ABSTRACT

Title of dissertation: APPLICATIONS OF ENSEMBLE FORECAST SENSITIVITY TO OBSERVATIONS FOR IMPROVING NUMERICAL WEATHER PREDICTION

Tse-Chun Chen, Doctor of Philosophy, 2018

Dissertation directed by: Professor Eugenia Kalnay
Department of Atmospheric and Oceanic Science

Massive amounts of observations are assimilated every day into modern Numerical Weather Prediction (NWP) systems, and more are being deployed. The large volume of data prevents thorough monitoring and screening (QC) the impact of each assimilated observation using standard observing system experiments (OSEs). The presence of so many observations also makes very difficult to estimate the impact of a new observing system using OSEs.

Forecast Sensitivity to Observation using adjoint formulation (AFSO, Langland and Baker, 2004) provides an efficient impact evaluation of each observation on forecasts. We propose 3 applications using the simpler ensemble formulation of FSO (EFSO, Kalnay et al., 2012) to improve NWP, namely (1) online monitoring tool, (2) data selection, and (3) proactive quality control (PQC).

We first demonstrate PQC on a simple Lorenz (1996) model, showing that EFSO is able to identify artificially ''flawed" observations. We then show that PQC improves the quality of analysis and forecast of the system, even if the observations

are flawless, and the improvement is robust against common sub-optimal of DA configurations in operation. A PQC update method reusing the original Kalman gain is found to be both accurate and computationally efficient.

EFSO and PQC are then explored with realistic GFS systems. A close-to-operation GFS-GSI Hybrid En-Var system is used to examine the data monitoring and selection applications. The benefit of the online observation monitoring and data rejection based on EFSO is very apparent. Identifying and deleting detrimental radiance channels results in a forecast improvement. Results obtained on a lower resolution GFS system show that PQC significantly improves the quality of analysis and 5-day forecasts for all variables over the globe. Most of the improvement comes from "cycling" PQC, which accumulates improvements brought by deleting detrimental observations over many cycles, rather than from deleting detrimental observations in the current cycle. Thus we avoid using "future data" in PQC and its implementation is shown to be computationally feasible in NCEP operations.

APPLICATIONS OF ENSEMBLE FORECAST SENSITIVITY TO OBSERVATIONS FOR IMPROVING NUMERICAL WEATHER PREDICTION

by

Tse-Chun Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Professor Brian Hunt
Professor Daryl Kleist
Professor James Carton
Professor Stephen Penny

© Copyright by
Tse-Chun Chen
2018

Dedication

To my father Wen-I, mother Hsueh-Chiao, brother Li-Chun, grandfather Shuei-Mu, and grandmother Chien-Sung.

Acknowledgments

There are many people I would like to thank that took part in this journey.

First, I would like to thank my advisor, Prof. Eugenia Kalnay, for guidance, mentoring, and encouragement. In every meetings and conversation we have had, Eugenia is always trusting with my capability and open to even the craziest idea I brought up. It is fair to say I couldn't have achieved what I have done without the positive atmosphere she created. More importantly, she set up an example of a great scientist. I learn from her fearless and proactive spirits, willingness to challenge the doctrines of the field, and take action when observing the need.

I am deeply grateful to the committee members, Profs. Brian Hunt, Daryl Kleist, Steve Penny, and James Carton, for constructive comments and suggestions. Special thanks to Steve on the inspiring conversations and discussions in and out of AOSC658E class. I would like to thank also Prof. Kayo Ide for many conversations, discussion in AOSC615 class, and organizing informative Weather-Chaos events providing opportunity to meet with many world renowned scholars. Thanks to my mentors in NTU, Profs. Hung-Chi Kuo, Jen-Ping Chen, and Chun-Chieh Wu for constant encouragement and inspiration.

The study would not be possible without the great support from JCSDA for accessing the S4 computer. Additionally, I am extremely grateful to the support from JPSS and Dr. Mitch Goldberg, and to Dr. James Jung for guidance and assistance on running GFS-GSI on S4. Drs. Daisuke Hotta and Guo-Yuan Lien provided fruitful discussions, guidance, and advises both scientifically and technically, espe-

cially on EFSO, GFS-GSI, and GFS-LETKF codes.

I would like to thank Drs. Krishna Kumar and Jordan Alpert for many discussions, assistance, and for collaborating on the PQC project. Also, I would like to thank Drs. Sean Casey, Lidia Cucurull, and Bob Atlas for the opportunity to collaborate on combining EFSO and OSSE.

I am grateful to all the guidance provided by Kathryn Shontz for discussions and proposal writing. Additionally, I am very lucky to have great colleagues and friends, Cheng, Takuma, Chu-Chun, Kriti, Eviatar, and Luyu for countless conversations and encouragements. Thanks to Adria for constant encouragement.

Also, I am thankful to Chia-Wei, Ho-Hsuan, Yuan-Ming, Shun-Nan, Wen-Hou, Yi, and Amy for accompanying some of my down times and listen to my complaint and murmurs, and my (ex-)roommates Wan-Chen and Tung-Chang for all the fun nights with animations and junk foods.

Finally and most importantly, I would like to express my deepest gratitude to my girlfriend Bor-Ting for being there with me through the happiest and hardest (my grumpiest) moments. Without her support, it would be ten times harder to achieve this milestone.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3D-Var	Three-Dimensional Variational data assimilation system
AI	Analysis Increment
AIRS	Advanced Infrared Sounder
AFSO	Adjoint Forecast Sensitivity to Observation
ATMS	Advanced Technology Microwave Sounder
CFSR	Climate Forecast System Reanalysis
CrIS	Cross-track Infrared Sounder
DA	Data Assimilation
DFS	Degrees of Freedom of Signal
ECMWF	European Centre for Medium-range Weather Forecast
EnKF	Ensemble Kalman Filter
EnSRF	Ensemble Square Root Filter
EFSO	Ensemble Forecast Sensitivity to Observation
EFSR	Ensemble Forecast Sensitivity to $\boldsymbol{R}$
ETKF	Ensemble Transform Kalman Filter
FSO	Forecast Sensitivity to Observation
GDAS	Global Data Assimilation System
GOES	Geostationary Operational Environmental Satellite
GFDPT	Global Forecasting Dropout Prediction Tool
GFS	Global Forecasting System (Global NWP system from NCEP)
GPSRO	Global Positioning System Radio Occultation
GSI	Gridpoint Statistical Interpolation
HIRS	High Resolution Infrared Radiation Sounder
HFSO	Hybrid Forecast Sensitivity to Observation
IASI	Infrared Atmospheric Sounding Interferometer
JCSDA	Joint Center for Satellite Data Assimilation
JMA	Japan Meteorological Agency
LETKF	Local Ensemble Transform Kalman Filter
MODIS	Moderate-Resolution Imaging Spectroradiometer
NASA	National Centers for Environmental Prediction
NH	Northern Hemisphere
NWP	Numerical Weather Prediction
OSE	Observing System Experiment
OSSE	Observing System Simulation Experiment
PIBAL	Pilot Balloon
PQC	Proactive Quality Control
QC	Quality Control
RMSE	Root Mean Square Error
S4	Supercomputer for Satellite Simulations and Data Assimilation Studies
SH	Southern Hemisphere
TR	Tropics
TRMM	Tropical Rainfall Measurement Mission
VIIRS	Visible Infrared Imaging Radiometer Suite
QOSAP	Quantitative Observing System Assessment Program

Chapter 1

Introduction

1.1 Background: Observing the Atmosphere

Since the start of the studying of the atmosphere, the observations have been the backbone of the field. Breakthroughs in the understanding of the atmosphere can be attributed (at least partially) to the advancement of the observing technology and data availability.

The first thermometer was generally credited to Galileo in around the year 1600 but did not become practical until 1700 when Fahrenheit invented the accurate mercury-based thermometer. Around the 1640s, pressure measurements were enabled by the barometers invented by Torricelli and Pascal. Wind observations were also recorded in support of sailing. Early observations were only available on the surface and limited to local regions until the development of the telegraph by Morse in around 1838. By the advancement of the communication technology, we can construct synoptic weather analysis for the first time. So a telegraphic network of observing stations was organized to "predict" storms for the downwind regions. In around 1900, we began to use balloons on measuring upper atmosphere conditions. The beginning of aviation brought an increasing need of upper air measurements, and the aircraft was started to be used as a platform for weather observations. The high-tension of military competition from World War II to the Cold War era boosted

the development of remote sensing technology including radar and satellite. These technological advancements extended our limited in situ observations beyond the horizon, and the number of observations underwent unprecedented bloom.

Allowed by these rapidly increasing number and abundance of observations, our knowledge of the atmosphere expands at a rate that it never did before. Together with the development of one of the first computers and the mathematical descriptions of the governing laws of the atmosphere, Charney et al. (1950) introduced the dawn of the numerical weather prediction (NWP). Under the international collaborative effort, nations assembled a global observing network sharing valuable routine in situ and remote sensing weather observations enabling NWP on the global scale. The initial conditions for the models were manually analyzed from the observations in the early days. As the advancement of the model and the overwhelming growth rate of the number of observations, this laborious task was eventually replaced and automated by a technique called data assimilation (DA), in which the model "assimilates" observations when the observations are combined with the previous model forecast to obtain the "best-estimate" state of the atmosphere.

The advancement of numerical weather prediction has generally been attributed to the improvement of the model, the advancement of data assimilation, and the increasing number and quality of the observations (e.g. Magnusson and Källén, 2013). The increase of the observations, especially from satellites, provides significant contributions to the increase of forecast skill. This contribution can be quantified by examining the forecast quality improvement of a reanalysis product such as ERA-interim dataset (Dee et al., 2011), which uses a frozen data assimilation and model

system. By eliminating the two also constantly improving factors, the advancement over time in forecast quality initiated from the reanalysis is purely from the changes of observations regarding quality and number (Magnusson and Källén, 2013). Supportive evidence is also shown with the rapid growing forecast skill over the Southern Hemisphere that almost caught up with that over the Northern Hemisphere (Figure 1 in Bauer et al., 2015). As of today, millions or even close to ten millions of observations are assimilated every 6 hours in NWP operational centers around the world (see NCEP observation counts at http://www.nco.ncep.noaa.gov/sib/counts/)

Due to the advancement in the observing technology, we are experiencing an unprecedented increase in the number of observations, mainly from hyperspectral satellite radiance instruments and satellite-derived atmospheric motion vectors. As we launch more new instruments with higher spatial, temporal, and spectral resolutions and run more accurate models with higher resolution and better representation of the atmospheric processes, we are entering the "Big Data Assimilation" era described in Miyoshi et al. (2016). It poses great technical and scientific challenges on "how to effectively make use of the massive amount of observations?".

1.2 Forecast Sensitivity to Observations

The effectiveness of data selection and quality control before DA is one of the key components towards better harvesting observation information in NWP (described in Chapter 3). To better perform the data selection process, observational impact evaluation is inevitable, and forecast sensitivity to observation (FSO) diag-

nostic can be instrumentally informative.

1.2.1 Overview

Addressing the need for impact evaluation of the massive amount of observations, generic forecast sensitivity to observations (generic FSO) is a computationally efficient diagnostic tool that maps the future forecast error changes to the assimilated individual observations. The mapping is constructed according to the gain matrix of analysis equation (the influence of each observation on the analysis, $\mathbf{x}^a = \mathbf{x}^b + \boldsymbol{K}\delta\mathbf{y}^o$) and the forecast sensitivity to analysis (error growth). The approaches for the computation of these two components can be differentiated into adjoint-based, ensemble-based, and hybrid-based methods as in standard DA classification. We will describe the different approaches in detail in Chapter 2. In this study, we adopted the ensemble-based method (EFSO) for its simplicity in implementations. This mapping enables an efficient and detail estimation of the impact of each observation on the future forecast. The efficiency and the granularity of the FSO diagnostic allow several immediate applications that could improve the NWP from the observation side.

1.2.2 Applications

In this study, we propose three FSO applications for improving NWP from the observation side. First, the computationally efficient FSO could be computed at least near real-time as an observation monitoring tool alongside with other existing

QC tools. This monitoring tool alerts the front-line forecaster to be aware of the up-coming degraded forecasts from currently assimilated detrimental observations, and it also allows the operation centers and the observation developer to improve the quality and the DA treatment of the observations. Second, FSO statistics should be included in data selection decisions since it provides efficient and detail information on the impact that was not available in the past. Taking FSO impact into account could greatly shorten the required time and improve the accuracy of the data selection process. Finally, Proactive Quality Control (PQC), a powerful fully flow-dependent QC scheme based on immediate FSO impact in each cycle, was proposed to improve the quality of NWP, but only non-cycling PQC was tested. We further explore the strategies of the cycling PQC and its feasibility in NCEP operational system.

1.3 Objectives

We propose to use FSO impact evaluation technique for coping with the massive amount of observations and improving the performance of NWP. More specifically, we would like to demonstrate the benefit in NWP of the three FSO-based applications, namely the (1) Data monitoring and selection, and (2) Proactive QC. Throughout the dissertation, we aim at exploring the issues to-be-addressed listed below and by the end showing evidence that the three applications improve the quality of NWP indeed.

- EFSO (Observation monitoring tool and data selection)

  - Why are there 50% of detrimental observations?
  - What are reasonable choices of the forecast verifying truth?
  - Is 6-hour verifying lead-time appropriate for the applications?
  - Does EFSO-based data selection improve the forecast?

- Proactive QC
  - What is the benefit of cycling PQC versus non-cycling PQC?
  - How to PQC the detrimental observations and update the correction?
  - Should we perform PQC with longer verifying lead-time?
  - How to make PQC feasible in operation?

Here we list out the questions to-be-answered as a guideline. Detailed discussions on the questions above and the experimental design will be provided in Chapter 3.

1.4 Outline

The dissertation is structured as the following. A review of FSO background, including the formulations of different approaches and all the related topics, is provided in Chapter 2. Chapter 3 introduces the proposed three FSO applications for improving NWP that includes the background of the applications and the questions we need to address. We first examine the applications in the simple Lorenz (1996) system as a proof of concept in Chapter 4. Chapter 5 and 6 present the results from

the realistic GFS model as evidence that the proposed applications could indeed improve NWP in the operational environment. The data monitoring and selection applications are shown in Chapter 5 using a close-to-operation system. Chapter 6 shows the results of PQC in realistic GFS model and validates the feasibility in operation. The summary and future directions are offered in Chapter 7.

Chapter 2

Generic Forecast Sensitivity to Observations: Review

The generic Forecast Sensitivity to Observations (FSO) technique and the Ensemble-based FSO (EFSO) in particular is the foundation of this study. This chapter is dedicated to briefly review the historical development of different approaches in the generic FSO family, followed by the formulation of each approach. The different choices for the verifying "truth" will also be reviewed. The relation between EFSO and degrees of freedom of signal (DFS) in ensemble formulation, a widely used method for radiance channel selection, is discussed. This relation will serve as the foundation of the data selection we perform in the results of Chapter 5.

2.1 Introduction

The main objective of the generic FSO is to provide efficient and detailed observational impact evaluation. Currently, observing system experiments (OSEs) is the prevailing approach for directly estimating the impact of a certain set of observations by examining the 5-day forecast skill score differences of the lengthy experiments spanning for months with and without such observation set (e.g. Bauer, 2009; Dumelow, 2002; Riishojgaard and Redder, 2008; Bi et al., 2011). The major drawback of the method is the high computational demand which in turn limits both its discernibility and efficiency. Evaluating the bulk impact of one specific

observation type requires at least one experiment in addition to the control, leading to low efficiency and high computational costs. Additionally, the length of the experiments depends on the magnitude of the impact to reach statistically significant conclusions. Observation sets with smaller impact require longer than average experimental periods. Geer (2016) demonstrated that 60 forecast samples are required to reach the significance of 95% confidence level for a 1% improvement in the 5-day forecast. This low discernibility is especially severe in the case where large volumes of observations are already present in the system. Hence, the high-cost of OSEs limits the applications to distinguish the detailed impact attribution within the massive amount of observations. This problem will become worse given the increasing trend of the observation counts.

To address these issues, Langland and Baker (2004) devised the Adjoint-based Forecast Sensitivity to Observations (AFSO) method that estimates *the impact of every single observation* onto the forecast at once by attributing forecast error changes back to each observation with the adjoint model, using a future analysis as the verifying truth. AFSO provides an efficient estimation of the impact of each observation on the quality of the forecast. However, one major drawback of the AFSO approach is the use of the adjoint model which is notoriously tricky both to develop and to maintain for keeping up with the updates in the nonlinear model, which also requires a considerable amount of resources. Comprehensive descriptions of physical processes are of critical importance in the accuracy of the propagation of sensitivity. It has been found that the diabatic component of AFSO impacts associated with moisture cannot be appropriately estimated without a complete representa-

tion of moist physical processes in the adjoint of the linearized model (Janisko and Cardinali, 2016). However, those are not accounted in the adjoint model used in most AFSO studies (e.g. Langland and Baker, 2004; Zhu and Gelaro, 2008) since the treatment of nonlinearities and discontinuities of the moist physical processes is very complicated.

An alternative approach, Ensemble-based FSO (EFSO; Liu and Kalnay, 2008; Li et al., 2010), for the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007) was developed using the ensemble forecasts and analyses instead of the adjoint model. Later a simpler and more general EFSO formulation was introduced in Kalnay et al. (2012), that can be applied to any form of Ensemble Kalman Filter (EnKF). The adjoint-free approach then naturally does not suffer from the issues related to the adjoint model, but localization in computing EFSO is needed to address sampling errors associated with insufficient ensemble size that is typical in general meteorological ensemble applications.

Recently, a hybrid approach was formulated (HFSO; Buehner et al., 2018) that is more consistent with ensemble-variational (EnVar) DA system. HFSO uses the ensemble forecasts to propagate the forecast error changes in time, as in EFSO. However, the minimization of the cost function is used in the attribution of forecast error changes to each observation, like in AFSO, instead of using the Kalman gain matrix as in EFSO. A benefit of using HFSO over EFSO in an operational EnVar DA system is that HFSO has a better representation of observational impact since the observation set assimilated in the variational component and the ensemble component are usually different both in assimilated types and number of observations.

Alongside with the observations, there are other sources of differences between the two components including the background error, variational bias correction, etc.

It should be noted that all three approaches have their advantages and drawbacks. The choice of the method should be in accordance with the DA system and the applications. Moreover, it is worth mentioning that the proposed applications in this study are not limited to EFSO and can be generalized to the other two approaches.

Note that the generic FSO technique is not a complete replacement of OSEs. The two techniques provide complementary information. The generic FSO computes the estimation using the original analysis gain matrix (as if all observations are present) and thus, can be viewed as a linear and non-cycling approximation of OSEs. In contrast, OSEs changes the gain by altering the assimilated set of observations and the background trajectory due to cycling DA. While OSEs provides the actual forecast response of denying a certain subset of observations with high computational costs, the generic FSO offers efficient and economical detailed estimation of each observational impact, before the forecasts mix the impact of all the observations.

2.2 Formulation of generic Forecast Sensitivity to Observations (FSO)

In this section, the general formulation for generic FSO will be derived, followed by the specific procedure in computing for AFSO, EFSO, and HFSO impacts. The notation below loosely follows Kalnay et al. (2012).

2.2.1 Generic FSO formulation

The concept behind generic FSO technique is to construct a mapping between the observations and the future forecast error changes due to the assimilation of those observations. It is natural to begin the derivation with the definition of the forecast errors changes.

The forecast errors for the same verification time t from 0 and from -6 hour analyses are respectively denoted as :

$$\mathbf{e}_{t|0} = \mathbf{x}^f_{t|0} - \mathbf{x}^v_t, \tag{2.1}$$

$$\mathbf{e}_{t|-6} = \mathbf{x}^f_{t|-6} - \mathbf{x}^v_t \tag{2.2}$$

where the $\mathbf{x}^f_{t|0}$ and $\mathbf{x}^f_{t|-6}$ represent the forecasts valid at time t initiated from time 0 and -6, respectively, and $\mathbf{x}^v_t$ is the verifying truth for the forecast valid at time t. In practice, the best option available for $\mathbf{x}^v_t$ in real-time is the analysis at time t, namely $\mathbf{x}^a_t$. Other choices of verifying truth will be discussed later in this Chapter.

The only difference between the two forecast errors is introduced by the data assimilation at time 0 and the forecast error changes can be measured with:

$$\begin{aligned}
\Delta e^2 &= \mathbf{e}^T_{t|0}\boldsymbol{C}\mathbf{e}_{t|0} - \mathbf{e}^T_{t|-6}\boldsymbol{C}\mathbf{e}_{t|-6} \\
&= (\mathbf{e}_{t|0} - \mathbf{e}_{t|-6})^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\
&= (\mathbf{x}^f_{t|0} - \mathbf{x}^f_{t|-6})^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) && (2.3) \\
&= (M(\mathbf{x}^a_0) - M(\mathbf{x}^b_{0|-6}))^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) && (2.4) \\
&\approx [\boldsymbol{M}(\mathbf{x}^a_0 - \mathbf{x}^b_{0|-6})]^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) && (2.5)
\end{aligned}$$

(linearization of model forecast)

where $\boldsymbol{C}$ is the chosen error norm matrix, $M(\mathbf{x})$ denotes integration of the nonlinear forecast model, and $\boldsymbol{M}$ and $\boldsymbol{M}^T$ are the tangent linear approximation and corresponding adjoint of the nonlinear forecast model $M(x)$.

By rewriting the general form of the analysis equation, we can see that taking the derivative with respect to the observational innovation $\delta\mathbf{y}_0^o$ on both sides of the general form of the analysis equation yields the sensitivity of analysis increment to observation innovation:

$$\begin{aligned}\mathbf{x}_0^a - \mathbf{x}_{0|-6}^b = \delta\mathbf{x}_0^a &= \frac{\partial\mathbf{x}_0^a}{\partial\mathbf{y}^T}\delta\mathbf{y}_0^o \\ &= \boldsymbol{K}\delta\mathbf{y}_0^o \end{aligned} \tag{2.6}$$

where $\delta\mathbf{x}_0^a$ denotes the analysis increment and the gain matrix $\boldsymbol{K} = \frac{\partial\mathbf{x}_0^a}{\partial\mathbf{y}^T}$ is the analysis sensitivity to observation that determines in general the weighting of observation information going into the final analysis. Note that each DA approach associates $\delta\mathbf{x}_o^a$ and $\delta\mathbf{y}_0^o$ differently.

