NCEP global forecast system. *Mon. Wea. Rev.*, **144**, 643–661, https://doi.org/10.1175/MWR-D-15-0149.1.

——, D. Hotta, E. Kalnay, T. Miyoshi, and T.-C. Chen, 2018: Accelerating assimilation development for new observing systems using EFSO. *Nonlinear Processes Geophys.*, **25**, 129–143, https://doi.org/10.5194/npg-25-129-2018.

Lorenc, A. C., and R. T. Marriott, 2014: Forecast sensitivity to observations in the Met Office global numerical weather

prediction system. *Quart. J. Roy. Meteor. Soc.*, **140**, 209–224, https://doi.org/10.1002/qj.2122.

Lorenz, E., 1996: Predicability—A problem partly solved. *Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, https://doi.org/10.1175/2010MWR3570.1.

Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quart. J. Roy. Meteor. Soc.*, **144**, 1667–1680, https://doi.org/10.1002/qj.3390.

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

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter.

Quart. J. Roy. Meteor. Soc., **140**, 2672–2679, https://doi.org/10.1002/qj.2343.

——, and ——, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus*, **68A**, 27885, https://doi.org/10.3402/tellusa.v68.27885.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Zhang, F., C. Snyder, J. Sun, F. Zhang, C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhu, Y., and R. Gelaro, 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Wea. Rev.*, **136**, 335–351, https://doi.org/10.1175/MWR3525.1.