Applying chain rules to decompose the forecast error changes and substituting

in equation 2.6 gives:

$$\Delta e^2 \approx \delta(\boldsymbol{M}\mathbf{x}_{\mathbf{0}}^{\mathbf{a}})^T \frac{\partial(\Delta e^2)}{\partial(\boldsymbol{M}\mathbf{x}_0^a)} \tag{2.7}$$

(linearization of forecast sensitivity)

$$= \delta\mathbf{x}_0^{aT}\boldsymbol{M}^T \frac{\partial(\Delta e^2)}{\partial(\boldsymbol{M}\mathbf{x}_0^a)} \tag{2.8}$$

$$\approx (\delta\mathbf{y}_0^o)^T \frac{\partial\mathbf{x}_0^{aT}}{\partial\mathbf{y}}\boldsymbol{M}^T \frac{\partial(\Delta e^2)}{\partial(\boldsymbol{M}\mathbf{x}_0^a)} \tag{2.9}$$

(linearization of analysis sensitivity)

$$= (\delta\mathbf{y}_0^o)^T\boldsymbol{K}^T\boldsymbol{M}^T \frac{\partial(\Delta e^2)}{\partial(\boldsymbol{M}\mathbf{x}_0^a)} \tag{2.10}$$

$$= (\delta\mathbf{y}_0^o)^T \frac{\partial(\Delta e^2)}{\partial\mathbf{y}} \tag{2.11}$$

where the $\frac{\partial(\Delta e^2)}{\partial(\boldsymbol{M}\mathbf{x}_0^a)} = \boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$ is the sensitivity of the forecast error changes to the forecast differences $\boldsymbol{M}\delta\mathbf{x}_0^a$ at time t, and $\frac{\partial(\Delta e^2)}{\partial\mathbf{y}} = \boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{C}(\mathbf{e}_{t|0}+\mathbf{e}_{t|-6})$, obtained by comparing equations 2.5 and 2.10, denotes the sensitivity of forecast error changes to the observation innovation. Equation 2.10 is the generic form of FSO showing that the forecast error changes can be decomposed into four factors: observation innovation, sensitivity of analysis increment to observation innovation (gain matrix), adjoint model, and forecast errors valid at time t. And the major differences between AFSO, EFSO, and HFSO are the different approaches in evaluating the last three factors, which will be further explained in the following subsections. Finally, the product of the last three factors can be viewed as the sensitivity of forecast error changes to the observation innovation in equation 2.11.

2.2.2 Adjoint FSO

The AFSO (Langland and Baker, 2004) was developed under the context of a variational DA system. Following the formulation in Lorenc and Marriott (2014), the sensitivity of forecast error changes to the observation innovation $\frac{\partial(\Delta e^2)}{\partial \mathbf{y}} = \boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$ is evaluated as a whole. The forecast sensitivity is projected backward to observation time using the adjoint model. Consistent to the structure of variational DA system, $\boldsymbol{K}^T$ is never explicitly evaluated but the entire term $\frac{\partial(\Delta e^2)}{\partial \mathbf{y}}$ is evaluated through the minimization of a modified cost function. The solver used to solve for the minimization problem in the variational system can be utilized, but the cost function is modified so that it solves for the gradient of forecast error change to observational innovation. (see Lorenc and Marriott (2014) for additional details).

2.2.3 Ensemble FSO

In EFSO formulation (Kalnay et al., 2012), the evaluation of $\frac{\partial(\Delta e^2)}{\partial \mathbf{y}} = \boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{C}$ $(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$ is much simpler in practice compared to that in AFSO since the adjoint model $\boldsymbol{M}^T$ is replaced by ensemble forecasts and the adjoint of gain matrix $\mathbf{K}^T$ can also be estimated from the ensemble analysis perturbations. The gain matrix $\boldsymbol{K}$ in EnKF that accounts for the relative accuracies of the background and the observation is defined by the background error covariance matrix $\boldsymbol{B}$, observation error covariance $\boldsymbol{R}$, analysis error covariance $\boldsymbol{A}$, and linearized observational operator

$\boldsymbol{H}$ with

$$\begin{aligned}\boldsymbol{K} &= (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1} \\ &= \boldsymbol{A}\boldsymbol{H}^T\boldsymbol{R}^{-1} \end{aligned} \tag{2.12}$$

given that $\boldsymbol{A} = (\boldsymbol{B}^{-1}+\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}$. The analysis error covariance $\boldsymbol{A}$ can be approximated by the sampled covariance of analysis perturbation $\boldsymbol{X}_0^a$ provided by EnKF as $\boldsymbol{A} \approx \frac{1}{K-1}\boldsymbol{X}_0^a\boldsymbol{X}_0^{aT}$ where K is the ensemble size. Then equation 2.12 can be rewritten as

$$\boldsymbol{K} \approx \frac{1}{K-1}\boldsymbol{X}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{H}^T\boldsymbol{R}^{-1} \tag{2.13}$$

With equations 2.6 and 2.13, the approximation of analysis perturbation in observation space $\boldsymbol{Y}_0^a \approx \boldsymbol{H}\boldsymbol{X}_0^a$, and forecast perturbation $\boldsymbol{X}_{t|0}^f = \boldsymbol{M}\boldsymbol{X}_0^a$, equation 2.5 can be rewritten as

$$\begin{aligned}\Delta e^2 &\approx [\boldsymbol{M}\boldsymbol{K}\delta\mathbf{y}_0]^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\ &\approx \frac{1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{M}^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \qquad (2.14) \\ &\approx \frac{1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_{t|0}^{fT}\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \qquad (2.15)\end{aligned}$$

where $\boldsymbol{X}_{t|0}^f$ represents the background perturbation initiated from time 0 and valid at time t.

The impact of each observation can then be obtained by decomposing the sum of the inner product of the innovation vector $\delta\mathbf{y}_0^o$ and the sensitivity vector $\frac{1}{K-1}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_{t|0}^{fT}\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$ into elements that correspond to each observation, so that $\frac{1}{K-1}\delta y_{0,l}^T[\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_{t|0}^{fT}\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})]_l$ represents the estimated impact of the l-th observation on the forecast error changes.

Due to the limited number of ensemble members, localization of error covariance is needed in EFSO formulation to reduce sampling noise. The localized form of the EFSO formulation becomes

$$\Delta e^2 \approx \frac{1}{K-1}\delta \mathbf{y}_0^{oT}[\rho \circ \mathbf{R}^{-1}\mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT}]\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \tag{2.16}$$

where ρ is the localization matrix that is multiplied to the error covariance element-wise. The localization function initially applied in DA will propagate with the evolving flow as model forecast advances. Several methods were introduced in accounting for such propagation. As mentioned in Hotta (2014), the ideal solution should be to evolve the localization at the initial time by Kolmogorov (Fokker-Plank) equation associated with the model, which is computationally impractical. Estimation of non-linear evolution of the initial localization function and a simple advection with group velocity methods were proposed in the original Kalnay et al. (2012) for the simple Lorenz '96 model, but both methods are practically impossible for realistic systems. Ota et al. (2013) explored in NCEP GFS model a simple method which moves the center of the localization function with the horizontal wind with success in better capturing the diurnal cycle in total observational impact comparing to the fixed localization; i.e., the total forecast error changes is better explained by the ensemble member with localization advection. With a simple two-layer model, Gasperoni and Wang (2015) employed a more sophisticated but computationally expensive Monte Carlo "group filter" technique in the improvement of tracking the evolution of the initial localization. However, Hotta et al. (2017a) showed that the simple advection with horizontal wind is accurate enough for forecast lead-times shorter than 24

hours (e.g. 6 hours) in the context of identifying detrimental observations.

2.2.4 Hybrid FSO

In the hybrid approach to FSO (Buehner et al., 2018), every step is identical to that in AFSO except for the propagation of forecast errors. Instead of projecting the forecast sensitivity backwards using the adjoint model, the ensemble forecasts are used to project the analysis increment forward to verification time t. More specifically, the variational system usually doesn't directly work in model state space but a transformed variable $\mathbf{v}$ where $\boldsymbol{B}_0^{1/2}\mathbf{v} = \delta\mathbf{x}_0^a$ and $\boldsymbol{B} = \boldsymbol{B}_0^{1/2}\boldsymbol{B}_0^{T/2}$ to improve conditioning of the minimization. Substituting this relation into equation 2.5 yields

$$\Delta e^2 \approx [\boldsymbol{M}(\boldsymbol{B}_0^{1/2}\mathbf{v})]^T C(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \tag{2.17}$$

$$\approx \mathbf{v}^T[\boldsymbol{M}\boldsymbol{B}_0^{1/2}]^T C(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \tag{2.18}$$

Here $\boldsymbol{M}\boldsymbol{B}_0^{1/2}$ can be approximated by $\boldsymbol{B}_t^{1/2}$ where $\boldsymbol{B}_t = \frac{1}{K-1}\boldsymbol{X}_{t|0}^f\boldsymbol{X}_{t|0}^{fT}$ that is based on the ensemble perturbation propagated to the verification time t with the nonlinear model. By this approximation, the need of adjoint model is then circumvented. Same as in EFSO, covariance localization is also needed in HFSO to suppress the noise arising from insufficient ensemble size.

2.2.5 Quantifying uncertainty in EFSO and HFSO impact

The major advantage of having an ensemble system is to have nonlinear realizations that provide uncertainty quantification in forecast and analysis. EFSO and HFSO both inherit the same advantage from the ensemble system that allows us to

estimate the uncertainty in the impact. In the analysis process, the impact of the same observation on each ensemble member is different since the background, the final analysis, and the subsequent forecasts are all different for each member. We can then utilize this information to quantify the uncertainty of the impact estimation from EFSO and HFSO. Using EFSO as an example, the deterministic mean impact can be computed using ensemble forecast mean to evaluate observational innovation $\delta \mathbf{y}_0^o = \mathbf{y}_0 - \boldsymbol{H}\mathbf{x}_0^f$ (observation minus background) and the forecast errors $\mathbf{e}_{t|t'} = \mathbf{x}_{t|t'}^f - \mathbf{x}_t^v$ in equation 2.16. On the other hand, the forecast terms can be replaced by individual forecasts from each ensemble member. In this sense, we are computing the observational impact for each ensemble member to achieve the uncertainty quantification. For HFSO, this replacement of ensemble mean with individual ensemble members changes the cost function for the minimization step, so we need to repeat this step for each ensemble member. Thus, even though this quantification is possible for HFSO, it may be computationally unfeasible. In contrast, the uncertainty quantification of impact is merely simple matrix multiplications for EFSO.

2.2.6 Comparison of the three generic FSO approaches

Here we briefly summarize the advantages and disadvantages of the three different approaches for generic FSO, which is generally similar to their corresponding DA methods.

The two main differences in the approaches are the sensitivity propagation

in time and the analysis sensitivity calculation. AFSO utilizes the adjoint model for the sensitivity propagation step and the variational DA system solver for the minimization of the cost function. The adjoint model requires massive efforts for both development and maintenance, and the representativeness of the error growth for different atmospheric processes is the key to accurate impact estimation. Besides, the implementation of minimizing the modified cost function for AFSO is non-trivial. By contrast, the calculation of EFSO involves only simple matrix multiplication, and almost all the information is readily computed on the fly with EnKF (some adjustments of the DA system needed for computing EFSO online will be discussed in next chapter). However, the localization of the covariance matrix is inevitable for the EFSO approach due to insufficient ensemble size. Several methods were proposed to account for the flow following localization in EFSO, but none of them is perfect. Fortunately, in a timescale as short as 6 hours, the localization advection by wind is accurate enough. HFSO requires a DA solver as do AFSO, but the ensemble replaces the adjoint model. Hence, the advantages and disadvantages are also inherited from both EFSO and AFSO. In EnVar system, HFSO may provide better representation compared to two other methods.

We emphasize here that all methods have their advantages and disadvantages, but the applications proposed in this study have no dependency on the choice of AFSO, EFSO, or HFSO. One can easily switch from one approach to another in those applications. The choice of the approach should be naturally based on the DA system configurations, the desired applications, and the implementation difficulties. The selection of EFSO in this study is based on its simplicity and elegance in

computation and implementation aspect.

For a better understanding of the generic FSO techniques, Dr. Auligné (JCSDA) and Dr. Gelaro (GMAO) leads an intercomparison project comparing the impact data from major operation centers, including NCEP, JMA, GMAO, NRL, and UK Met office. Each center uses different DA methods, models, quality control, and FSO approaches. Hence, the intercomparison project could provide insight into how the observational impacts subject to all those differences.

2.2.7 Definition of the detrimental/beneficial observations

The generic FSO technique assigns an error impact value to each observation. This value represents the estimated forecast error change due to the assimilation of the corresponding observation. The positive (negative) impact values indicate that the forecast error increases (reduces) by assimilating such observations, and hence, we call them detrimental (beneficial) observations. This terminology was introduced in Hotta et al. (2017) to avoid the confusion between "positive EFSO impact value", which indicates a detrimental impact, with positive (beneficial) impact.

2.3 Verification of forecast error in generic FSO

In standard formulations for generic FSO, the analysis is used as the verifying truth for the forecast errors. Recently, some studies proposed using alternative sources for verification truth, including observations and independent analyses. In this section, we discuss the use of different verifying truth in generic FSO. Also, a

numerical comparison of different verifying truths will be provided in Chapter 4.

2.3.1 Observations as verifying truth

Recent studies (Sommer and Weissmann, 2016; Verlaan and Sumihar, 2016; Cardinali, 2018) proposed using observations as verifying truth. Here the derivation follows the same notation as before. Starting with the definition of forecast error changes again as before:

$$\begin{aligned}\Delta e^{o2} &= (\mathbf{e}^o_{t|0} - \mathbf{e}^o_{t|-6})^T(\mathbf{e}^o_{t|0} + \mathbf{e}^o_{t|-6}) \\ &= [\boldsymbol{H}_{veri}(\mathbf{x}^f_{t|0} - \mathbf{x}^f_{t|-6})]^T(\mathbf{e}^o_{t|0} + \mathbf{e}^o_{t|-6}) \qquad (2.19) \\ &\approx (\delta\mathbf{y}_0)^T\boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{H}^T_{veri}(\mathbf{e}^o_{t|0} + \mathbf{e}^o_{t|-6}) \qquad (2.20)\end{aligned}$$

where the superscript of o denotes the forecast error $\mathbf{e}^o_{t|0} = H_{veri}(\mathbf{x}^f_{t|0}) - \mathbf{y}_t$ verified by observation, H_{veri} and $\boldsymbol{H}_{veri}$ represents the nonlinear observation operator and the corresponding linearized operator for the verifying observations, and $\mathbf{y}_t$ stands for the observations available as verifying truth at time t. We set the original weighting matrix $\boldsymbol{C} = \boldsymbol{I}$ for each observation since the impact unit naturally depends on the type of observation and it is sometimes difficult and misleading to unify the units of different types of observation. By comparing equations 2.10 and 2.20, it is clear that the two formulation is almost identical except for the addition operator of $\boldsymbol{H}^T_{veri}$. Applying to EFSO formulation and verifying with all available observations, we get

$$\begin{aligned}\Delta e^2 &\approx \frac{1}{K-1}\delta\mathbf{y}^{oT}_0[\rho \circ \mathbf{R}^{-1}\mathbf{Y}^a_0\mathbf{X}^{fT}_{t|0}\boldsymbol{H}^T](\mathbf{e}^o_{t|0} + \mathbf{e}^o_{t|-6}) \qquad (2.21) \\ &\approx \frac{1}{K-1}\delta\mathbf{y}^{oT}_0[\rho \circ \mathbf{R}^{-1}\mathbf{Y}^a_0\mathbf{Y}^{fT}_{t|0}](\mathbf{e}^o_{t|0} + \mathbf{e}^o_{t|-6}) \qquad (2.22)\end{aligned}$$

Although it appears to be slight modifications in the formulation. There are several differences between verifying with analysis and with observations: (1) verified area of the forecast, (2) resulting impact units, and (3) the accuracy of the truth.

When we choose to verify the forecasts with observations, only the forecasts in the observed area are taken into account. A drawback could be the forecasts originated from the data-rich area are not accounted for because the downstream area has less or no observations. The resulting impact estimation does not include the impact of those observations located at the immediate upstream of a data-scarce area. A typical example is the observation impact from the east coast, which is upstream of the oceans (no dense in situ observing network available) of the continents in the mid-latitudes will not be fully taken into account. Another drawback is the unit of the resulting impact estimation that depends on the observation used for verification, and it can be difficult to interpret and aggregate the results. Lastly, the accuracy of the impact estimation of generic FSO depends on the accuracy of the verifying truth. So by nature, the analysis should be more accurate than the observations if the DA system works properly unless the quality of the analysis is severely degraded by large model bias.

2.3.2 Independent source of analysis as verifying truth

Recently, it is being proposed that replacing the self-analysis with an independent analysis from another source could benefit the generic FSO impact estimation since the independent analysis is uncorrelated with the forecasts or is simply more

accurate by nature as truth (Kotsuki et al., 2018, Baker, 2017, personal communication). Kotsuki et al. (2018) have shown that switching the verifying analysis from native analysis to ERA-Interim can result in substantial changes both in individual observational impact and the aggregated impact, especially for moisture component. However, there are not much of in-depth studies that evaluate the impact of the correlations of the verifying truth and the forecasts in generic FSO computation. A comparison of using dependent and independent analysis will be examined in Chapter 4 with the Lorenz (1996) model.

2.3.3 Dependency of generic FSO on Error Metric

As hinted by different choices of the verifying truths, the generic FSO technique is error metric-dependent. For realistic systems, the natural selection of error metric is the moist total energy norm (Ehrendorfer et al., 1999) that includes the variables of interest in most meteorological applications. Sometimes, the dry energy norm (excluding specific humidity) is used when the moist processes are not well-represented in the forecast sensitivity using the adjoint model. However, different error metric should be considered when applying this technique to a specific purpose. For example, one might be more interested in the errors in the hurricane track and intensity prediction when evaluating the observing system that is designed specifically for the improvement of hurricane prediction. Hence, the choice of error metric depends on the application.

2.4 Degrees of Freedom of Signal (DFS) and Its Relation to generic FSO

The purpose of this section is to relate generic FSO with DFS so that it can serve as a generalized framework, which allows a better understanding of both generic FSO and DFS.

2.4.1 Introduction to DFS

The Degrees of Freedom of Signal (DFS; also called analysis sensitivity, self-sensitivity, or information content) originated from a diagnostic tool for statistical multiple-regression analyses that provides an influence estimate of individual data on the analysis. Cardinali et al. (2004) proposed an approximate method to calculate DFS under the context of 4D-Var DA system. Liu and Kalnay (2009) then formulated the DFS in EnKF framework. For the DFS discussion, we loosely follow the formulation introduced in Liu and Kalnay (2009).

The analysis equation 2.6 ($\mathbf{x}_0^a - \mathbf{x}_{0|-6}^b = \boldsymbol{K}\delta\mathbf{y}_0^o$) can be rewritten into the following form:

$$\mathbf{x}_0^a = \boldsymbol{K}\mathbf{y}_0^o + (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\mathbf{x}_0^b \tag{2.23}$$

Applying the linearized observation operator $\boldsymbol{H}$ on both sides gives the same equation in observation space:

$$\mathbf{y}_0^a = \boldsymbol{H}\mathbf{x}_0^a = \boldsymbol{H}\boldsymbol{K}\mathbf{y}_0^o + \boldsymbol{H}(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\mathbf{x}_0^b \tag{2.24}$$

The *influence matrix* S^a, which determines the sensitivity of the analysis to individual data, then can be defined as

$$\boldsymbol{S}^a = \frac{\partial \mathbf{y}_0^a}{\partial \mathbf{y}_0^o} = [\boldsymbol{HK}]^T \approx \frac{1}{K-1}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{Y}_0^{aT} \tag{2.25}$$

The diagonal elements of the influence matrix are the analysis self-sensitivities and the off-diagonal elements are the cross-covariance sensitivities. The information content of observations is usually computed as the trace of the diagonal elements of the target observations $tr(\boldsymbol{S}^a)$. From equation 2.25, it is quite straightforward to see that DFS is simply the transformed gain $\boldsymbol{K}$ which is determined by the ratio of background and observation uncertainties as in equation 2.12 since this is the way DA system determines how much a single observation influences the final analysis with equation 2.6.

Many researchers perform data selection for satellite radiance instruments such as IASI and AIRS based on DFS as a quantified measure (e.g. Rabier et al., 2002; Rodgers, 1996). More details on the selection procedure will be reviewed in the next Chapter when we introduce existing data selection methods for radiance.

It is worth noting that DFS has been found to be useful as an ordering method for serial ensemble square root filter (EnSRF). Several studies confirmed that using reduction of the error variance $\frac{\boldsymbol{HAH}^T}{\boldsymbol{HBH}^T}$ as an ordering method improves the analysis accuracy significantly (Kotsuki et al., 2017; Nerger, 2015; Whitaker et al., 2008). Moreover, recently it has been recognized that the reduction of error variance is directly related to DFS: $\frac{\boldsymbol{HAH}^T}{\boldsymbol{HBH}^T} = 1 - DFS$ (Hotta 2016, personal communication).

2.4.2 Relating EFSO to DFS

The generic FSO in special form can be directly related to DFS. For simplicity, we show the relation using the ensemble formulation for both FSO and DFS (Liu et al., 2009). With equation 2.14 and 2.25, setting $\boldsymbol{C} = \boldsymbol{H}^T\boldsymbol{H}$, and verifying at the initial time $t = 0$, (so that $\boldsymbol{M} = \boldsymbol{I}$,$\mathbf{e}_{0|0} = 0$, $\mathbf{e}_{0|-6} = -\delta\mathbf{x}_0^a$) yields :

$$\Delta e^2 \approx \frac{1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{M}^T\boldsymbol{C}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \tag{2.26}$$

$$\approx \frac{1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{H}^T\boldsymbol{H}(-\delta\mathbf{x}_0^a) \tag{2.27}$$

$$\approx \frac{-1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{K}\delta\mathbf{y}_0^o \tag{2.28}$$

$$\approx -\delta\mathbf{y}_0^{oT}\boldsymbol{S}^a\boldsymbol{S}^{aT}\delta\mathbf{y}_0^o \tag{2.29}$$

$$\approx -\delta\mathbf{y}_0^{aT}\delta\mathbf{y}_0^a \tag{2.30}$$

where $\delta\mathbf{y}_0^a = \boldsymbol{H}\delta\mathbf{x}_0^a = \boldsymbol{H}\boldsymbol{K}\mathbf{y}_0^o = \boldsymbol{S}^{aT}\delta\mathbf{y}_0^o$ is the analysis increment in observation space. This shows the DFS is indeed included in EFSO when the impact is projected onto observational space at verification time $t = 0$. The physical meaning is quite clear that DFS is simply the analysis sensitivity to observation in observational space which is part of EFSO impact estimation for $t = 0$ (analysis sensitivity to observations or analysis influenced by observations).

To extend this derivation further, we set the verification time from 0 back to t again and substitute in $\mathbf{e}_{t|-6} \approx (\boldsymbol{M}\delta\mathbf{x}_0^a + \mathbf{e}_{t|0}) = (\boldsymbol{M}\boldsymbol{K}\delta\mathbf{y}_0^o + \mathbf{e}_{t|0})$:

$$\Delta e^2 \approx \frac{1}{K-1}\delta\mathbf{y}_0^{oT}\boldsymbol{R}^{-1}\boldsymbol{Y}_0^a\boldsymbol{X}_0^{aT}\boldsymbol{M}^T\boldsymbol{H}^T\boldsymbol{H}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \tag{2.31}$$

$$\approx \delta\mathbf{y}_0^{oT}(\boldsymbol{K}^T\boldsymbol{M}^T\boldsymbol{H}^T)\boldsymbol{H}(-\boldsymbol{M}\boldsymbol{K}\delta\mathbf{y}_0^o + 2\mathbf{e}_{t|0}) \tag{2.32}$$

$$\approx -\delta\mathbf{y}_0^{oT}\boldsymbol{S}^f\boldsymbol{S}^{fT}\delta\mathbf{y}_0^o + \delta\mathbf{y}_0^{oT}\boldsymbol{S}^f\boldsymbol{H}(2\mathbf{e}_{t|0}) \tag{2.33}$$

where $\boldsymbol{S}^f = \frac{\partial \mathbf{y}_0^f}{\partial \mathbf{y}_0^o} \approx \boldsymbol{K}^T \boldsymbol{M}^T \boldsymbol{H}^T$ ($\because \delta \mathbf{y}_0^f = \boldsymbol{H}\boldsymbol{M}\delta \mathbf{x}_0^a = \boldsymbol{H}\boldsymbol{M}\boldsymbol{K}\mathbf{y}_0^o$) denotes the influence matrix for the forecasts, or forecast sensitivity to observation in observational space.

DFS has been shown to be an effective way to perform data selection. Further, it will be shown in Chapter 5 that generic FSO can also be used to perform data selection. Moreover, the information not only includes DFS sensitivity measure but the actual impact. This difference is an additional advantage over DFS that generic FSO takes into account the actual response of assimilation of observation and whether the future evolution of such response agrees with the verifying truth instead of just the sensitivity. It is possible that the initial response at assimilation time will decay within short forecasts and it might also be possible that the response will increase the forecast error in the future. Additionally, the probability distribution of innovation $\delta \mathbf{y}_0^o$ should ideally be consistent with the prescribed observation error covariance $\boldsymbol{R}$ and the background error covariance, but the true $\boldsymbol{R}$ is never known, and it is very likely to have certain types of observations with inconsistently large (or small) magnitudes of innovation $\delta \mathbf{y}_0^o$ resulting in under- (or over-) estimation of the observational influence. By computing $\delta \mathbf{y}_0^{oT} \boldsymbol{S}^f$ instead of just $\boldsymbol{S}^a$, we ensure the actual influence of the observations (not the expected influence) are estimated.

2.5 Summary

In this chapter, we offered a brief historical review on the development of the generic FSO family with different approaches along with the derivation of formulation, followed by a discussion of the advantages and limitations in all approaches.

The choice of verifying truth for generic FSO is discussed. And additionally, we demonstrated that DFS is indeed included in FSO in ensemble formulation. This connection between generic FSO and DFS serves as a foundation of the data selection application in Chapter 5.

In the next chapter, we will introduce the two main applications of the generic FSO in NWP and discuss the methodologies used in the study.

Chapter 3

Generic Forecast Sensitivity to Observations: New Applications

In this Chapter, we will extend the discussion of generic FSO to the applications in NWP. In this study, two applications for improving NWP is proposed: (1) Efficient Data monitoring and selection and (2) Fully flow-dependent Proactive Quality Control (PQC) to avoid the flow-dependent detrimental observations. The first two sections will be dedicated to the introduction of the applications, followed by a section describing the questions that need to be addressed in this dissertation.

3.1 Online Observation Monitoring Tool and Data Selection Based on EFSO

We propose to routinely compute the EFSO impact to monitor the observations online. The forecast errors originated from the detrimental observations are likely to continue growing beyond EFSO verification time. Hence the online monitoring of EFSO impacts may alert the forecasters of a forthcoming forecast degradation event such as the notorious forecast skill dropouts that occasionally reduce the forecast skill to a low level. These dropout events are also the motivation for the development of PQC (e.g. Ota et al., 2013; Hotta et al., 2017a).

In fact, there are already many data quality monitoring routines in operation trying to screen out some of these potentially detrimental observations, either

through online decisions based on some thresholds or a quasi-steady predefined QC decisions. The latter is also called data selection which is of critical importance for the success of NWP systems. Merely assimilating all available observations into the system without any quality control is not only computationally expensive but also very likely to degrade the quality of the DA and the associated forecast products. Also, maintaining and renewing the instruments is very costly in terms of human and financial resources. Thus, the evaluation and controlling of the observational impact on the quality of NWP are becoming more critical. We will describe in the following the various types of quality check deployed as part of the data assimilation procedure in operational routines. Then the use of EFSO to improve this data monitoring and selection process will be discussed.

3.1.1 Present Observational QC and Screening Techniques

In this subsection, we briefly review existing QC and screening methods that ensure the quality of observations going into the DA system. The implementations are slightly different in each operational center and here we follow the documentation from European Centre for Medium-Range Weather Forecast (ECMWF; 2016). The "fixed" observation screening in ECMWF consists of two parts: (1) the generic independent screening and (2) the dependent screening. In addition, we will also introduce a technique called variational QC that was developed to deal with erroneous observations within variational DA system.

3.1.1.1 Generic independent screening

In this part, the decisions are made independently for each observation without involving other observations or previous screening decisions.

1. Completeness check: the observational reports with incomplete entries are rejected.

2. Blacklist: the observations are rejected based on a blacklist library. The library is constructed with two parts. Firstly, complicated data selections such as radiance channel selections are implemented via the blacklist check. Secondly, the library is updated with monthly monitoring of the stations that provide excessively noisy or biased data.

3. Background quality control: the observations are discarded if the background departure (or observation innovation) $\delta \mathbf{y}_0^o = \mathbf{y}_0^o - H(\mathbf{x}_0^b)$ exceeds a predefined threshold determined by the error variance of both the background and the observation.

3.1.1.2 Dependent screening

The dependent screening integrates the information from the previous independent screening and other observations.

1. Vertical consistency check: for multi-level reports, the observation is rejected if the adjacent four layers were discarded in the previous checks.

2. Duplication check: this checks for duplication of reports and discards one of them.

3. Redundancy check: for some report types, only the report closest to the analysis time with most trusted data from the same station are retained, and the others are discarded as redundant reports.

4. Thinning: horizontal thinning is performed for observation types that provides spatially dense observations, such as aircraft, radiance, and scatterometer data.

The online monitoring tool and data selection based on EFSO proposed in this study can significantly improve the efficiency and the accuracy of the construction of the blacklist library mentioned above. Owing to its computational efficiency, EFSO can be used as an online monitoring tool. So the detrimental subset of observations can be identified and avoided through frequent updates of the blacklist library based on EFSO impact. On the other hand, complex data selection decisions for existing or new instruments can also be achieved through aggregation of EFSO impact for an extended period (e.g., one month). This long record of EFSO impact allows identifying specific subsets of observations that keeps degrading the forecasts. Both the EFSO-based monitoring and the data selection can provide an efficient and accurate improvement to current procedures.

3.1.1.3 Existing data selection (blacklist) methods

Present data selection approach can be differentiated into physics-based, OSEs-based, and statistics-based methods. The physics-based data selection typically does not have standard procedures. It is usually based on the physical characteristics of the observation or the environmental condition of the observed quantity. A perfect example is the NOAA/NESDIS method for channel selection (Gambacorta and Barnet, 2013), which is mainly based on our knowledge of spectral properties of the data. This approach requires a comprehensive understanding of the spectral properties in both the reality and the model. OSEs impact evaluation is another widely used data selection method. It is, however, limited by its high computational costs as mentioned in Chapter 2.

We now focus on the statistics-based method used mainly in satellite radiance channel selection. The main reason for the need for the statistical methods are the multi-channel and hyperspectral radiance instruments. Nowadays, around 80% of the total data comes from these space-borne instruments in the form of radiance and derived products, such as feature-tracking winds. These instruments include (but are not limited to) the High Resolution Infrared Radiation Sounder (HIRS) with 20 channels, the Advanced Infrared Sounder (AIRS) measuring 2378 channels, the Infrared Atmospheric Sounding Interferometer (IASI) measuring 8461 channels, the Advanced Technology Microwave Sounder (ATMS) with 22 channels, the Visible Infrared Imaging Radiometer Suite (VIIRS) with 22 channels, and the Cross-track Infrared Sounder (CrIS) with 1305 channels. These unprecedentedly large volume

of data poses challenges in both data transmission and effective assimilation, mainly due to the operational constraint on releasing forecast product on time. The data needs to be downsized while preserving most of the information contained.

The goals of data selection are to (1) ensure data quality, (2) guarantee the representativeness of the model and data, (3) avoid redundant information, and (4) prioritize observations with most information content. The most obvious goal is to prevent bad data that contains large gross error or is associated with faulty instruments from going into DA system. Additionally, we need to account for the imperfections in the model. There are many satellite channels with spectral properties that are not well represented in the model, such as surface conditions, the effect of hydrometeors in the atmosphere. Also, there is a considerable portion of redundant information among the channels. These redundant observations could result in long processing time wasting computational resources with no gain and the assimilation of correlated observations, which can degrade analysis quality if not appropriately addressed. The last is to identify high impact data and prioritize them so that the most information content is assimilated under the time constraint.

Hence the problem is the choice of optimal selection of an acceptable number of channels for assimilation. Here we briefly introduce the existing statistical data selection methods.

One commonly used data compression method is principal component analysis (PCA) via eigenvector decomposition. The data, while retaining most of the variability and information contained, is compressed to a smaller volume (e.g. Goldberg et al., 2003).

Rabier et al. (2002) compared several techniques for channel selection for hyperspectral instruments and concluded that the iterative DFS (information content)-based method proposed by Rodgers (1996) is the most optimal way for channel selection. The process goes through many iterations of selections of channels with the highest DFS. By the definition of DFS, the selected channels are the ones that the analysis is most sensitive to. However, data selection using DFS guarantees only the assimilation of the most influential data on analysis, not the resulting accuracy of analysis nor the impact of the data on the subsequent forecast. It is possible that the seemingly large initial impact on the analysis lies in the decaying modes and dies off in a short period. Moreover, the initially substantial impact could turn out to be detrimental to subsequent forecasts in the worse case scenario. Hence, we propose to utilize EFSO to improve the data selection results by actually evaluating the impact of the data on future forecasts.

3.1.1.4 Variational QC

An alternative approach to deal with observations having a large departure from the corresponding model forecast was proposed in variational context. Under the assumption of Gaussian distributed observational error, very large departure is rare and contributes to unrealistically large analysis increment that degrades the analysis. The variational QC (Anderson and Järvinen, 1999; Ingleby and Lorenc, 1993) assumes that the observational error is constituted by random error (Gaussian distribution) and gross error (constant distribution with bounded interval).

A drawback of this approach is that the error distribution goes back to Gaussian again for very large error. A Huber norm approach (Tavolato and Isaksen, 2009, 2015) was first developed for 4D-Var that adjusts the observational error covariances based on the departure from current analysis solution so that low-quality observations are not completely discarded but weighted down. The ensemble version of this approach was then invented by Roh et al. (2013), but the method uses departure from background instead of using the departure from the analysis. The Huber-type methods then developed into a mathematical framework of so-called "Robust" data assimilation with respect to outlier data (Rao et al., 2017).

3.1.2 Using EFSO for Data Monitoring and Selection

We proposed to establish online monitoring of EFSO impact and design data selection based on long-term EFSO statistics from the monitoring tool in complement to minimum OSEs. To avoid degrading the quality of NWP from blindly assimilating all observations, operational centers have implemented similar approaches for data selection (blacklist in the previous section). The design of these approaches relies heavily on OSEs and DFS. A significant number of OSEs are performed for each observing systems in selecting of the subset of the observations that should be assimilated. Radiance channel selection is a typical example of such kind of approach, and it is of critical importance to decide whether specific radiance channels should be assimilated. The task becomes even more challenging with hyperspectral instruments, such as Infrared Atmospheric Sounding Interferometer (IASI) and At-

mospheric Infrared Sounder (AIRS), which measure thousands of channels at once. Because of the limited discernibility of OSEs, it is challenging to identify a subset of detrimental channels that may be concealed in a generally beneficial observation set. Additionally, the design of the QC for the specific observing system could take up to a year due to the high computational cost of OSEs and prolong the assimilation of new systems. Using the instrument-integrated precipitation map data from Tropical Rainfall Measuring Mission (TRMM) as an example, Lien et al. (2018) demonstrated that using EFSO complemented with a minimum number of OSEs can significantly accelerate the data selection development of new observing systems. We propose to use the similar paradigm to identify detrimental subsets of data using EFSO and perform a minimal number of OSEs just for verification of the EFSO-based selection. We will show examples in Chapter 5 of radiance channel selection that detrimental channels can be identified even within a massive amount of generally beneficial channels, and that this approach can not only accelerate the implementation of the new observing system but also find detrimental channels in the existing operational data selection.

3.1.3 Extra requirements for online EFSO calculations

Using NCEP as an example, there are few necessary adjustments to the operational system for computing EFSO online in addition to the actual computation of EFSO. For EFSO verified at 6 hours, the deterministic forecast needs to be extended from the original 9 to 12 hours. Additionally, observation diagnostic files

containing the innovation and background perturbation in observation space needs to be stored for EFSO computation (to prevent redundant processing). If a longer verifying lead-time is required, both the integration length of the deterministic and ensemble forecasts need to be adjusted accordingly. For example, 24-hour EFSO requires a 30-hour length of deterministic forecast and the 24-hour length of ensemble forecasts (from the original 9-hr length). In this regard, the 6-hour verifying lead-time may be the preferable choice.

3.2 Proactive Quality Control

Taking advantage of immediate EFSO impact in each DA cycle, PQC is an EFSO-based fully flow dependent QC. It was pioneered by Ota et al. (2013) and Hotta et al. (2017a) targeting the alleviation of the forecast skill dropout problem (or forecast busts called by ECMWF). Some of the Global Forecasting System (GFS) forecast skill dropout events result from assimilating faulty observations found by the Global Forecast Dropout Prediction Tool (GFDPT; Kumar et al., 2017) team at National Center of Environmental Prediction (NCEP). They found the GFS forecasts regain skill by ingesting pseudo-observations generated from the European Centre for Medium-Range Weather Forecasts (ECMWF) analyses showing some evidence that the culprit may not the model but the assimilation. It was also found in some dropout events that it can be alleviated by (subjectively) removing suspicious radiance channels. While the role of imperfections in the model cannot be completely eliminated, they show some evidence that faulty observations may be

the culprit for some dropout cases. Further investigation is needed to conclude with statistical significance. Trying to alleviate the skill dropout problem, Hotta et al. (2017a) rejected observations having a detrimental EFSO impact on pre-identified forecast skill dropout regions and thus improved the 5-day forecasts. We introduce in the following subsections the algorithm, data denial strategies, and PQC update methods used in both Hotta et al. (2017a) and this study. Also, possible shortcuts to reduce the computational costs in operational context will also be discussed. We will also compare the differences between PQC and another similar technique called key analysis error (Klinker et al., 1998; Isaksen et al., 2005).

3.2.1 PQC Algorithm

The essential concept of PQC is to utilize the immediate EFSO impact as observational QC for each DA cycle for the identification of detrimental observations. The analysis is then modified to avoid the impact of those identified detrimental observations. It should be noted that EFSO cannot be computed until the next analysis becomes available for forecast error verification. The PQC algorithm can be summarized with figure 3.1. It inserts additional steps (verifying analysis for EFSO, EFSO computation, PQC analysis update, and the forecast from the updated analysis) into a standard DA cycle.

Suppose the desired PQC time is $t = 0$ in a DA system with 6-hour assimilation window, the PQC steps are:

1. Run standard DA cycle from $t = -6$ to $t = 6$ to get the analysis at $t = 6$ for

verification.

2. Obtain 12-hour and 6-hour forecasts from $t = -6$ and $t = 0$ respectively.

3. Perform 6-hour EFSO with the above information to determine which observations should be rejected at $t = 0$ using the selected data denial method.

4. Update the analysis to avoid the rejected observations (described later in this section).

5. Repeat the forecast from the PQC corrected analysis to carry on the improvement.

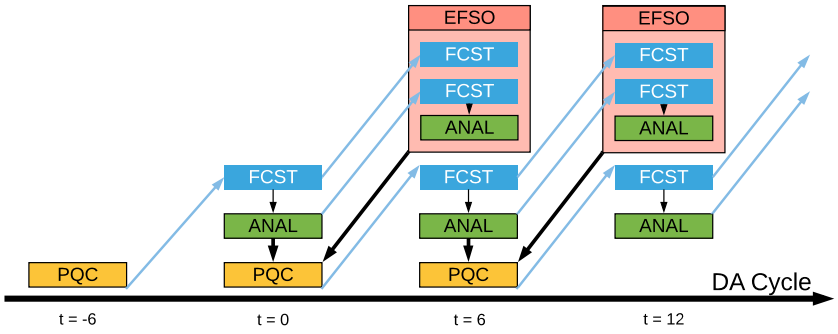


Figure 3.1: Flowchart of cycling Proactive QC algorithm (adapted from Hotta et al. (2017a)).

3.2.2 Data denial strategy

One of the main focus of Hotta et al. (2017a) is the comparison of several data denial strategies. The strategy determines the observations to be rejected in Step 3 of the algorithm. Here we introduce two strategies, namely PQC-Hotta and Threshold (PQC-THR).

- **PQC-Hotta:**

  EFSO verified after 6 hours is performed for a pre-identified forecast skill dropout regions in Hotta et al. (2017a). The regions are defined as sliding latitude-longitude rectangular cells of the size of $30^o \times 60^o$, where either the ratio of 6-hour forecast errors to the mean 6-hour forecast error in that region or the ratio of 6-hour forecast errors to the 12-hour forecast errors is larger than twice of their standard deviations (Ota et al., 2013). In Hotta et al. (2017a), several data denial strategies were tested, and it is concluded that rejecting all detrimental observations from observing systems having a net detrimental impact on the skill dropout regions performs the best.

  The advantage is that only a limited number of observations and model space are involved in the calculation which saves computational time. However, a major drawback of this strategy is that the EFSO impact only takes into account the forecast error changes inside the rectangular cells that suffered from skill dropout. It is very likely for the observations to have impact range exceeding the rectangular cells (for variational DA and also for ensemble DA if the localization cutoff length is large). Besides, the PQC corrections are

limited only to the regions showing mature sign of forecast skill dropout. This regionally confined correction may be fine if the goal is to just alleviate the identified skill dropout event, but the potential corrections not directly related to this identified skill dropout event are left out, and the seemingly detrimental observations could be largely beneficial outside of target cell. Also, multiple EFSO calculations are needed if there are more than one skill dropout regions.

- **Threshold(PQC-THR):**

  To avoid the drawback of regional EFSO in PQC-Hotta, we perform 6-hour EFSO for the entire globe and reject any observation that is more detrimental than a chosen threshold for EFSO detrimental impact. This strategy is the one we used throughout the dissertation. We will explore the dependency of PQC improvement on tuning the thresholds in Chapter 4 with the Lorenz (1996) system and Chapter 6 with the GFS.

3.2.3 PQC update methods

A major focus of this study is to compare the performance of five possible PQC analysis update methods defined as follows (also summarized in Table 3.1):

- PQC_H: This method avoids the influence of detrimental observations by deleting their corresponding columns from the observational operator $\boldsymbol{H}$, which is equivalent to ignoring the observation. This is the method adopted in Ota et al. (2013) and Hotta et al. (2017a) for PQC. Since it is easy to implement, this method is also one of the commonly used methods for data denial. A ma-

jor disadvantage of this method is that it is necessary to repeat the analysis step to obtain the PQC update.

- PQC_R: This method removes the influence of detrimental observations by increasing the corresponding observational error covariance $\boldsymbol{R}$. It can be viewed as a soft version of PQC_H with the flexibility of tuning the magnitude of the detrimental influence. It is also commonly used for data denial and easy to implement. It has the same drawback as does PQC_H, requiring the repetition of the analysis step.

- PQC_K: This method was proposed in Ota et al. (2013) and Hotta et al. (2017a) as an approximation to PQC_H by computing the PQC correction to analysis using the same gain matrix $\boldsymbol{K}$ but ignoring the columns associated with the detrimental observation.

  Mathematically,

$$\begin{aligned} \mathbf{x}_0^{PQC\_K} &= \mathbf{x}_0^b + \boldsymbol{K}[\delta\mathbf{y}_0^o - \begin{bmatrix} \mathbf{0} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix}] = \mathbf{x}_0^a - \boldsymbol{K} \begin{bmatrix} \mathbf{0} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix} \\ &= \mathbf{x}_0^a - \frac{1}{K-1}[\boldsymbol{X}_0^a \boldsymbol{Y}_0^{aT}]\boldsymbol{R}^{-1} \begin{bmatrix} \mathbf{0} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix} \end{aligned} \tag{3.1}$$

  where the superscript "deny" denotes the vector of rejected observation. However, this method is, in fact, more consistent with EFSO since it also assumes the gain matrix to be the same for impact estimation. In Section 4.4.2, we will show that this method performs much better than PQC_H and PQC_R, and does not require repeating the process of computing the gain matrix $\boldsymbol{K}$

(analysis step) so that it significantly reduces the computational burden.

- PQC_BmO: Another approach, following the thought of not repeating the analysis step and the idea of serial EnKF, is to treat the original analysis as background and assimilate the innovation (Observation minus Background, OmB) associated with detrimental observations again but with the opposite sign (BmO), thus canceling the influence of those observations.

  This is equivalent to

$$\begin{aligned}
\mathbf{x}_0^{PQC_BmO} &= \mathbf{x}_0^b + \boldsymbol{K}\delta\mathbf{y}_0^o + [\mathbf{0} \quad -\boldsymbol{K}_{PQC}] \begin{bmatrix} \delta\mathbf{y}_0^{o,keep} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix} \\
&= \mathbf{x}_0^b + [\boldsymbol{K} - [\mathbf{0} \quad \boldsymbol{K}_{PQC}]] \begin{bmatrix} \delta\mathbf{y}_0^{o,keep} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix}, \\
\boldsymbol{K}_{PQC} &= (\boldsymbol{A}^{-1} + \boldsymbol{H}_{PQC}^T \boldsymbol{R}_{PQC}^{-1} \boldsymbol{H}_{PQC})^{-1} \boldsymbol{H}_{PQC}^T \boldsymbol{R}_{PQC}^{-1}
\end{aligned}$$

  The $\boldsymbol{H}_{PQC}$ and $\boldsymbol{R}_{PQC}$ represents the corresponding observation operator and observational error covariance for the rejected observations, and the innovation vector $\delta\mathbf{y}_0^o$ is sorted from the observations we keep $\delta\mathbf{y}_0^{o,keep}$ to the ones we reject $\delta\mathbf{y}_0^{o,deny}$. The $\boldsymbol{K}_{PQC}$ is the Kalman gain using original analysis as background. The effective gain matrix is the same for the block associated with beneficial observations, and a correction term is added for that associated with detrimental observations. We can express PQC-modified block (associated with detrimental observations) in the hybrid gain formulation of Penny (2014):

$$\hat{\boldsymbol{K}} = \beta_1 \boldsymbol{K}' + \beta_2 \boldsymbol{K}_{PQC} + \beta_3 \boldsymbol{K}_{PQC} \boldsymbol{H}_{PQC} \boldsymbol{K}' \tag{3.2}$$

, where $\beta_1 = 1, \beta_2 = -1, \beta_3 = 0$, and $\boldsymbol{K}'$ is the columns of $\boldsymbol{K}$ associated with detrimental observations. Note that PQC_BmO becomes PQC_K if the original background is used to construct $\boldsymbol{K}_{PQC}$.

- PQC_AmO: This method is a variant of the PQC_BmO. The only difference between the two is the definition of the innovation. Innovation in PQC_BmO is defined as the observation minus the original background whereas that in PQC_AmO as observation minus the original analysis.

  Mathematically,

$$\mathbf{x}_0^{PQC\_AmO} = \mathbf{x}_0^a - \boldsymbol{K}_{PQC}[\mathbf{y}_0^{o,deny} - \boldsymbol{H}_{PQC}\mathbf{x}_0^a]$$

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \boldsymbol{K}[\mathbf{y}_0^o - \boldsymbol{H}\mathbf{x}_0^b]$$

$$\Rightarrow \mathbf{x}_0^{PQC\_AmO} = \mathbf{x}_0^b + [\boldsymbol{K} + [\mathbf{0} \quad \alpha\boldsymbol{K}_{PQC}(\boldsymbol{I} - \boldsymbol{H}_{PQC}\boldsymbol{K})]] \begin{bmatrix} \delta\mathbf{y}_0^{o,keep} \\ \delta\mathbf{y}_0^{o,deny} \end{bmatrix}, \quad \alpha = -1$$

  The $\mathbf{y}_0^{o,deny}$ represents the rejected observations. It is clear in this form that the analysis increments associated with beneficial observations are still the same as original, but the ones associated with detrimental observations are changed. The modified block of the gain for detrimental observations of PQC_AmO also has the exact same form as the hybrid-gain formulation of Penny (2014) with the scaling parameter $\alpha = -1$. The corresponding parameters in the form of equation 3.2 are $\beta_1 = 1, \beta_2 = -1$, and $\beta_3 = 1$.

The choice of PQC method differs mostly in the computational requirement and the changes in $\boldsymbol{K}$. The PQC_H and the PQC_R require the most computational resources, while PQC_K poses the lowest computational burden. Regarding the

change in the gain matrix $\boldsymbol{K}$, it is modified the most in PQC_H, PQC_R, and moderately in PQC_BmO and PQC_AmO. PQC_K does not alter the gain matrix $\boldsymbol{K}$ at all. We will discuss more on the importance of the change in $\boldsymbol{K}$ in Section 4.4 when comparing the performance of the methods.

Table 3.1: Summary of PQC update methods

Methods	Mechanism	Change in $\boldsymbol{K}$	Change in Spread	Repeat analysis	Computational cost
PQC_H	Recompute $\boldsymbol{K}$ without observation	Large	Increased	Yes	High
PQC_R	Recompute $\boldsymbol{K}$ with up-weighted $\boldsymbol{R}$	Large	Increased	Yes	High
PQC_K	Reuse the original EFSO $\boldsymbol{K}$	None	None	No	Low
PQC_BmO	Assimilate background minus observation	Low	Reduced	(Serial update)	Medium
PQC_AmO	Assimilate analysis minus observation	Low	Reduced	(Serial update)	Medium

3.2.4 A unique form of PQC

It is worth noting here that another interesting application of EFSO is the ordering of observations in serial EnSRF found by Kotsuki et al. (2017). Recall that in Chapter 2, we mentioned that a DFS-based ordering method proposed by Whitaker et al. (2008) is shown to improve the performance significantly. In Kotsuki et al. (2017), several ordering methods were compared including the DFS-based and the EFSO-based method, and the results show that the EFSO-based ordering from most beneficial to most detrimental observation outperforms all other methods. It should be pointed out that the EFSO-based ordering method implemented by Kotsuki et al. (2017) is, in fact, a special and weak form of PQC for serial EnSRF. In serial assimilation methods, the analysis is updated as each observation being assimilated. Naturally, the background spread gets smaller as more observation added. The effect is that the observations assimilated later in the order are not trusted as much even if the error $\boldsymbol{R}$ is the same. Hence the EFSO-based ordering method gradually weighs down the influence of the detrimental observations.

3.2.5 Possible shortcuts to reduce computational burden for PQC in NCEP operational framework

Here we talk about implementing PQC into current NCEP operational framework (see Chapter 5 for the description of the system) and possible shortcuts to minimize the resources needed for PQC.

As mentioned earlier in the algorithm, the full PQC procedure requires EFSO

impact which requires the next available analysis for verification. After determining which observations should be rejected, the analysis is corrected with any of the PQC update methods to avoid the rejected observations, and then better analysis is obtained. To carry on the improvement to the next cycle, the deterministic and the ensemble forecasts from the PQC improved analysis are also required. For the operational implementation of PQC, two shortcuts were proposed by Hotta et al. (2017a) but were never tested. One of the shortcuts PQC-K, as mentioned earlier, turns out to be more accurate compared to the straightforward approach PQC-H. One additional shortcut is proposed in this study. We will verify the validity of the two possible shortcuts at the end of Chapter 6.

3.2.5.1 Using GFS analysis as verifying truth

The most time-consuming part of PQC is the wait for next available analysis for verification. Utilizing NCEP GDAS/GFS dual analysis configuration, Hotta et al. (2017a) proposed that we can save around 2.5 hours by verifying with GFS early analysis instead of GDAS final analysis. A summary schematic of the dual track configuration together with PQC using GFS early analysis as verifying truth is shown in Figure 3.2.

A possible drawback of the replacement is that the accuracy of EFSO depends on the accuracy of verifying analysis and switching to GFS analysis might lead to less accurate EFSO impact. However, this should be only minor changes since GFS early analysis ingests more than 70-80% of the observations available to GDAS

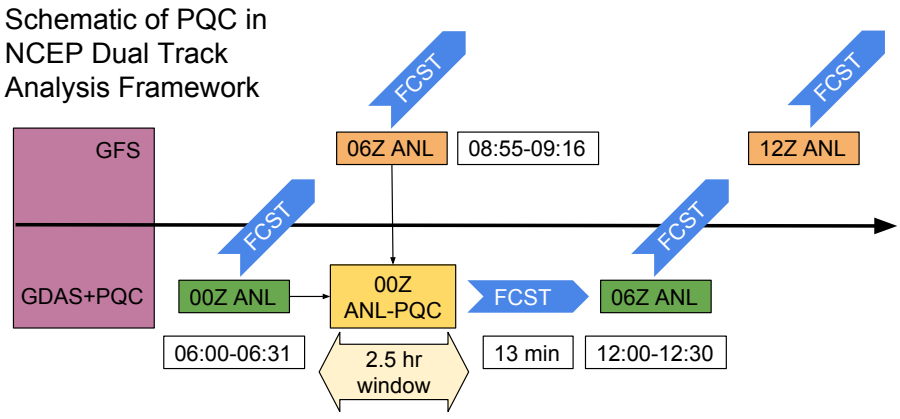


Figure 3.2: Schematics of job flow of PQC using GFS analysis as verifying truth in NCEP GDAS/GFS dual analysis framework

nowadays and lead to small differences between the two. Also, PQC mostly targets the large outliers which should not be very sensitive to the switch of the verifying analysis. (As we will show in Chapter 4 and 6, most of the PQC improvement is from rejecting about 10% of the most detrimental observations and hence not affected by the slight differences.)

3.2.5.2 Reuse the original ensemble forecasts

Repeating the ensemble forecasts from the PQC analysis is very expensive and time-consuming. We argue that this is not necessary since the ensemble forecasts provide only the background error covariance to the deterministic hybrid En-Var system. The changes in ensemble background error covariance should be small enough, and it is not worth extra computational resource to repeat the ensemble forecasts.

3.2.5.3 Reuse the original Kalman gain (PQC-K update method)

Repeating the analysis process is computationally expensive, and the entire process is to obtain implicitly or explicitly the gain matrix $\boldsymbol{K}$. Hence, the reuse of the original Kalman gain, the PQC-K update method we introduced earlier, was proposed in Ota et al. (2013) and Hotta et al. (2017a) as a degraded approximation of the standard PQC-H update method. As we will show in Chapter 4 and 6, PQC-K is, in fact, superior to PQC-H update concerning both the forecast improvement and the computational cost.

In summary, the shortcuts proposed in Hotta et al. (2017a) and this paper includes replacing GDAS analysis with GFS analysis for EFSO verification and not repeating ensemble forecasts after PQC. They will be verified at the end of Chapter 6.

3.2.6 Relation of PQC-K to key analysis error

PQC_K rejects the detrimental part of analysis increments (AIs) in unstable modes via observation space. In this regard, PQC can be viewed as close relative to "key analysis errors" method that perturbs the initial condition based on the adjoint linear sensitivity to future forecast error (Klinker et al., 1998; Isaksen et al., 2005). However, there are fundamental differences between the two. EFSO, and hence PQC, maps forecast error changes, the results of nonlinear model propagation of AIs, back to individual observations and the associated AIs, whereas the "key analysis error" identifies perturbations in the initial condition that could potentially reduce the future forecast error. Additionally, since the norm of PQC corrections is bounded by the magnitude of total analysis increments (the difference between the background and the analysis), they are free from the additional constraints on the size of the correction required in "key analysis errors".

3.2.7 A smoother aspect of PQC

We would like to mention the smoother aspect of PQC as it has been brought up several times in personal communications (e.g., Jeff Whitaker, 2017). In PQC, the main improvement comes from bringing in future information, the verifying analysis, and change the assimilated observation set in accordance to their corresponding "forecast increment". In the smoothers, the analysis is also created by including observations from the future. Hence, there could be a connection between PQC and a smoother since they both bring in future information that should be

further explored.

3.3 Scientific Problems to be Resolved

In this section, we further discuss the questions that should be answered for the (1) data monitoring and selection and (2) PQC applications introduced.

3.3.1 EFSO (Observation monitoring tool and data selection)

3.3.1.1 Why are there 50% of detrimental observations?

It is a well-known feature that nearly 50% of the observations are identified as detrimental in any variant of FSO from every operation center using different models and DA schemes. This high detrimental percentage is indeed counterintuitive and surprising to the community not familiar with FSO since every valid observation is generally considered informative to the system. Many explanations have been provided. A detail literature review alongside with our alternative explanation using Lorenz (1996) system is offered in Chapter4.

3.3.1.2 What are reasonable choices of the forecast verifying truth?

In Chapter 2, we reviewed various choices of the verifying truth for EFSO computation, including the native analysis from the DA system, observations, and the analysis from independent sources. Each option has its advantages and disadvantages. The standard EFSO uses the native analysis which is convenient and presumably more accurate than the observation. On the other hand, using the ob-

servation for verification could save the time waiting for the next analysis and avoid the influence of model error. Moreover, the potential issue caused by the correlation between the forecast and the native analysis can be avoided by using an independent source of analysis. To address this question, we will explore choices with Lorenz (1996) system in Chapter 4.

3.3.1.3 Is 6-hour verifying lead-time fair for the applications?

Under the context of data monitoring and selection, there are various choices of the lead-time for forecast verification. It has become customary to use 24 hours as the verifying lead-time. The reason behind this choice is to avoid the influence of diurnal cycle and capture the long-term dynamical evolution. As explained earlier, at least t+6 hours of deterministic forecast and t hours of ensemble forecast are required for computing EFSO verified with t hours lead-time analysis. Hence, the forecast lengths must be extended for computing EFSO with longer lead-time under current operational setup in NCEP. Additionally, the validity of the advecting localization for the ensemble approach and the linearized models for the adjoint approach deteriorates with longer lead-times, compensating the benefits of having longer lead-times. Hence, it worth exploring the agreement of data selection using different verifying lead-time.

3.3.1.4 Does EFSO-based data selection improve the forecast?

The most important question to answer for the data selection application is that if the EFSO-based selection improves the forecast. The current selection of the radiance channels, for example, was based on a combination of the comprehending knowledge on the spectral properties, degrees of freedom of signal (DFS), OSEs, and OSSEs, and it is being verified routinely with daily forecasts. On the other hand, we can only afford to perform channel-subtraction from the default selection since the addition of channels are prohibited by our limited computational resources.

3.3.2 Proactive QC

3.3.2.1 What is the benefit of cycling PQC versus non-cycling PQC?

It has already been shown the significant improvement from non-cycling PQC in Ota et al. (2013) and Hotta et al. (2017a). We hypothesize the benefit of cycling PQC in addition to the immediate improvement as in non-cycling PQC is the improved background from the accumulation of past PQC corrections. The accumulation of corrections requires that the improvement of PQC last for several the subsequent DA cycles. This accumulation of improvement is critical since the immediate correction is not affordable in operation. Cycling PQC is more desirable for operational implementation if the improvements do accumulate.

3.3.2.2 How to PQC the detrimental observations and update the correction?

In Ota et al. (2013) and Hotta et al. (2017a) , a tremendously large effort was spent on designing the data-denial strategies, which complicates the system. We propose a simple alternative using a globally fixed threshold of the EFSO impact for rejecting observations. Besides, there are on average 50% of detrimental observations in each cycle as described earlier and we need to explore the sensitivity of the cutoff threshold. Lastly, several PQC update methods are devised with their advantages and disadvantages. Using the Lorenz (1996) system in Chapter 4, we will conclude the superior method regarding the improvement in the forecasts.

3.3.2.3 Should we perform PQC with longer verifying lead-time?

Same as the question discussed in section 3.3.1.3, but with additional complexity. In addition to the previous discussion, PQC requires computation of EFSO in real-time. However, we need to wait for the verifying analysis valid at the chosen lead-time, so PQC becomes more infeasible as the EFSO verifying lead-time increases. By design, the 6-hour PQC is the most feasible configuration. On the other hand, 6-hour PQC may suffer from the influence of diurnal cycle and degrade the longer-term forecast for the short-lived improvement that dies off within 24 hours.

3.3.2.4 How to make PQC feasible in operation?

In section 3.2.5, several shortcuts are proposed to lower the computational burden and shorten the delay from performing PQC in NCEP operational setup. These shortcuts include replacing GDAS with GFS analysis for the verifying truth, reusing the original ensemble forecast in EnVar system, and reusing the original Kalman gain to perform the PQC update (PQC-K method). In Chapter 6, we will explore the validity of all these shortcuts.

3.4 Summary

In this chapter, we continue from the FSO discussion in Chapter 2 and introduce the proposed applications in more detail. A background review of the data selection and quality control method is provided for the data monitoring and selection applications. We also introduce the PQC algorithm together with several shortcuts to lower the computational burden in NCEP operation configuration.

In the next chapter, the applications will be explored in a simple Lorenz (1996) system coupled with ensemble transform Kalman filter for showing a proof of concept and investigate the sensitivity of the applications on configurations that would not be affordable in a realistic system.

Chapter 4

Simple-model Experiments with Lorenz 1996 Model

4.1 Introduction

Proactive Quality Control (PQC; Ota et al., 2013; Hotta et al., 2017a) based on EFSO was proposed aiming to resolve the forecast skill dropout issues (Kumar et al., 2017) through identification and rejection of detrimental observations that may be harmful to the forecast. Ota et al. (2013) showed using the Global Forecasting System (GFS) from the National Centers for Environmental Prediction (NCEP), that denying the detrimental observations identified by EFSO with 24-hr verification lead-time reduced forecast errors in several forecast skill dropout cases. Hotta et al. (2017a) successfully showed with 20 forecast skill dropout cases that the forecast errors are also reduced by rejecting detrimental observation from EFSO verified with only 6 hours. Hence, it was further proposed that PQC would be affordable in operational cycling to reduce or avoid skill dropouts in an online fashion. A major potential benefit in cycling-PQC is that the improved forecast may serve as a better background and subsequently lead to improvement in the following analyses and forecasts. However, cycling-PQC has not been thoroughly tested yet. Idealized simulation experiments in a controlled environment can provide insights on how to optimally set up cycling PQC for realistic models.

In this Chapter, the primary goals are to examine the EFSO characteristics,

compare the PQC update methods, and perform PQC sensitivity tests. We implement EFSO and PQC on the simple Lorenz (1996) model (hereafter, L96) with 40 variables coupled with ensemble transform Kalman filter (ETKF; Bishop et al., 2001) for DA. This ideally controlled environment allows the separation of factors contributing to the errors in DA, EFSO, and PQC that are entangled together in realistic systems. The experimental setup, including the Lorenz (1996) model and the DA method, will be described in Section 4.2. In Section 4.3, we explore some of the characteristics of EFSO in this system. Section 4.4 shows the results of the PQC performance using various configurations as well as its sensitivity to a suboptimal DA system. We summarize the findings of the study in Section 4.5.

4.2 Experimental Setup

4.2.1 Lorenz 1996 system

The main purpose of this study is to test cycling-PQC in a simple idealized system, which allows to separate the impact of each factor clearly and perform sensitivity tests more efficiently compared to a realistic system such as the GFS. To achieve this goal while remaining relevant to the realistic application in the atmosphere, we choose the one-dimensional simplified atmospheric model from Lorenz (1996) which resembles some of the large-scale atmospheric behavior and error growth characteristics. It is a model of N variables $x_1, ..., x_N$ governed by N equations:

$$\frac{dx_n}{dt} = x_{n-1}(x_{n+1} - x_{n-2}) - x_n + F$$

where the quadratic terms on the right-hand-side simulate advection while conserving total energy $\sum_{n=1}^{N} x_n^2/2$, whereas $-x_n$ and F represents dissipation and external forcing that drives the chaotic dynamic. We follow the commonly used configuration as in Lorenz and Emanuel (1998). The constant forcing term F is set to 8, so that the error doubling rate corresponding to the leading Lyapunov exponent λ_1 is about 0.42 model time units or 2.1 days, assuming that 0.05 model time unit is equivalent to 6 hours in physical space based on the error growth rate. This time scale for error doubling is approximately consistent with that of the large-scale atmosphere in the mid-latitudes. The time integration uses the 4th order Runge-Kutta scheme with a time step of $\Delta t = 0.05$ model time units. We will be using this $\Delta t = 1$ step as basic time unit throughout the paper instead of the commonly adopted conversion to physical 6 hours since this can be misleading especially when the dynamical timescale is not necessarily relevant to the error doubling timescale. The model dimension of $N = 40$ is chosen as in common practice.

Each experiment is initialized from a randomly chosen state and spun up for 500 time-steps, allowing the ensemble members to converge to the model attractor. For the control and PQC experiments, an additional 500-step spin-up for DA is performed. Each experimental period is 5000-step long after spin-up. A "truth" run without DA and PQC is performed to generate the observations and verify the performance of the experiments. Following Lorenz and Emanuel (1998), the observations are generated at $\Delta t = 0.05$ intervals by adding to the truth random observational noise drawn from $\mathcal{N}(\mu, R)$. Unless otherwise stated, observations errors are generated with $\mu = 0$ and $R = 0.01$.

For "flawed" (imperfect) observing system experiments, we modify the values of μ and R to make them inconsistent with the prescribed observational error covariance matrix $\boldsymbol{R}$. The sensitivity to the spatial coverage of the observing network is also tested. If the number of observing locations is not equal to 40, their distribution is randomly chosen from a uniform distribution for each DA cycle.

We chose to perform ensemble transform Kalman filter (ETKF; Bishop et al., 2001) with a perfect model and ensemble size of 40 members because we are interested in assessing the EFSO and PQC performance without the need for localization and inflation, which are designed to deal with insufficient ensemble sizes and model error in EnKF (Liu and Kalnay, 2008).

4.2.2 Ensemble Transform Kalman Filter

The Ensemble Kalman Filter (EnKF) is one of the prevailing methods for data assimilation, combining a model forecast with observations to construct a linear least-square error estimation of the true state or analysis. An ensemble of Monte Carlo simulations initiated from K perturbed states form the flow dependent error covariance of the forecast (also known as background error covariance $\boldsymbol{B} = \frac{1}{K-1}\boldsymbol{X}^f\boldsymbol{X}^{fT}$, where $\boldsymbol{X}^f$ is the background perturbation matrix, whose columns are the forecast ensemble perturbations with respect to the mean $\mathbf{x}^f$), thus accounting for model uncertainty and cross-variable correlations. It is then used together with the observational error covariance $\boldsymbol{R}$ to combine the background state $\mathbf{x}^f$ and the observations $\mathbf{y}$ into the analysis state $\mathbf{x}^a$. The analysis equation can be written

as

$$\mathbf{x}^a = \mathbf{x}^f + \boldsymbol{K}(\mathbf{y} - \boldsymbol{H}\mathbf{x}^f) \tag{4.1}$$

$$\begin{aligned}\boldsymbol{K} &= (\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1} \\ &= \boldsymbol{A}\boldsymbol{H}^T\boldsymbol{R}^{-1}\end{aligned} \tag{4.2}$$

where $\boldsymbol{A} = \frac{1}{K-1}\boldsymbol{X}^a\boldsymbol{X}^{aT}$, $\boldsymbol{X}^a$, and $\boldsymbol{H}$ represent the analysis error covariance, analysis perturbations and observation operator, respectively.

The actual implementation of the analysis equation has many variations. In this paper we adopt the ETKF formulated by Bishop et al. (2001). The analysis equation then becomes

$$\boldsymbol{X}^a = \boldsymbol{X}^f\boldsymbol{T} \tag{4.3}$$

$$\boldsymbol{T} = \boldsymbol{C}\boldsymbol{\Gamma}^{-1/2}\boldsymbol{C}^T \tag{4.4}$$

where $\boldsymbol{T}$ represents the ensemble transform matrix. $\boldsymbol{C}$ and $\boldsymbol{\Gamma}$ are computed through eigenvalue decomposition:

$$\boldsymbol{C}\boldsymbol{\Gamma}\boldsymbol{C}^T = \boldsymbol{I} + \frac{1}{K-1}(\boldsymbol{H}\boldsymbol{X}^f)^T\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{X}^f \tag{4.5}$$

where $\boldsymbol{I}$ is the $K \times K$ identity matrix. Note that Hunt et al. (2007) showed that the transform matrix $\boldsymbol{T}$ can be interpreted as the weighting for each ensemble members according to the observations. The members closer to the observations receive higher weights. In addition, a localization version of ETKF was proposed in Hunt et al. (LETKF; 2007). In high-dimensional applications, LETKF instead of ETKF is generally adopted to suppress the under-sampling noise arising from insufficient ensemble size. We use LETKF as the DA system for GFS model in Chapter 6.

4.3 Results: Characteristics of EFSO Impacts

In the first part, we examine the EFSO sensitivity to the verifying lead-time. The dependence on the choice of lead-time is of interest since it may be determined by an intrinsic dynamical timescale, which is relevant to the optimal design of PQC. In the second part, we explore the choice of EFSO verifying truth. Finally, we discuss the causes for the observed low percentage of beneficial EFSO impacts. We offer an explanation supported by the results in the idealized system in the third part.

4.3.1 EFSO sensitivity to verifying lead-time

We begin by comparing EFSO impacts with various forecast evaluation lead-times. In Hotta et al. (2017a), a considerable effort was devoted to validating the use of a lead-time of only 6 hours, needed for PQC given that this is the usual data assimilation interval. By contrast, operational FSO reports commonly use 24 hours lead-time to avoid the errors introduced by diurnal cycle. Also, having to track the flow-following localization function in a high dimensional system, complicates the applicability of EFSO, requiring large computational resources for computing forecast impacts at longer lead-times (Kalnay et al., 2012; Ota et al., 2013; Gasperoni and Wang, 2015). By contrast, the L96 system is governed by dynamics with a single timescale and hence is free of multiple timescale issues. We purposely select to use ETKF with full ensemble size, which requires no localization, to focus on the pure EFSO characteristics in an ideal environment. With this configuration, we can

examine just the dependence of EFSO on the evaluation lead-time.

Figure 4.1 shows a typical snapshot of the 4.1a EFSO impact and 4.1b the normalized (with $l2$ norm) impact of observations at cycle 2200, with lead-times ranging from 1 to 40 time steps. The normalized impact $\hat{e}_{ijk} = \tilde{e}_{ijk}/\sqrt{\sum_j \tilde{e}^2_{ijk}}$, where $\tilde{e}_{ijk}$denotes the EFSO impact of j-th observation with lead-time of k in i-th cycle. The impact is normalized for each lead-time since it is typically an order of magnitude larger with longer lead-time due to nonlinear error growth. We are more interested in the agreement of the sign and the relative magnitude of the impacts between observations at each lead-time rather than the difference in absolute magnitude across lead-times, which is dominated by nonlinear error growth. Note that the signs of the impact are mostly consistent throughout the lead-times for large impacts (e.g., the 9th and 19th observation) and some rather large ones. However, there are several observations (e.g., the 2nd and 3rd observation) with small to medium impacts that decreased in magnitude at the beginning and then changed signs after 5 steps of lead-time.

We take the mean of the normalized impacts of all lead-times ($< \hat{e}_{ij} >= \sum_k \hat{e}_{ijk}$) and use it as the consensus among all lead-times to quantify the agreement between EFSO impact from each lead-time. Figure 4.2 shows the correlation between the mean normalized impacts of each observation and the actual impacts with each lead-time. It is clear that the impact of short lead-time does not correlate well with the mean impact and neither does the impact of the very long lead-time. The correlation peaks at around lead-time of 21 steps, suggesting that EFSO is not able to fully capture the longer-term dynamical evolution with short lead-times,

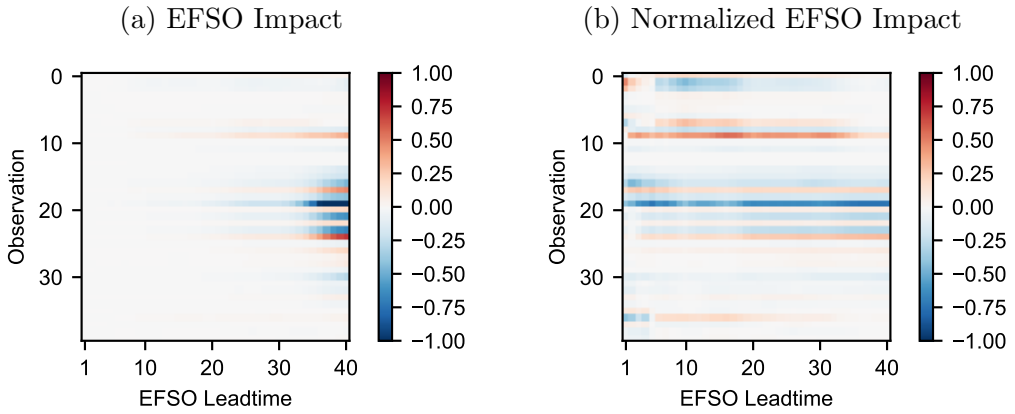


Figure 4.1: Snapshots of EFSO impact as a function of lead-time for each observation at 2200 cycle. Both (a) the impact and (b) the impact normalized with respect to each lead-time are shown.

whereas the growth of nonlinearity limits the accuracy of EFSO for lead-times that are too long. As a result of the balance between the two factors, the ideal lead-time to be used settles at around 21 steps. It is arguably true that the result depends heavily on the model dynamics and may not be generalizable to other models. For this model, 21-step lead-times seems to represent well of the consensus among all lead-times and, as we will see later, this is consistent with the PQC results.

4.3.2 Choice of verifying truth

In this system, we have the luxury of a controlled environment where the truth and error are known, providing the opportunity to investigate the choice of verifying truth. The options are the native analysis, observations, and the independent analysis. As discussed in Chapter 3, there are advantages for verifying with

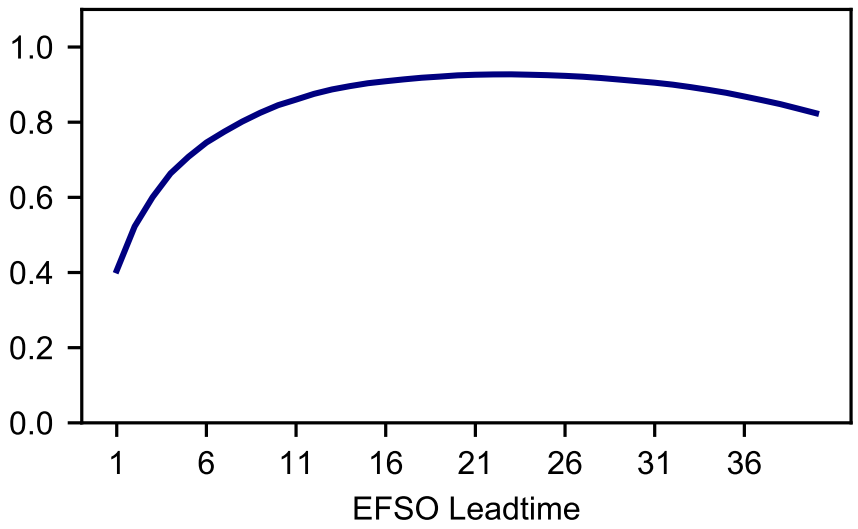


Figure 4.2: Correlation between the normalized mean EFSO impact and the actual impact with each lead-time

the "alternative truths" other than the native analysis. Using observations as the "truth", EFSO is free of any dependencies on models and data assimilations. The advantages are that the EFSO impact evaluation is not contaminated with model error and that it is free of the delay from waiting for DA to complete. Another choice is using an independent analysis, having the advantage of being uncorrelated with the forecasts. An obvious drawback is a dependence on the source (most likely another operation center) producing the analysis. In this demonstration, the independent analysis is generated by adding random perturbations to the analysis with the magnitude of ensemble spread at each grid point to represent a possible analysis having uncorrelated errors.

Figure 4.3 compares the choices by computing the correlations of EFSO verified with the choice and that with the actual truth. The comparison allows a direct

evaluation of the choices with the corresponding accuracy of EFSO impact. We can see that the differences between the choice are the most significant for shorter lead-times, where the observation is less accurate compared to the native analysis. This result is not surprising since the analysis should be more accurate than the observation by nature. The independent analysis follows closely with the native analysis. As we increase the verifying lead-time, the differences decrease when all correlation coefficients converge to one. This behavior indicates that as we increase the lead-time, the verifying truths become more accurate relative to the forecast, and essentially any choice is almost as good as the actual truth as the forecast error is sufficiently large.

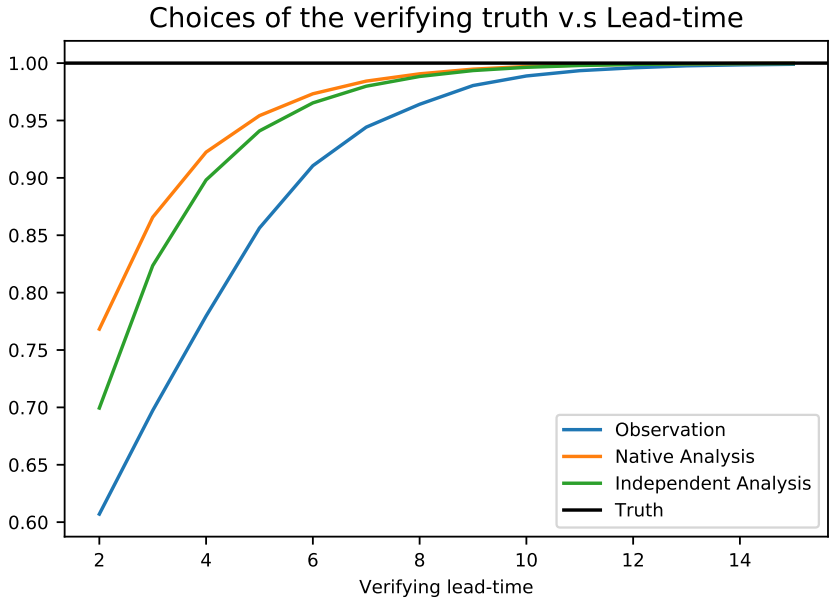


Figure 4.3: Comparisons of the choices of verifying truth by computing the correlations of the EFSO verified with the alternative truth with that verified with the actual truth for various verifying lead-times.

One main proposed advantage of using observations as verifying truth is that it is free of model dependency and hence could become a superior choice if model error is present. In Figure 4.4, the EFSO verified with observations is only superior to the analysis when the forecast model forcing deviates significantly from the truth (large model error). We should note here that the system experienced filter divergence using that particular forcing for the forecast. The lesson we learned by this exercise is that the analysis is always a better choice if the DA system works properly, and the choice of verifying with observations is only superior when the filter diverges, which is very unlikely in the operational system. Also, the EFSO impact verified with analysis seems to be insensitive to the model error until the filter diverges.

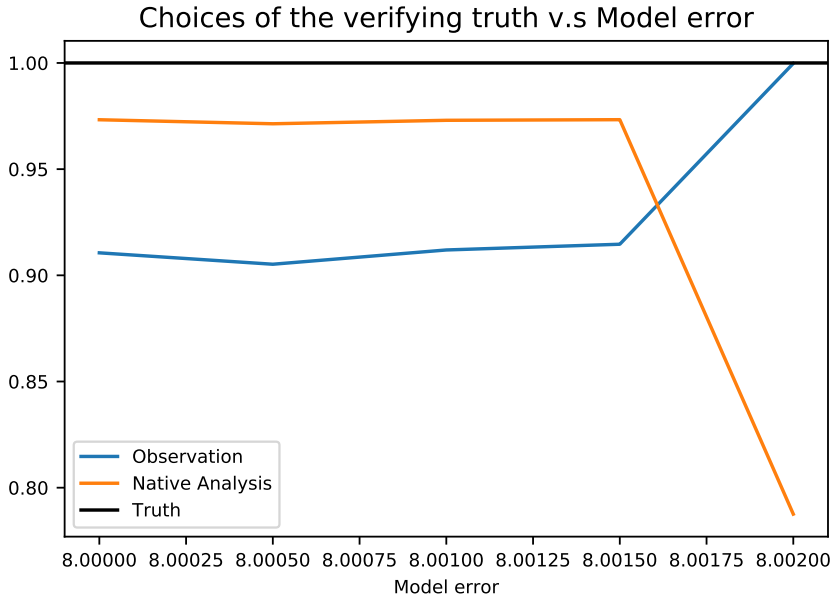


Figure 4.4: Same as in Figure 4.4, but with various model forcing for the forecast model that deviates from the truth. Here we show the 6-step EFSO.

4.3.3 Low beneficial percentage of FSO impact

One puzzling property of FSOs is the low percentage of beneficial observations. From every reporting operational center, the beneficial percentage is only slightly more than 50% no matter which type of FSO was used to make the calculation. Many discussions appear in several studies about the reason why so few of the observations are beneficial. It has been attributed to inaccurate verifying analysis (Daescu and Todling, 2010), incorrect assignment of background and observation error covariance matrix, and the distribution of analysis increment in dynamical modes with different growth rate (Lorenc and Marriott, 2014). These factors all contribute to the determination of the beneficial percentage, but as we will show later, the beneficial percentage is still low even without those proposed suboptimal configurations in our ideal system, suggesting that they are not the dominant factors. Gelaro et al. (2010) mentioned the beneficial percentage could be increased to 60%-65% with a scalar DA system when the background and the observation accuracy are comparable.

Here we offer an alternative hypothesis that the relative difference between the quality of background and that of the observation is a critical factor contributing to the low beneficial percentage of observations. Figure 4.5, is a schematic of a group of unbiased observations centered around the truth and a background forecast initiated 6 hours before lies somewhere on this one-dimensional model space with a distance to the truth. It is clear that the forecast error determines the distribution of innovation on each sign and mainly, the number ratio of the majority to the minority

separated by the background. It is the nature of DA that the majority group contains observations closer to the truth and the associated AIs reduce the analysis error and the subsequent forecast error. By definition, in FSO, beneficial impacts are assigned to the observations in the majority group. From this point of view, we can reach two conclusions about FSO assuming that the quality of observations is fixed and the DA system is working correctly: (1) the beneficial percentage depends on the average quality of the background, and (2) the beneficial percentage cannot be lower than 50% and it can be asymptotically approaching 100% if the background quality is deficient compared to the observational error, where no observation error is more significant than background error.

To test this hypothesis, we leverage the simple L96 system by computing the beneficial percentage with various background quality generated by varying the length of DA window. With figure 4.6a, the averaged beneficial percentage of 5000 DA cycles are plotted against DA windows from once every forecast step to every 100 forecast steps. Beneficial percentage increases with DA window monotonically from 55% and saturates at around 92% when the interval is larger than 60. In figure 4.6b, we tested the impact of observation error standard deviation from 10^{-5} to 10^{0} and the beneficial percentage drops from 67% to 55%. It is not possible to change the observational quality while keeping background quality constant in our system and, hence, the beneficial percentage saturates at 67%, a rather low value compared to 92%, which is a clear demonstration that the relative difference between the quality of backgrounds and observations is the dominant mechanism for the low beneficial percentage in this simple system. Besides, we can also view beneficial percentage

from the information contribution perspective. The low beneficial percentage indicates that the DA system relies heavily on the background while high beneficial percentage suggests more information coming from the observations. Thus, the low beneficial rate of observations in operational NWP system is very likely the consequence of the high quality of the backgrounds such that only slightly over 50% of observations are beneficial in FSO.

4.4 Results: Proactive QC

In this section, we examine the performance of non-cycling PQC in which the improved forecast is not used as background for the next analysis, following Hotta 2017. The first experiment of cycling PQC is performed, and we compare the different PQC update methods introduced in Chapter 3 where each uses different mechanisms to avoid the impact of the detrimental observations. The sensitivity of PQC performance to the choice of EFSO lead-time and the amount of rejected observations are investigated. Lastly, the robustness of PQC is tested with the different sources of imperfections in the DA system, relevant to operational applications.

4.4.1 Non-Cycling PQC

Non-cycling PQC means the PQC-improved forecast is not used as background in the following DA cycle. We begin with using PQC_H, which is usually how data denial (QC) are performed in operations and then examine its performance for different configurations of PQC. In Hotta et al. (2017a), one of the key highlights

is the design of the data denial strategy. The EFSO impact was used as a guide on the denying priority, but not every detrimental observation was rejected. It was intuitively believed that rejecting nearly 50% of the observations will lead to forecast degradation. Additionally, every observation, by the nature of DA, provides an extra piece of information that should be useful in estimating the true state. Given the advantages of this simple low dimensional system, here we can test the sensitivity in a great granularity of PQC_H performance to the number of rejected observations. In figure 4.7, we show how the non-cycling PQC_H improves or degrades the forecasts by varying the number of rejected observations ordered from the most detrimental to the most beneficial ones. In this case, all the observations are perfectly consistent with prescribed error covariance $\boldsymbol{R}$, meaning none of them are flawed. We observe a forecast error reduction mainly resulting from the rejection of the 4 most detrimental observations. Then the error is insensitive to the additional rejection of observations. And not until we reject the last few very beneficial observation we observe that a rapid error growth takes place. This is in agreement with the EFSO impact for the observations that the impact of the most beneficial and detrimental observations are orders of magnitudes larger than that of the insignificant observations. It is also consistent with the improvement increased in Hotta et al. (2017a) that the error reduction by PQC is well preserved and even amplifies in magnitude as the forecasts advance (Note the vertical axis is in log scale.). It should be noted here that in a non-cycling PQC, the impact of using different EFSO lead-times is not significant (not shown) which is consistent with our intuition that different lead-times perturb the rejecting order slightly in most cases and only make significant

changes in rare situations, thereby the effect can only be observable after cycles of accumulation. This result is a demonstration of the ability of EFSO to identify the very detrimental observation from a pool of observations. Also, the result confirms the speculation that only a few very detrimental observations should be rejected for optimal performance of PQC_H.

4.4.2 Cycling PQC: Update Methods

Figure 4.8 compares the performance of all proposed PQC methods using 6-steps and with varying percentages of rejected observations. Note that the magnitude of observational impacts are different and the percentages of beneficial observations could vary from 30% to 70% from cycle to cycle. There could be fewer observations with large detrimental impact for some of the cycles and more on others. Hence it is not desirable to reject the same amount of observations for each cycle. Here we construct a range of thresholds corresponding to the 0th to 100th percentiles of EFSO impacts obtained from a control experiment of 5000 cycles (Table 4.1). Then PQC rejects observations with an impact above the threshold, where the 10th percentile threshold rejects the top 10% of the most detrimental observations and the 90th percentile threshold keeps only the top 10% of the most beneficial observations. The lead-time here is chosen to be 6 steps rather arbitrarily, and the sensitivity to the lead-time will be examined later. Since the Kalman gain $\boldsymbol{K}$ of PQC_R approaches that of PQC_H asymptotically with increasing observational error, it is not surprising to see that the PQC_H and PQC_R methods

perform more or less the same regarding both analysis and 30-step forecast error reduction. The errors are reduced the most when rejecting 10% of the observations for the two methods. This result is consistent with the one from the non-cycling experiment where 10th percentile threshold rejects around 4 observations on average. It is somewhat surprising that PQC_K, PQC_BmO, and PQC_AmO all outperform PQC_H and PQC_R, which are the two most commonly used data denial methods. For the analysis quality improvement, the obvious choice of the threshold shifts towards 20%. PQC_K does not show any degradation of analysis until rejecting more than 60 % of the observations, whereas PQC_BmO and PQC_AmO stop showing improvement after 50% and even suffer from filter divergence beyond 60%. For the forecast quality improvement, the dependence of PQC_BmO and PQC_AmO on the thresholds are qualitatively similar to that in analysis performance. It is very surprising to find that PQC_K has nearly no dependence on the thresholds between the 10th and 60th percentile, especially when compared to the 10% optimal choice for PQC_H and PQC_R.

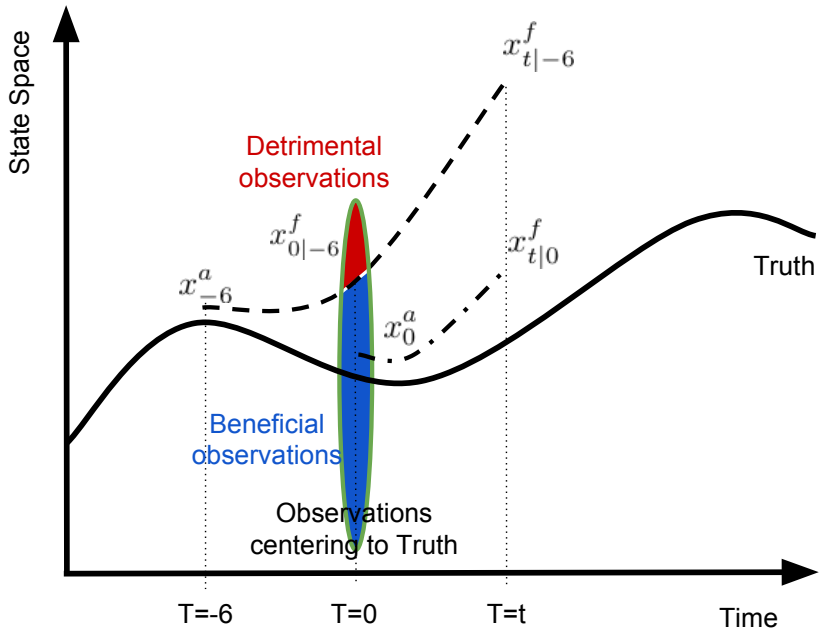


Figure 4.5: Schematic of EFSO computation in a 1-D model. The black dash curve represents the previous forecast trajectory initiated from $T = -6$ and the black dash-dot curve represents the current forecast initiated from $T = 0$. The two both start close to the truth (black solid) and deviate further in time. The assimilated observations are distributed within the green oval that centers at the true state with the probability proportionate to the width representing a Gaussian distribution of error. In EFSO calculation, the observation in the red (blue) shaded area will be quantified as detrimental (beneficial).

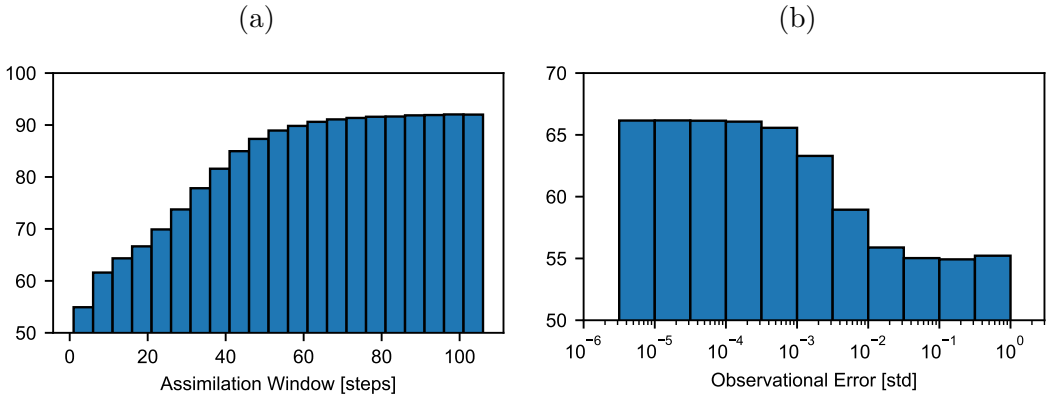


Figure 4.6: Beneficial percentage of EFSO impact as a function of (a) background quality represented by length of assimilation window and (b) observational error.

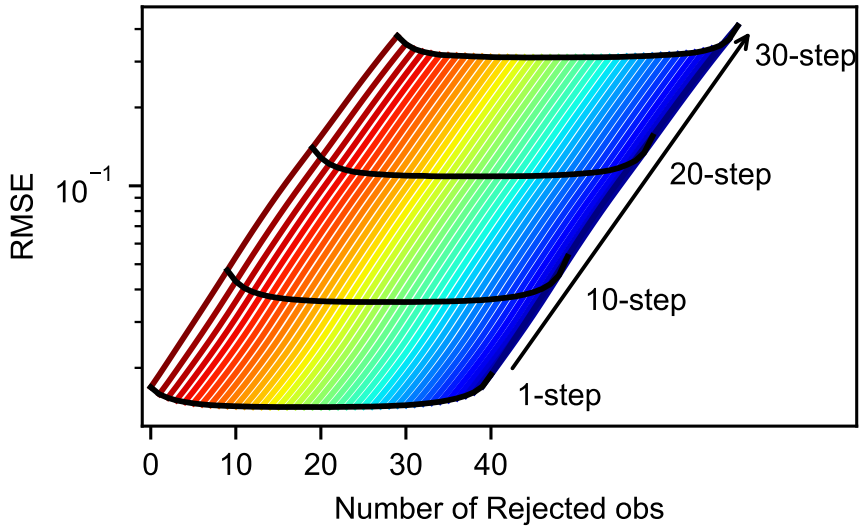


Figure 4.7: Forecast RMSE from single cycle PQC_H as a function of number of rejected observations. The colored lines represent the forecast error evolution from PQC analysis corresponding to increasing number of rejected observations from red to blue. Different colors simply distinguish individual experiments with different numbers of rejected observations. The black lines mark the forecast error at 1, 10, 20, and 30 steps.

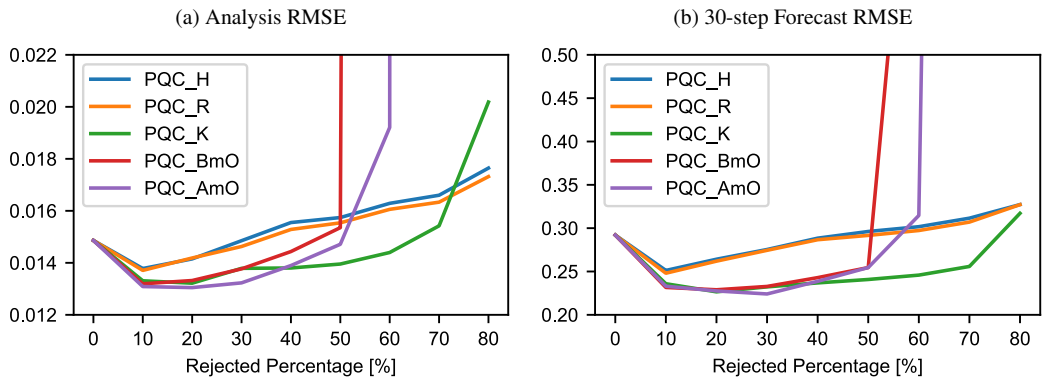


Figure 4.8: Performance of 6-step PQC with all 5 methods in terms of (a) analysis RMSE and (b) 30-step forecast RMSE as a function of rejection percentage.

Table 4.1: Rejecting threshold based on EFSO impact (Jkg^{-1}) as a function of EFSO lead-time and rejecting percentile

Lead-time (steps)	00-th	20-th	40-th	60-th	80-th	100-th
6	4×10^{-1}	3×10^{-4}	3×10^{-5}	-7×10^{-5}	-5×10^{-4}	-6×10^{-1}
11	3×10^{0}	1×10^{-3}	9×10^{-5}	-2×10^{-4}	-1×10^{-3}	-5×10^{0}
16	3×10^{1}	3×10^{-3}	2×10^{-4}	-4×10^{-4}	-4×10^{-3}	-3×10^{1}
21	5×10^{1}	7×10^{-3}	6×10^{-4}	-9×10^{-4}	-9×10^{-3}	-6×10^{1}

Intuitively, the "flat bottom" of PQC_K (rather than the "check mark" shape of PQC_H and PQC_R) is more consistent with the estimated impact of the observations since the magnitude of the impacts between 10th to 60th percentile is small compared to that of those below the 10th percentile. Hence, it should be insensitive ("flat bottom") to rejecting those observations between 10th to 60th percentile, explaining why the results are better for PQC_K than for PQC_H since PQC_K is more consistent with the nature of the computation of EFSO and the estimated impact. Note that EFSO is simply an ensemble-based linear mapping between forecast error changes and the observational innovation, which is associated with analysis increments (AIs) through the gain matrix $\boldsymbol{K}$. It provides the estimated impacts of each observation in the presence of all other assimilated observations, and hence the impacts remain valid as long as $\boldsymbol{K}$ does not change much. However, PQC_H and PQC_R significantly change $\boldsymbol{K}$ when rejecting some observations, thereby the accuracy of the estimated EFSO impacts becomes lower, and the PQC based on those impacts does not work as desired. The total AIs obtained at the end of the update consists of the AIs contributed by each observation, and it is the AIs that determines the forecast error change rather than the observation innovation. Hence, PQC should target the AIs corresponding to the detrimental observations rather than the observations themselves. So simple data denial by manipulating $\boldsymbol{H}$ and $\boldsymbol{R}$ does not necessarily reject the AIs that lead to forecast degradation especially when rejecting an excessive number of observations. PQC_K, by contrast, uses the same $\boldsymbol{K}$ to reject the exact detrimental AIs identified by EFSO and ends up with even larger improvements. Interestingly, PQC_K was originally proposed as an "ap-

proximation" of PQC_H to avoid the large computational cost of recomputing the analysis in realistic system applications (Hotta et al., 2017a; Ota et al., 2013). In reality, as discussed above, PQC_K is, in fact, much more accurate in the context of PQC based on EFSO. Besides, the observations with the largest impacts contribute to AIs among the most unstable modes, while the less impactful observations are associated with the neutral and stable modes which have little or no error growth. Hence, after rejecting the few very detrimental AIs, it does not matter much whether those less impactful AIs are rejected since the difference is very unlikely to grow in the future, thereby showing the "flat bottom" feature in the center of Fig. 4.8.

For PQC_BmO and PQC_AmO, they change $\boldsymbol{K}$ in a less radical fashion by "assimilating" new observations into the original analysis and yield improvements similar to PQC_K with a small number of rejected observations. However, they suffer from filter divergence easily with a large number of rejected observations since the ensemble becomes overly confident due to the "additional" assimilation of opposite innovations. It is worth noting that the commonly observed difference in the impact estimated by EFSO and observing system experiments/ data denial experiments corresponds to the difference in PQC_K and PQC_H.

4.4.3 Cycling PQC: Sensitivity to Lead-times

In this section, we explore the sensitivity of cycling PQC_H, PQC_K, and PQC_AmO to the rejecting threshold and more importantly, to the EFSO lead-time. We refer to PQC based on a t-step EFSO as t-step PQC hereafter. Beginning

with PQC_H shown in Figure 4.9, it seems to be quite difficult to conclude the relation between the PQC_H performance and the length of lead-time directly from the analysis error reduction. It is also counter-intuitive to find that 21-step PQC_H performs worse than that with only 6 steps. However, the dependence of the performance on forecast error reduction can be easily summarized as follows. The forecast quality increases with the lead-times up to 16 steps and distributes then remains the same with 21 steps. The result suggests the optimal choice of EFSO lead-time settles between 16 to 21 steps, best describing the impact corresponding to the underlying dynamical evolution. Short lead-time PQC seemingly reduces the analysis or even short-term forecast error but may be irrelevant to long-term error growth. This speculation can be confirmed by the comparison between the performances of 6-step and 21-step PQC_H. It is clear that a considerable portion of error reduction by 6-step PQC_H is distributed within the stable subspace and decays over time. Additionally, the corrections of 21-step PQC for rejecting 10% of the observations does not reduce much of the total analysis error compared to other lead-times, but it turns out those are the most relevant to error growth in long-term, and rejecting them leads to huge forecast improvement.

Now we show in Figure 4.10 the sensitivity of the performance of PQC_K to the lead-time and the rejecting threshold. It is qualitatively consistent with the result of PQC_H, where the maximum forecast error rather than analysis error reduction increases with lead-time and saturates at 16 to 21 steps. There is a general feature in forecast error reduction shared among all lead-times. The error drops dramatically when rejecting with 10th percentile followed by the "flat bottom" feature when

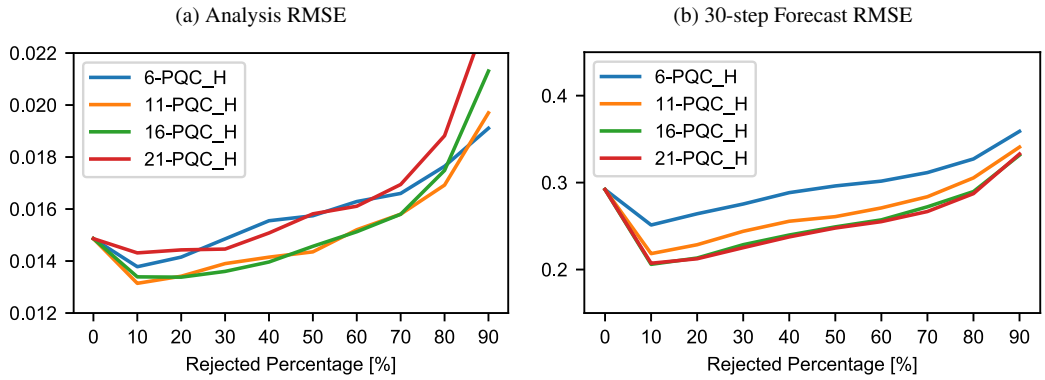


Figure 4.9: Comparison of cycling PQC_H performance using various EFSO leadtime as a function of rejection percentage in terms of (a) analysis and (b) 30-step forecast analysis RMSE.

increasing the rejecting percentile. Then the filter diverges when rejecting beyond a critical percentile. The only differences are the percentile where PQC_K leads to filter divergence and the magnitudes of error reduction. The critical threshold gradually approaches from the 80th percentile for 6 steps to 50th percentile for 21 steps.

This dependence of PQC_K on the rejecting percentile can be explained with Figure 4.5, which is a schematic of EFSO computation in one-dimensional model space with a group of unbiased observations centering around the truth and forecasts initiated from $T = -6$ and $T = 0$ lying somewhere with a distance to the truth. We can think of the 1-dimensional model space in the figure being aligned with the fastest error growing subspace (1st Lyapunov vector). So the error growth depends on the balance between the detrimental AIs and the beneficial AIs. The outermost observations and the corresponding AIs in detrimental direction are the

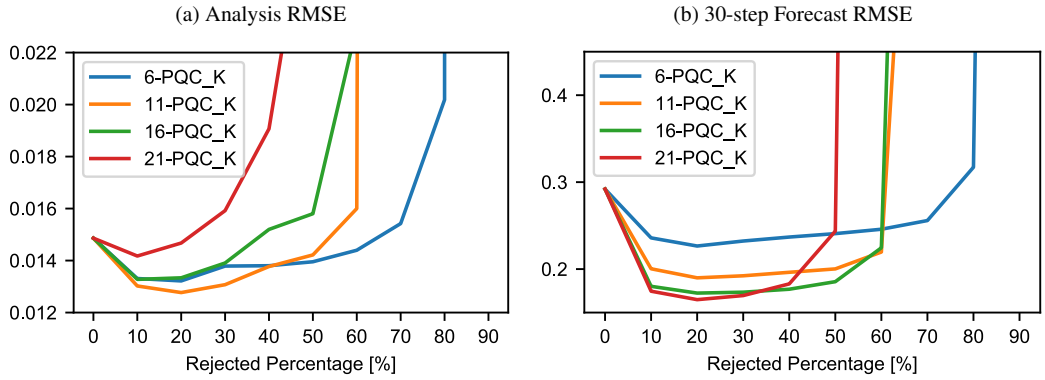


Figure 4.10: Same as Fig. 4.9, but for PQC_K.

primary drivers that deviate the forecast from the truth trajectory, whereas the beneficial observations draw the forecast towards the truth. Ideally, we should reject the detrimental AIs because we could end up with only the unstable growing "beneficial" AIs when more are rejected. The error drop with rejecting with 10th percentile results from getting rid of the largest detrimental AIs and the "flat bottom" is associated with those non-growing AIs. Once rejecting more than 50%, the average beneficial percentage, the unstable growing "beneficial" AIs take over and lead to error growth in the "beneficial" direction. Note that EFSO with shorter lead-time cannot differentiate between observations with small to medium impact. So the direct consequences of using shorter lead-time is a smaller error reduction caused by rejecting some of the beneficial AIs while keeping some of the detrimental ones unintentionally. It also means that when rejecting more than 50% of the observations, the remaining AIs are not entirely pointing to the "beneficial" direction and delay the occurrence of the divergence.

With Figure 4.11, we show the PQC_AmO performance sensitivity to the

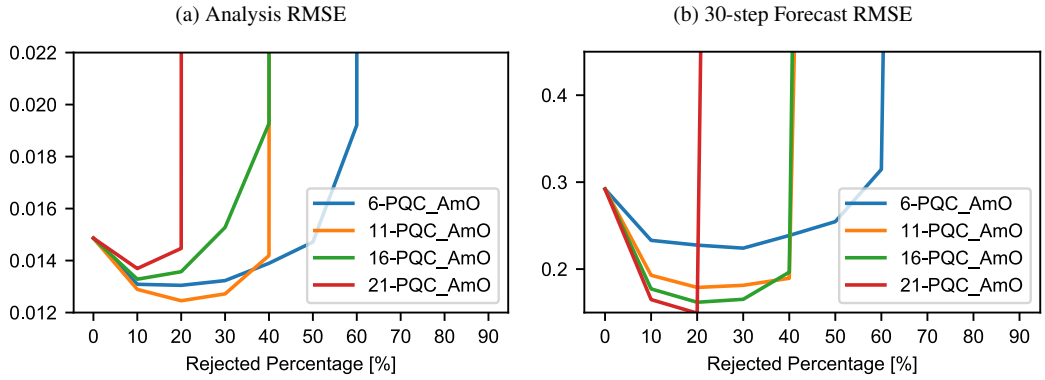


Figure 4.11: Same as Fig. 4.9, but for PQC_AmO.

choice of lead-time and rejecting percentile. It is clear that it performs comparably well or even better than PQC_K in some configurations. However, suffering from further contraction on the ensemble spread by assimilation of the detrimental innovation with the opposite sign, the method can easily lead to filter divergence associated with overly confident ensemble members.

4.4.4 PQC sensitivity to the change in Kalman gain

In Chapter 3, we mention the changes in Kalman gain matrix $\boldsymbol{K}$ introduced by each PQC update methods. Here we examine the impact of the changes in $\boldsymbol{K}$ on the improvement in the resulting analysis and forecast. To separate the PQC corrections on the mean trajectory and the Kalman gain (essentially the analysis covariance matrix; $\boldsymbol{K} = \frac{1}{K-1}\boldsymbol{A}\boldsymbol{H}^T\boldsymbol{R}^{-1}$), the PQC corrections are applied only to the mean of the ensemble, and the perturbations remain the same as the original.

Figure 4.12 shows the improvements of the PQC corrections only on the mean state. Interestingly, we found the PQC improvement mainly comes from the mean

correction rather than the covariances, suggested by the fact that all update methods converge towards PQC_K, and the changes in the covariances by PQC lead to suboptimal performance. From the figure, it is clear that when the covariance is unchanged, the maximum improvement was provided by PQC_AmO, and that PQC_H contributes the least improvements (but still improve a lot from the original PQC_H). However, it should be noted that PQC_K is much computationally efficient than the others and does not require the extra step to keep the original covariance matrix.

4.4.5 PQC with suboptimal DA Systems

To remain relevant to applications in an operational environment, we would like to explore PQC in suboptimal conditions including imperfect model, flawed observing system, DA window, and various sizes of ensemble and observing network. For the rest of the paper, we will be using 6-step PQC, which also improves the quality of the forecast (though not as much as 21-step PQC), but is less computationally expensive.

In high dimensional complex chaotic systems, we do not have the luxury of using a sufficiently large ensemble size because of the limited computational resources. It is important to examine the performance of PQC with different numbers of ensemble members. We tested a wide range of ensemble sizes (from 5 to 640), shown in figure 4.13 . The experiments with ensemble size less than 40 suffered, as expected, from filter divergence, since no localization was applied. Surprisingly, PQC

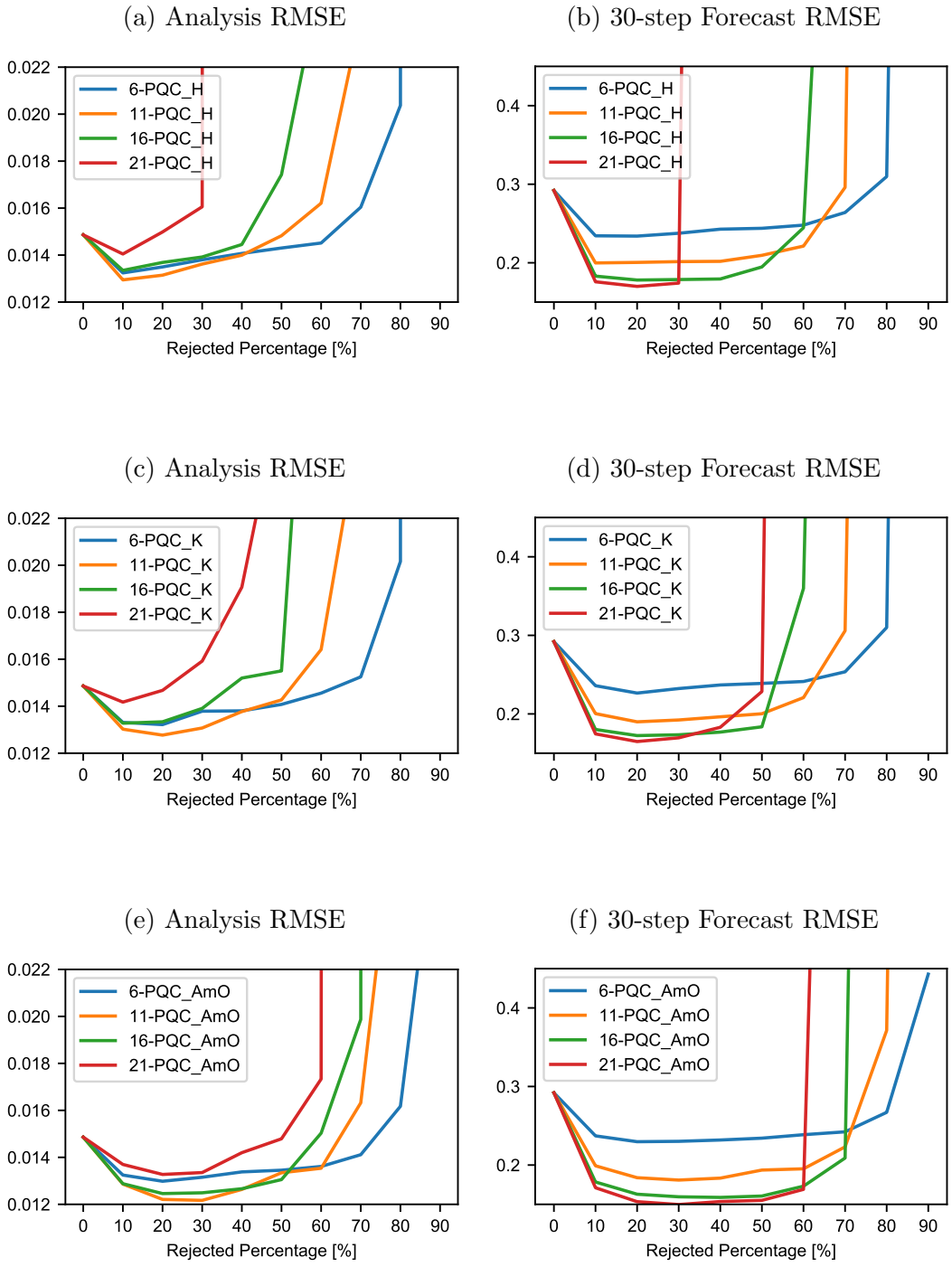


Figure 4.12: PQC_H (top), PQC_K (middle), and PQC_AmO (bottom) performance in terms of (left) analysis RMSE and (right) 30-step forecast RMSE when only the mean is updated.

can reduce the analysis error significantly even though the filter still diverges. With around 40 ensemble members, ETKF works well, and PQC improves the quality of analysis as expected, whereas the analysis error shows a slight increase as the ensemble size doubles monotonically beyond 40, which is a characteristic unique to the family of ensemble square root type of filter and has been documented in literature (e.g. Lawson and Hansen, 2005; Ng et al., 2011). The prevailing explanation is the high probability of having ensemble outliers leading to ensemble clustering, which can be ameliorated by applying additional random rotation to the transform matrix, but this is not of our interest in this study. The PQC analysis error also increases with ensemble size, but still smaller than the control error except for PQC_H with a size larger than 320.

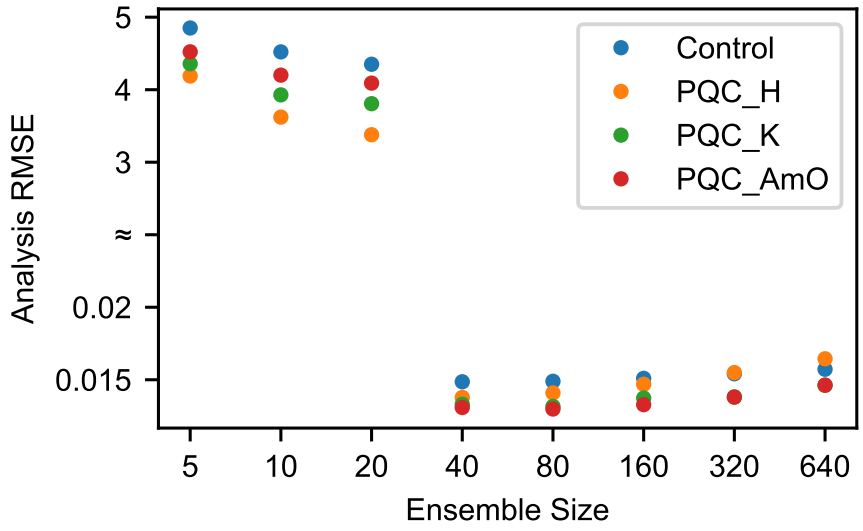


Figure 4.13: PQC performance in terms of analysis RMSE as a function of ensemble size.

In a "realistic" application, the state is only partially observed, and hence it is important to show whether PQC also benefits the system even when the size of

the observing network does not match the size of the model. For the extreme case where only 5 observations are available (12.5% observed), PQC_H seems to degrade both the analyses and the forecasts while PQC_K and PQC_AmO still improve the forecasts. This result again shows that PQC operates on the AIs rather than the observation itself since PQC_H degrades the system by changing the gain matrix $\boldsymbol{K}$ significantly. For the observing network with a size larger than 5, the improvements with PQC_K and PQC_AmO are somewhat similar, showing that PQC performs well with a wide range of observing network size. In Figure 4.14 (c) and (d), we show the PQC performance sensitivity to the length of the DA window, which determines the nonlinearity of the model increments. Both the analysis and the 30-step forecast RMSE are reduced by PQC although showing the same increase as the control with increasing DA window.

So far, we have been using observations with errors consistent with $\boldsymbol{R}$, meaning the observing system is flawless. However, in the real world, the observational error covariances are never truly known but are just estimations which deviate from the truth, no matter how sophisticated the techniques applied. It is worth noting here that an ensemble forecast sensitivity method was proposed recently that provides a way for fine-tuning of $\boldsymbol{R}$ (EFSR; Hotta et al., 2017b). In addition to the inaccuracy in error covariances, observational bias may pose an even greater danger of degrading the filter performance. To examine PQC under the influences of flawed observing systems, random errors are added to every observation of the 10th grid observation and biases to observation of 30th grid separately in two sets of experiments similar to those in Liu and Kalnay (2008). The average EFSO impact for each grid is shown in

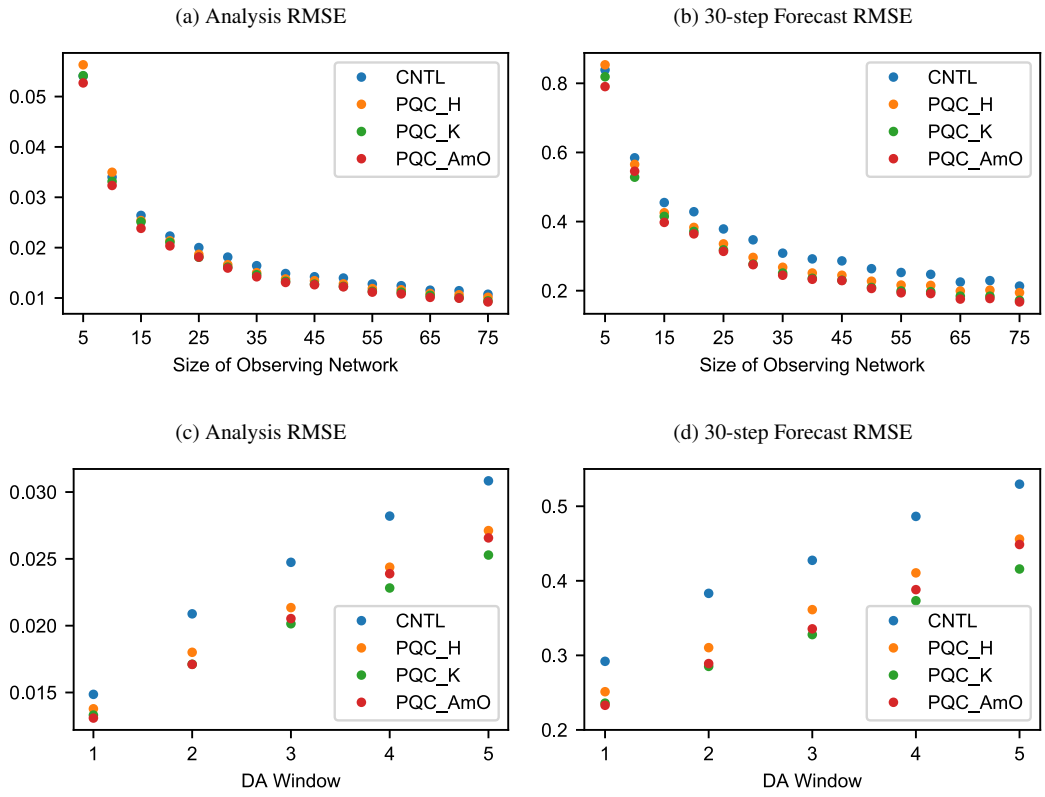


Figure 4.14: Performance of PQC_H, PQC_K, and PQC_AmO in terms of (left) analysis RMSE and (right) 30-step forecast RMSE as a function of: (top) size of observing network and (bottom) DA window.

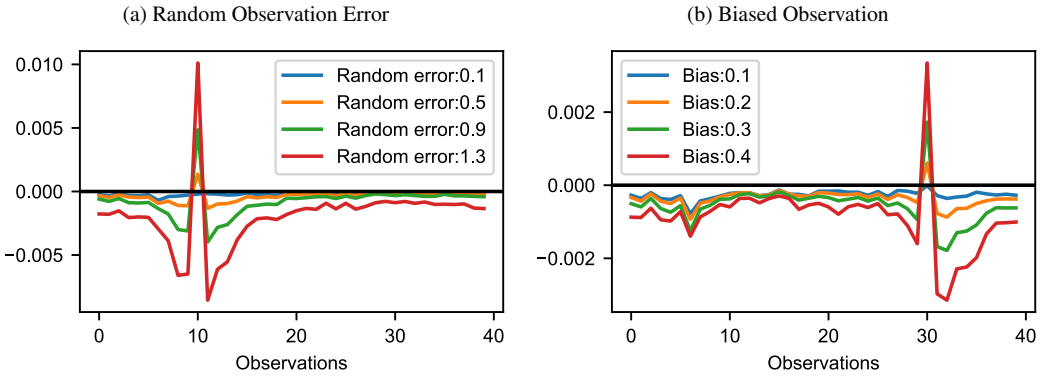


Figure 4.15: Mean EFSO impact for each observation through out 5000 cycles of experiment that assimilates flawed observations: (a) observation at 10th grid point with various random error larger than specified in observational error covariance matrix $\boldsymbol{R}$ and (b) observation at 30th grid point with various systematic bias.

Fig 4.15 for a range of the flaw magnitudes. The flawed observations are successfully identified, indicating that EFSO can also be used as data selection or QC method as described in Lien et al. (2017). Both the detrimental impact of the flawed grid and the beneficial impact of the neighboring grids increase with the magnitude of the flaw. Fig 4.16 (a)-(d) summarizes the responses of PQC to both types of the observational deficiency, and it is clear that analysis and forecast error reduction by PQC increases with the magnitude of the flaws. For biases more significant than 0.5 the filter suffer from divergence, but PQC becomes stable for this bias, which is similar to what we observe in some other border cases (not shown).

Besides flawed observations, model error is another source of error that deserves particular attention. We examine the response of the control and PQC by setting the forcing term F to be slightly different from the nature run, which is both

the verifying truth and where the observations were drawn from. Fig 4.16 (e) and (f) visualize both the control and PQC error in analysis and forecast increases with the model error that will eventually lead to filter divergence. It is shown that PQC improvements are almost invariant with the model error.

In this section, we have shown that PQC improves the quality of analysis and the forecast even in suboptimal DA system. The improvement is even more significant when the imperfections are originated from flawed observations. In the biased observation case, it is shown that PQC provides the extra stability that avoids filter divergence. Additionally, we found that PQC is not sensitive to the quality of verifying analysis for EFSO computation since the forecast error is much larger than the analysis error due to the sub-optimalities in the system.

4.5 Summary and Discussion

In this study, we explore the characteristics of EFSO and PQC with different update methods and the sensitivities to the configuration using the simple Lorenz (1996) model with ETKF as the data assimilation system.

For the characteristics of EFSO, we explored the sensitivity of EFSO impact on the length of verifying lead-time and the choice of verifying truth. More importantly, we found an alternative explanation for the low beneficial percentage of FSO impact that is determined by the relative accuracy between the forecast and the observation.

We examine the performance of PQC_H, PQC_R, PQC_K, PQC_BmO, and PQC_AmO using various configurations of the EFSO lead-time and rejecting per-

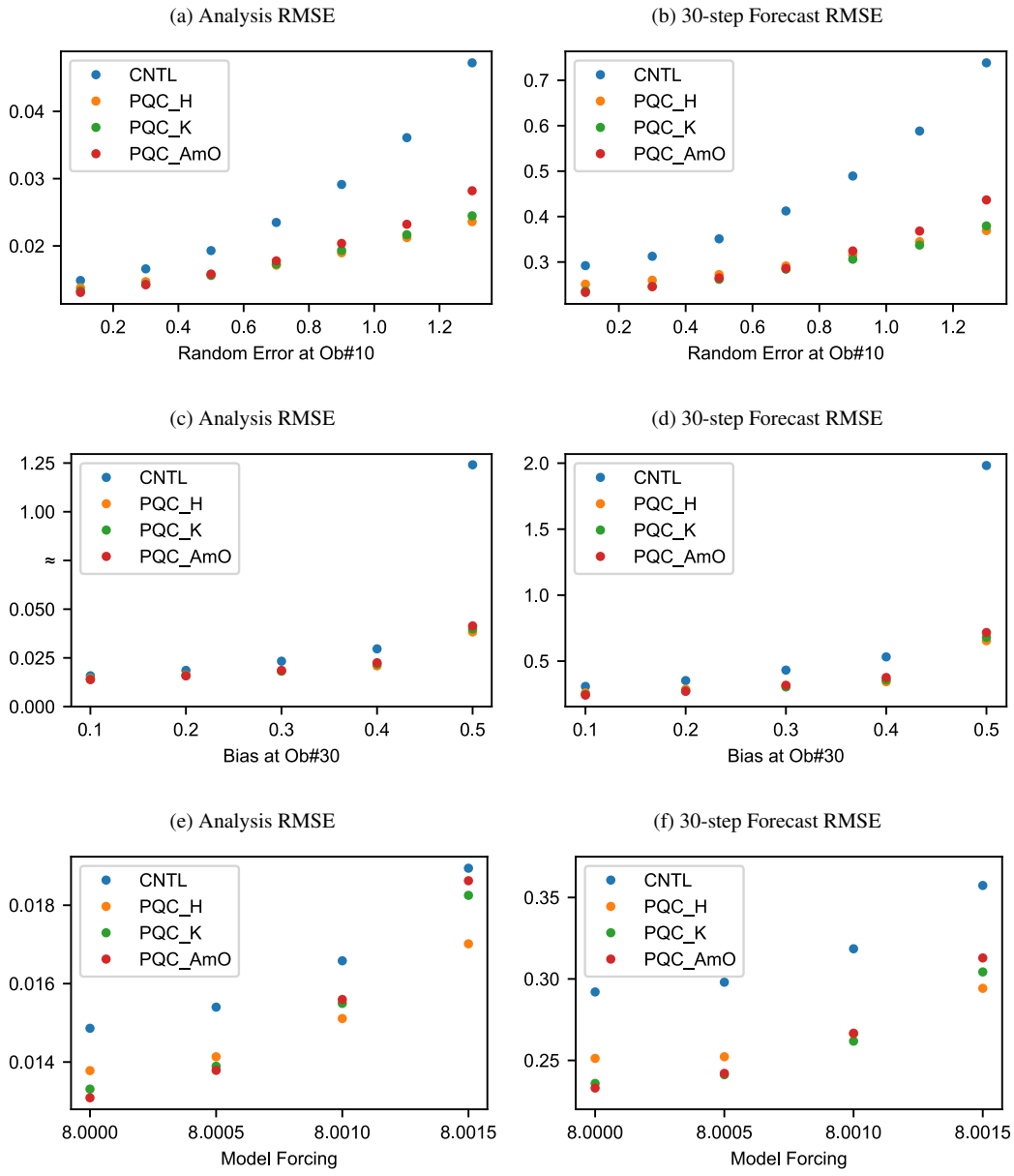


Figure 4.16: PQC_H, PQC_K, and PQC_AmO performance in terms of (left) analysis RMSE and (right) 30-step forecast RMSE in suboptimal DA system as a function of: (top) random error larger than specified in observational error covariance matrix $\boldsymbol{R}$ at 10th grid point, (middle) systematic observational bias at 30th grid point, and (bottom) model forcing F error.

centile. We show that PQC_H and PQC_R are suboptimal and computationally more expensive than other methods for repeating the analysis process. While the PQC_AmO and PQC_BmO can easily lead to filter divergence due to over-confidence, we found that PQC_K has the best performance because it rejects the analysis increments (AIs) contributed by detrimental observations without changing the gain matrix, hence it is consistent with EFSO. We find that in the Lorenz (1996) system, even in the absence of flawed observations, the PQC rejection of the 10% most detrimental observations significantly improves the forecasts. Between 10 % and 50 %, however, the denial of more observations have little effect, and beyond 50 % rejection, the forecasts deteriorate significantly. The improvement of PQC increases with the forecast length, but it saturates around 16-20 time steps.

We also examine PQC performance for suboptimal DA setup with varying ensemble and observing network size, DA window, biased observation, random error inconsistent with $\boldsymbol{R}$, and model error. The results show that all the sub-optimality leads to degradation of the control analysis and forecast quality, but PQC still improves the quality even in the extreme cases of filter divergence. The improvement grows with the magnitude of the flaw in the observations.

We would like to point out that we deliberately avoided the issues associated with localization accounting for insufficient ensemble sizes, and dynamical systems with multiple timescales (all of which will be encountered in the GFS model used in Chapter 6.). For multiple time scale systems, it is clear that the optimal lead-time should be long enough to capture the error growth in the timescale we are interested in, but a shorter-than-optimal EFSO lead-time could still improve the system.

Also, to implement PQC with a serial type ensemble filter, the PQC_K method may not be optimal since the real contribution of each observation to the AI is determined by the intermediate gain matrix $\boldsymbol{K}$ and cannot be well represented by just the final $\boldsymbol{K}$. On the other hand, in a variational DA system where the gain matrix is not available, the PQC_AmO or its variant may be more appealing than in ensemble system due to its simplicity and lower computational cost.

In summary, we have demonstrated the beneficial impact of applying PQC in Lorenz (1996) system. PQC improves the system in the presence of flawed observations, showing its potential of being a care-free and automated QC scheme on the fly with the DA system. More importantly, even in a flawless system, PQC still improves the quality of analysis and forecast by eliminating "harmful" growing components of analysis increment. We will continue to explore PQC in Chapter 6.

In the next chapter, we will examine and demonstrate the usefulness of FSO applications in data monitoring and selection.

Chapter 5

Realistic-model Experiments with NCEP GFS Model I: Online Monitoring Tool and Data Selection

5.1 Introduction

In Chapter 4, we demonstrate using the Lorenz (1996) system that EFSO is capable of identifying the artificial errors added to the observations. There are more complexities in realistic DA systems, including the spatiotemporal inhomogeneity of the observation distribution and various types of observations using different measuring techniques. Going from the simple system to quasi-operational system in this application is a major step forward.

On the average, there are about 50% detrimental observations, which is also true even for a very beneficial subset of observations. It is almost impossible to find a subset of meaningful size containing pure beneficial or detrimental observations. By aggregating data into subsets, we are mostly looking at the statistical properties of the subsets and their expected contribution to the forecast, where a net beneficial subset still contains observations that tend to reduce more than increase the forecast error. The generic FSO technique provides an efficient way to identify the beneficial as well as the detrimental subset of observations. As explained in Chapter 3, we proposed to include the aggregation of EFSO into data selection process. However,

we will demonstrate only the subtraction of identified detrimental subset from the assimilated observations due to the limited computational resource.

5.2 Experimental setup

A brief description of the experimental setup in this chapter is provided below, including the two DA systems used to generate EFSO datasets for the two periods of study and the EFSO setup.

Two sets of the DA system were adopted in this chapter for EFSO impact evaluation for the two experimental periods, 00Z Jan 01, 2012 - 18Z Feb 09, 2012 and 00Z Jun 1, 2017 - 00Z Jun 27, 2017, respectively. We deliberately selected winter season and summer season for the two periods. To stay as close as possible to the configuration in NCEP operation, the Gridpoint Statistical Interpolation (GSI) 3D-Var based ensemble-variational (En-Var) hybrid DA system (Wang et al., 2013; Kleist, 2012; Kleist and Ide, 2015a,b) is used with 80 ensemble members. The En-Var system linearly combines the static and the ensemble background covariance in the cost function with a prescribed weight. In this study, the static part is set to contribute only 0.25 and 0.125 (*beta1_inv*) to the hybrid background covariance for the 2012 and 2017 dataset respectively. Usually, the variational component is in a higher horizontal resolution while the resolution is only half of that in the ensemble component to lower the computational cost. The main difference between the two experiments is the update of the GSI (from v2012 to v2016) so that data from the more recent observing systems can be assimilated, and the data selection, as well as

routine QC, are also updated. Another major difference is the ensemble component of the hybrid system. For the 2012 period, we adopted Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007) as in the system used in Hotta et al. (2017a). It is the same as the ETKF used in Chapter 4, but with localization to suppress sampling noise due to the rank deficiency in background error covariances. The operational serial ensemble square root filter (EnSRF Whitaker et al., 2008) is used in the 2017 period to compare the difference. The covariance localization was based on the fifth-order polynomial localization function (Gaspari and Cohn, 1999) with 2,000 km of horizontal cutoff length and two scale height in the vertical. The inflation was the same as in Wang et al. (2013) applying both relaxation-to-prior-spread (RTPS; Whitaker and Hamill (2012)) multiplicative inflation with a relaxation parameter of 0.85 and an NMC-type, additive inflation with a scaling parameter of 0.32. For bias correction, the corrections for mass, scan angle, and emissivity are turned on for the 2017 dataset, but correction of emissivity was not available for the v2012 system. Also, following the configuration of Hotta (2014), the scan angle correction is also turned off in generating the 2012 dataset. Despite all the differences, the identified detrimental channels for the two datasets turn out to be quite similar as we will see in the result section.

For the forecast model component, we used the GFS from NCEP with horizontal resolutions of T254 for deterministic and T126 for ensemble forecasts for the 2012 period. In the 2017 period, a higher horizontal resolution of T670 and T254 is used for deterministic and ensemble forecasts. The GFS model has 64 vertical levels. The observations assimilated in this study is the same as those assimilated

Table 5.1: Experimental setup for the two datasets

	EXP. 2012	EXP. 2017
Period	00Z Jan 10, 2012 - 18Z Feb 09, 2012 (Winter)	00Z Jun 1, 2017 - 00Z Jun 27, 2017 (Summer)
Model	GFS T254/T126 L64	GFS T670/T254 L64
DA	LETKF/3D-Var Hybrid GSI v2012	EnSRF/3D-Var Hybrid GSI v2016

into the operating system at that time. The differences between the two datasets are summarized in Table 5.1.

For the EFSO computation, the Moist Total Energy error norm (MTE; Ehrendorfer et al., 1999) was chosen to measure the observational impact. A simple scheme for advection of localization function with the horizontal wind vector is used (Ota et al., 2013). The forecasts were verified with the high-resolution final GDAS analysis.

5.3 Results: Online Monitoring Tool

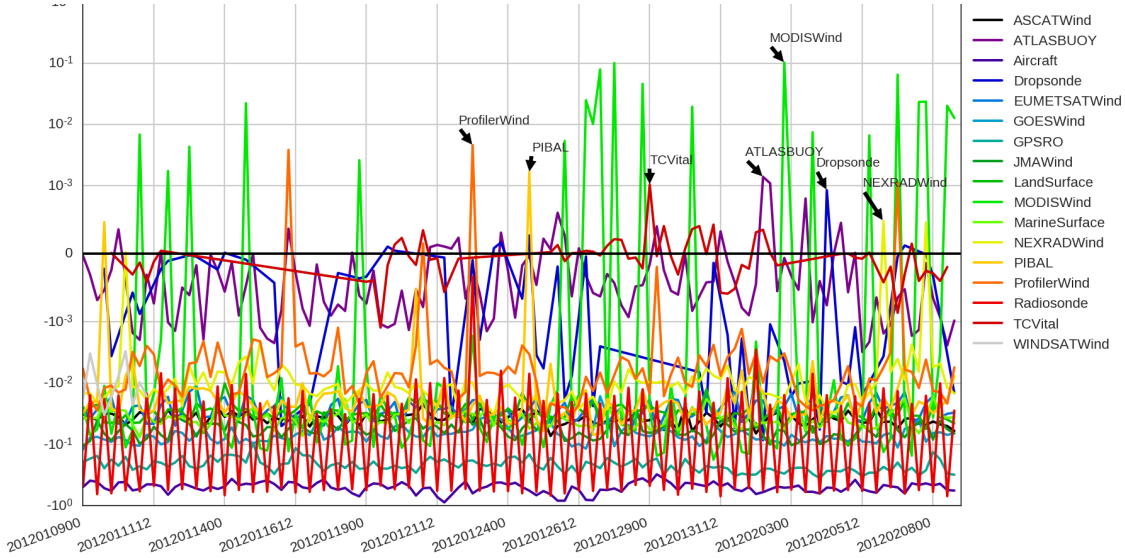
It is very common in the literature to aggregate the (E)FSO impacts over a long period with respect to observing systems (e.g., Lorenc and Marriott, 2014; Ota et al., 2013; Hotta et al., 2017a). This is owing to the desire in comparisons of

the average forecast impact between observing systems and it also reflects the customary way of viewing the observational impact on forecast improvement in OSEs. We are used to the aspect that each observing instruments have a constant impact evaluation over time unless there are changes in the instrument or algorithm level. We found that simple aggregation of 6-hour EFSO impact for each observing system as a function of time as shown in Figure 5.1 is very informative and serves as a first order online monitoring tool on the quality of assimilating the observations in complement to current QC monitoring tool. Several operational centers reported that most of the routinely ingested observing systems are beneficial for short-term forecasts (24 hours) by aggregating (E)FSO impact over a long period of time (Lorenc and Marriott, 2014; Ota et al., 2013; Cardinali, 2009). We also concluded the same in this study that most routinely ingested observing systems are overall beneficial (figure 5.1). It is clear that the top-3 beneficial non-radiance systems are commercial aircraft reports, GPS radio occultation (GPSRO) data, and radiosondes. With EFSO, the fluctuation of the impact of radiosondes is due to the fact that there are much fewer launches at 06Z and 18Z than at 00Z and 12 Z UTC. In addition, we found that EFSO without aggregation in time identified several observing systems became detrimental at different occasions, in what we called detrimental episodes. These systems include Atlas buoys, dropsondes, MODIS polar winds, NEXRAD radar winds, pilot balloons, and profiler winds in alphabetical order. For the satellite radiance instruments, it does not show as many detrimental episodes as the non-radiance systems. However, if we display the same monitoring for the *individual* channels, it does reveal even more detrimental episodes in specific channels (not

shown). With this figure, we show that EFSO can reveal such detrimental episodes from observing systems that are beneficial in general. We argue that these detrimental episodes are highly flow-dependent (supported later in this section) and difficult to identify without EFSO. This is an example of how (E)FSO could be used as an online monitoring tool for tracking the quality of assimilating each observation. And EFSO provides innovative aspects in evaluating the impact of an instrument that is not constant over time but depends on the flow condition and background quality. This monitoring also provides alarms for the forecasters about the upcoming degraded model forecast due to the detrimental observations. Besides, long-period accumulations the EFSO data can assist the model and observation developers to improve their products.

Among those observing systems associated with detrimental episodes, MODIS polar winds contributed not only a considerable number of the events but also the largest ones. In figure 5.2, the geographic distribution of the 6-hour impact of non-radiance observations is shown for 18Z Feb 6, 2012, one of the more massive detrimental episodes. It can be quickly noticed that there is a considerable portion of detrimental observations, which is consistent with other studies that the average detrimental percentage of observations is nearly 50%, though the exact percentage may vary from cycle to cycle. Although given the fact that close to 50% of the observations were identified to have a detrimental impact, most of the regional net impacts of all the nearby observations have a beneficial effect, especially for those regions that are densely observed such as North America and Europe (Figure 5.3). The discussion on the reason why so few of the observations are beneficial

(a) Non-Radiance observing systems



(b) Satellite Radiance instruments

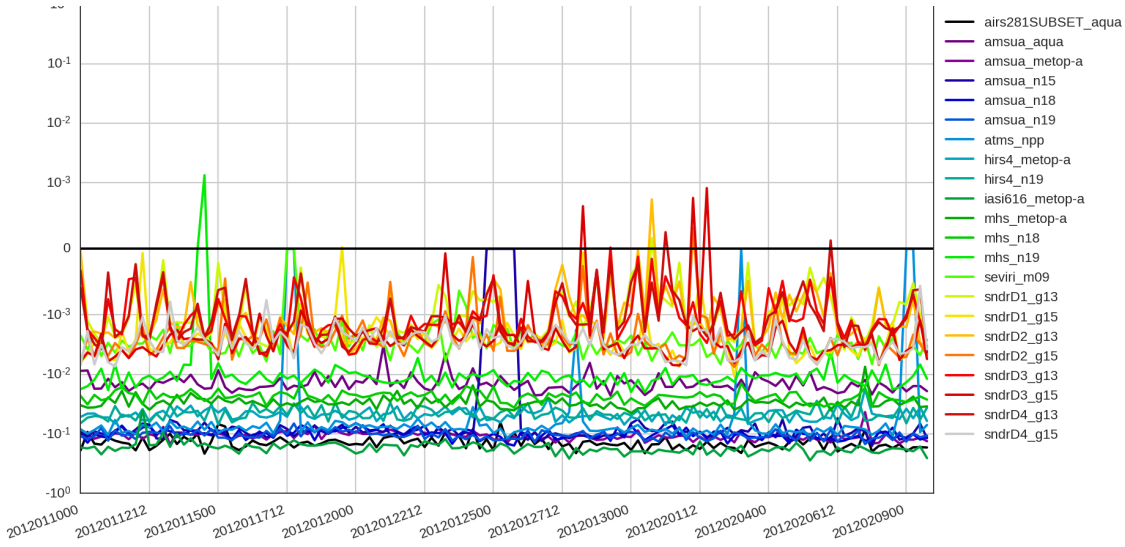


Figure 5.1: The time evolution of 06-hr total impact of each non-radiance observing system (top) and Satellite Radiance instrument (bottom) for 1-month period. Positive (negative) values represents the detrimental (beneficial) impact. The annotation with arrow indicates the systems associated with detrimental episodes.

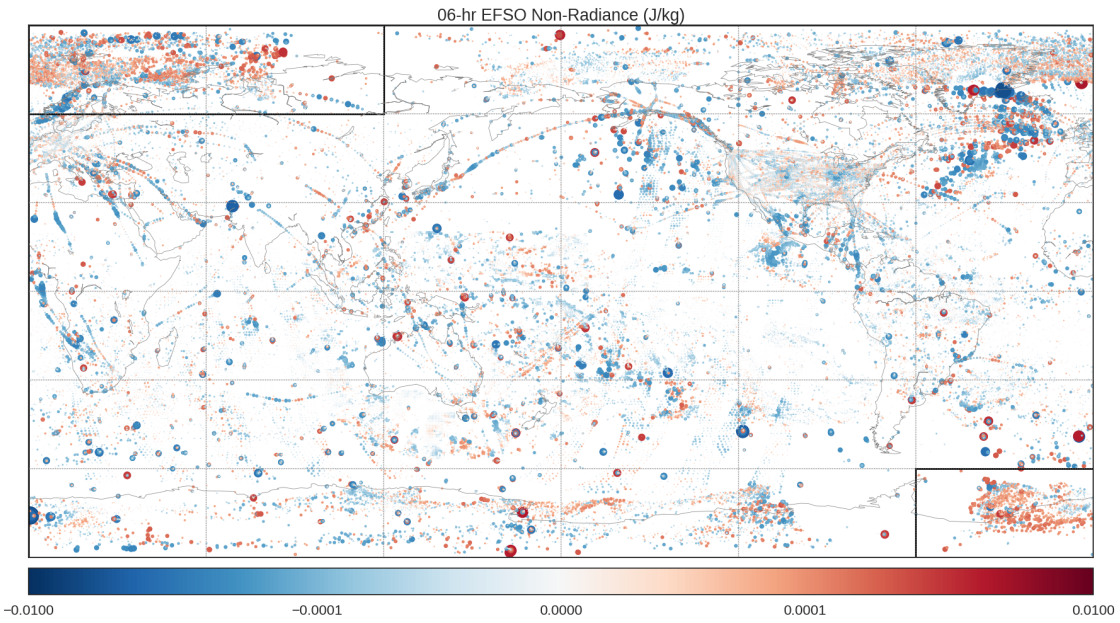


Figure 5.2: Geographic distribution of the EFSO impact of each non-radiance observation on the 6h forecast at 18Z Feb/06/2012. Each dot represents one observation. Blue indicates a beneficial and red a Detrimental observation. The size is logarithmically proportional to the magnitude of the impact. Black boxes in the South Atlantic and Eurasia enclosed the clustered detrimental MODIS polar winds.

is offered in Chapter 4. In this particular case shown in figure 5.2, there are two major regions, located north of Eurasia continent and the southern Atlantic Ocean that are clustered with detrimental observations. These observations are mostly from MODIS winds, which contributed to the detrimental episodes. We emphasize again here that the MODIS winds are beneficial on average and these transient flow dependent detrimental observations are difficult to identify without (E)FSO.

Finding that the MODIS winds are one of the significant contributors of detri-

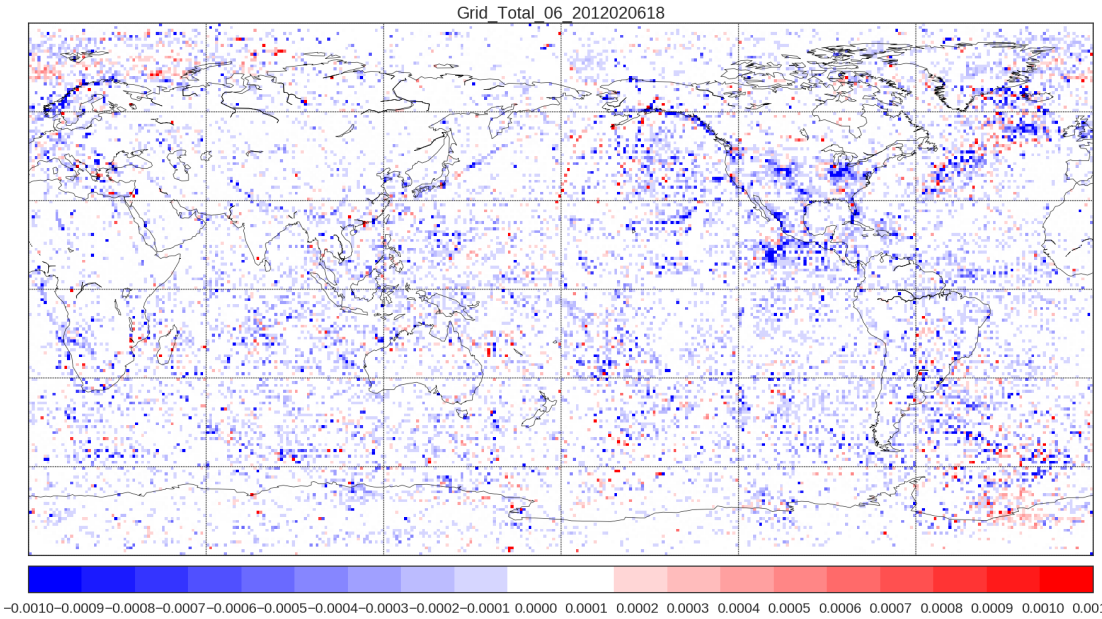


Figure 5.3: Geographic distribution of total 6-hour EFSO impact $[Jkg^{-1}]$ of all observations within $1^o \times 1^o$ grid box at 18Z Feb/06/2012. Blue (red) represent beneficial (detrimental) impact.

mental observations was quite surprising, given the fact that they are highly valuable in providing critical wind profiles over near-polar regions, where other non-radiance observations are rare. Most major NWP operational centers, including NCEP, NASA, ECMWF, JMA, and UK Met Office, have reported an increase in average forecast skill scores after assimilation of MODIS winds (Sarrazin and Zaitseva, 2004; Le Marshall et al., 2008; Bormann and Thépaut, 2004; Kazumori and Nakamura, 2004; Riishojgaard and Zhu, 2004). The usual approach in measuring forecast impact from assimilating specific observation subset is through OSEs as mentioned earlier. The reports of the beneficial impact of MODIS winds estimated

in OSEs are consistent with aggregating EFSO impact in time. However, we have shown that there is a subset of MODIS winds with detrimental impact hidden among the beneficial ones, demonstrating the usefulness of EFSO in monitoring the quality of assimilation of observations, and the same goal would be practically impossible to achieve with OSEs.

5.4 Results: Data Selection

Here we demonstrate how a generic FSO can identify a detrimental subset of the observations.

5.4.1 Detrimental subset of MODIS winds

Here we demonstrate how long-period EFSO statistics can be used to assist data selection. We begin with the MODIS winds by aggregating EFSO impact with multiple variables. It is possible to obtain the flow-dependent conditions for those detrimental observations appeared in figure 5.1 and 5.2. We show in figure 5.4 that this can be done by aggregating EFSO impacts against desired variables, namely innovation and the observed wind direction in this case. The observations are separated into subgroups according to the type and the hemisphere it was located. The wind components (U and V) were also separated, and a huge difference between the two was found. The dominant pattern is what we called innovation bias, which means that the sign of impacts (either beneficial or detrimental) depends strongly on the signs of innovation. Also, the innovation bias also depends on the observed

wind directions. One interesting feature is that the pattern of cloud tracking winds was very similar to that of water vapor tracking winds in clear sky, while the water vapor tracking winds in the cloudy sky do not share the similarity. It is worth noting that the same analysis was applied to geostationary satellite winds including GOES, JMA, and European satellites, which all shared the same feature tracking algorithm, but none of these biases appeared (Figure 5.5). MODIS winds are one of the examples of this innovation bias. There are other observing systems including some satellite radiance observations that demonstrated similar kind of innovation bias (e.g. Groff et al., 2017). While the cause of these biases needs further investigation and beyond the scope of this paper, we demonstrated that EFSO can find the flow-dependent conditions for detrimental observations.

5.4.2 Evaluation of the contribution of the radiance channels

The EFSO statistics also provide efficient guidance on radiance channel selection, which is especially true and useful for the advanced instruments that provide high spectral resolution, where channel-wise OSEs are practically impossible.

5.4.2.1 Channel evaluation for 2012 dataset

We first show the channel evaluation for several instruments of interest assimilated in the 2012 dataset. To demonstrate the usefulness of the generic FSO impact in data selection process, we are interested in the hyperspectral instruments as well as other multi-channel instruments that cover roughly the same spectrum such as

AIRS, IASI, HIRS, and GOES sounders.

In Figure 5.6a and 5.6b, the EFSO impact evaluation for each assimilated channel of the hyperspectral AIRS and IASI is shown together with its peak pressure level of the weighting function. As expected, most of the assimilated channels are primarily beneficial. However, there are plenty of channels that contribute a detrimental impact to the forecast. For IASI, those detrimental channels distributed within the wavelength range of $8 - 11\mu m$. Most AIRS channels in the same wavelength range are not assimilated, but many of the AIRS channels shorter than $8\mu m$ are (Not for IASI). Only two channels around $4.58\mu m$ show detrimental impact for AIRS. All these detrimental channels have strong weightings on the conditions at the low-level or surface, where the brightness temperature of clouds is very close to that of the surface, which indicates that low clouds may contaminate these channels.

We then show the geographical source of these detrimental channel observations in Figure 5.6c and 5.6d. Interestingly, it reveals specific regions where these detrimental measurements were taken. For AIRS, it is clear that the detrimental impact mainly comes from the northern tropical Pacific and Atlantic. Besides, these "detrimental" channels provide neutral to positive impact at higher latitudes, demonstrating again the detailed evaluation FSO provides. For IASI, the Australia continent and the oceans at low latitudes constitutes the source of detrimental channels. The surface radiation representation of the Australia continent is notoriously difficult due to its special silica surface condition.

We moved on to the multi-channel instruments covering the similar range in spectral space. Figure 5.7 is the same as 5.6, but for GOES sounders and HIRS.

Although the number of channels provided by these instruments is far less than that by AIRS or IASI, FSO impact can still provide useful information for data selection process. It shows there is a common detrimental channel #13 ($4.57\mu m$) across all three of the instruments. Channel #8 ($11.03\mu m$) for GOES13 sounder is also detrimental. We noticed that the wavelength of the detrimental channel #13 is very close to the one shown up in AIRS. Also, it is clear that the source is also from the northern tropical oceans, offering the opportunity to cross-validate the detrimental impact of the similar waveband from multiple instruments.

The coherent spatial structure of the source of detrimental channels and the cross-validation between instruments indicate that there is a common problem contributing detrimental impacts. This cause of the detrimental impact is unlikely to be instrument dependent but may be related to the model representation or the data processing procedure that is improper for the waveband.

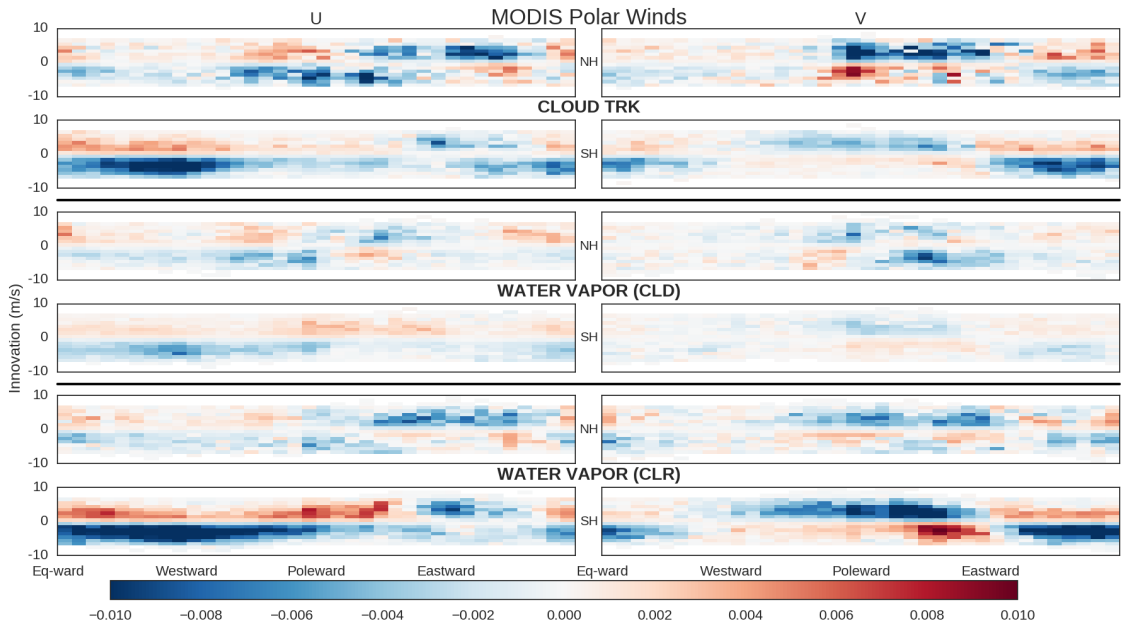


Figure 5.4: EFSO impact of MODIS polar wind against innovation and observed wind direction. Blue (red) represents beneficial (detrimental) impact. The left and right column display the impact of the zonal and meridional component of the wind respectively. The top, middle, and bottom groups of two rows are for the three types of MODIS wind observations: cloud tracking, cloudy water vapor tracking and clear water vapor tracking winds. The winds are also separated into two hemispheres where the Northern (Southern) Hemisphere displays in the top (bottom) row in each group.

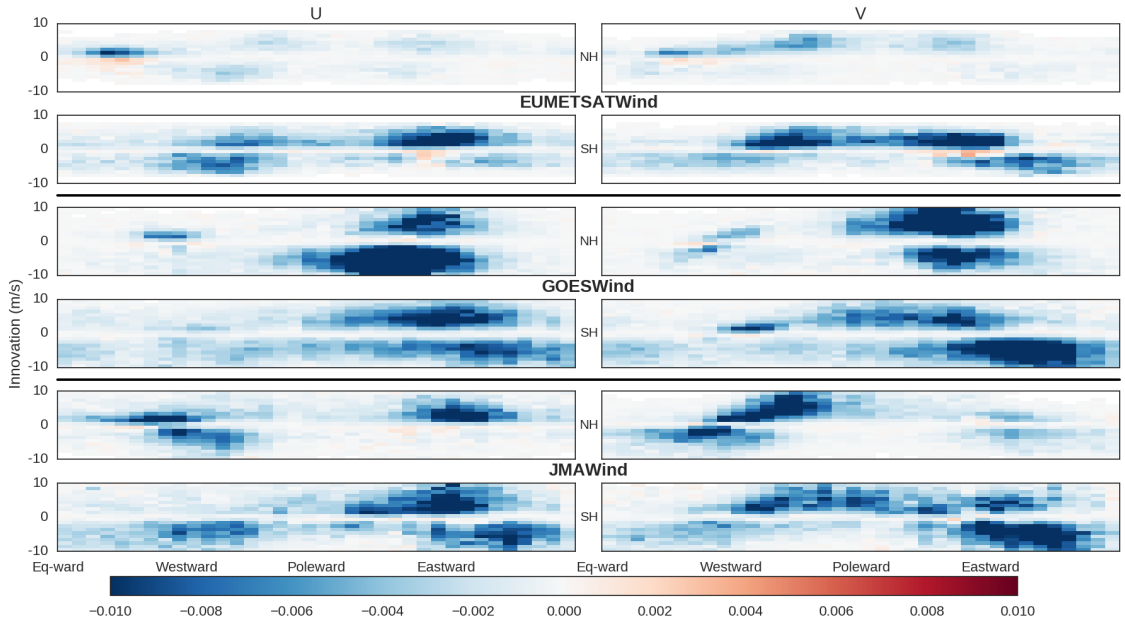


Figure 5.5: Same as figure 5.4, but for geostationary satellite winds. The top, middle, and bottom group represents data from EUMETSAT, GOES, and JMA.

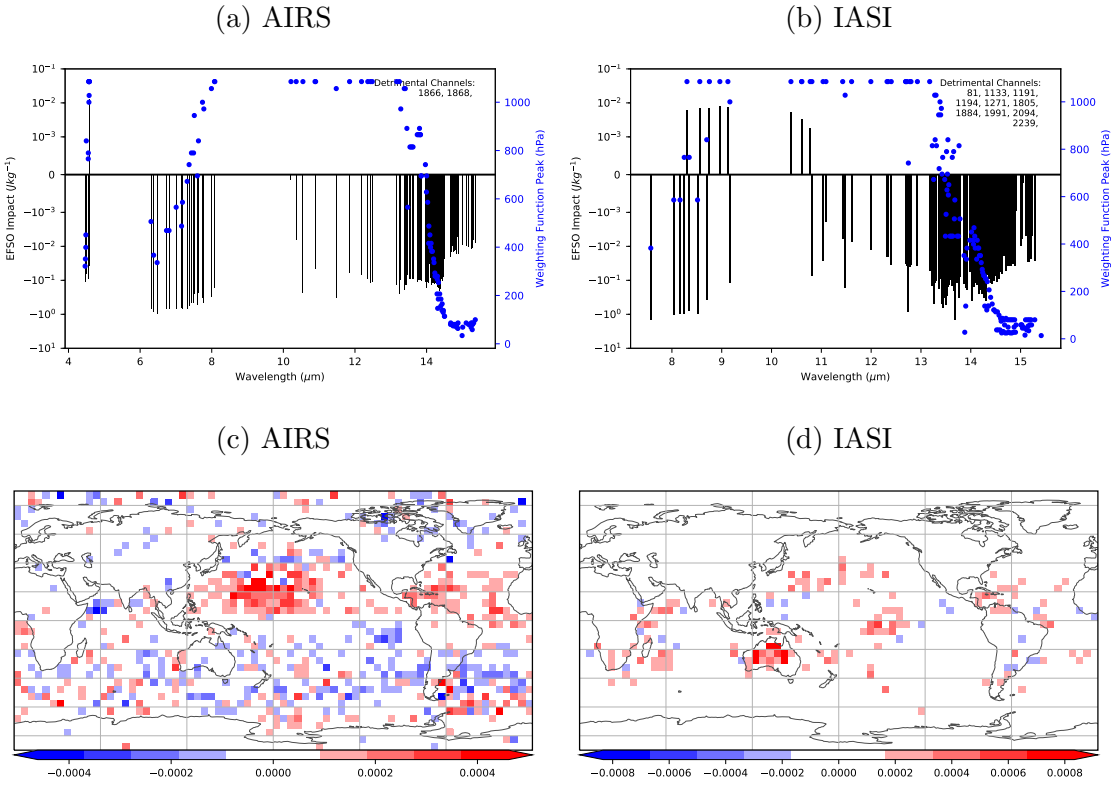


Figure 5.6: (Top) Channel wise 6hr-EFSO impact evaluation of 2012 dataset for hyperspectral instruments: AIRS and IASI. Each black line represents an assimilated channel and shows the EFSO impact $[Jkg^{-1}]$. The blue dots indicate the weighting function peak pressure level [hPa] of the channels. Net detrimental channels are listed on the upper-right corner. (Bottom) Geographic distribution of the impact $[Jkg^{-1}]$ of the net detrimental channels.

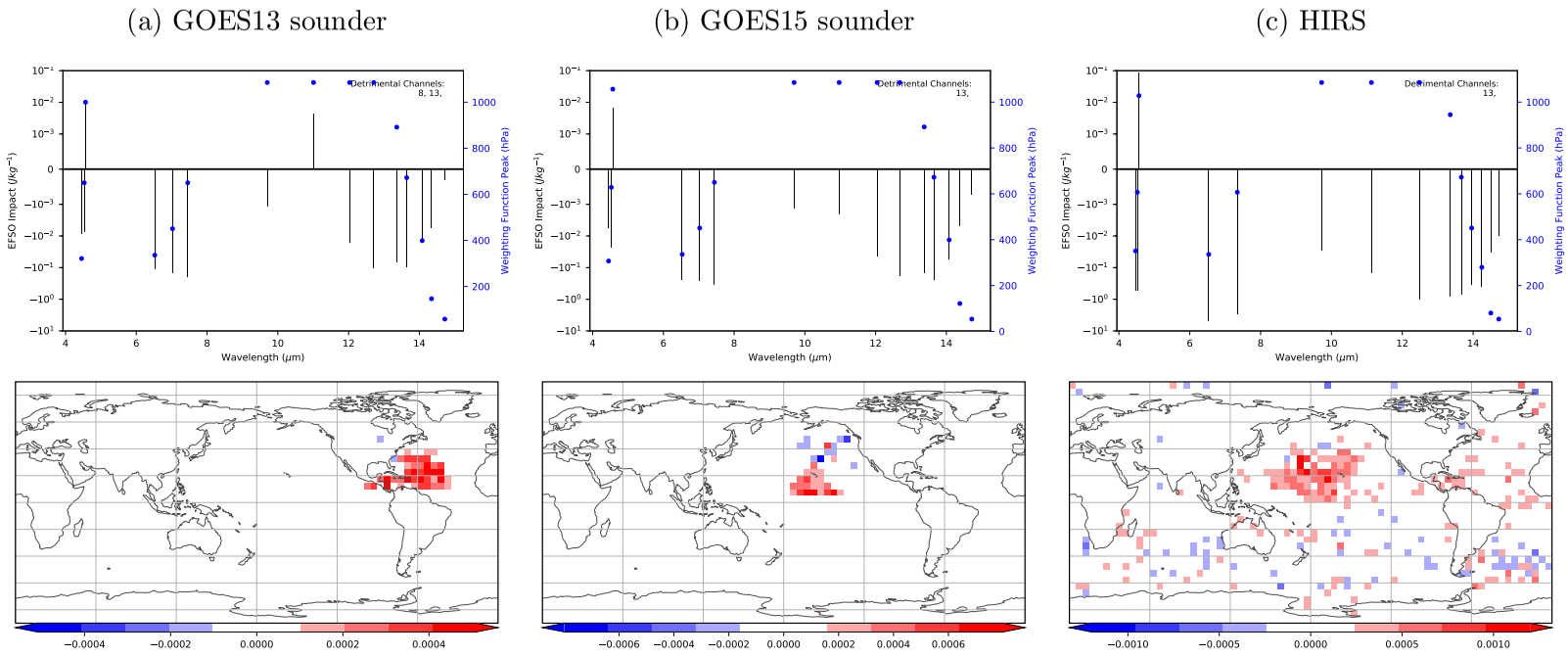


Figure 5.7: Same as 5.6, but for multi-channel instruments: GOES13 sounder, GOES15 sounder, and HIRS.

5.4.2.2 Sensitivity to EFSO verifying lead-time

It has become customary to compute the impacts with 24-hr lead-time to avoid the influence of diurnal cycle. However, there are several drawbacks to compute FSO with longer lead-time owing to the validity of linear approximations made in both the ensemble and variational approach and the additional computational costs.

We now show the same aggregation in spectral space but with 24-hr lead-time in Figure 5.8 and compare the differences between the aggregation results from the 6-hr (Fig. 5.6 and 5.7) and that from the 24-hr lead-time. Almost every detrimental channels identified with 6-hr EFSO shows up in 24-hr EFSO except for channel #81 from IASI (The detrimental impact of that channel is minimal). Hence, the detrimental channels with 6-hr EFSO are quite robust. There are additional detrimental channels identified in IASI and the GOES sounders. They are channel # 1027 for IASI and channel #1, #2, #9, and #15 for the GOES sounders. Most of them are also surface sensitive channels except for #1 and #2 for GOES sounders, which surprisingly are high-level channels.

This exercise shows that 6-hr EFSO provides robust identification of the detrimental channels that are not dependent on the verifying lead-time. Also, the 24-hour EFSO reveals a few more detrimental channels.

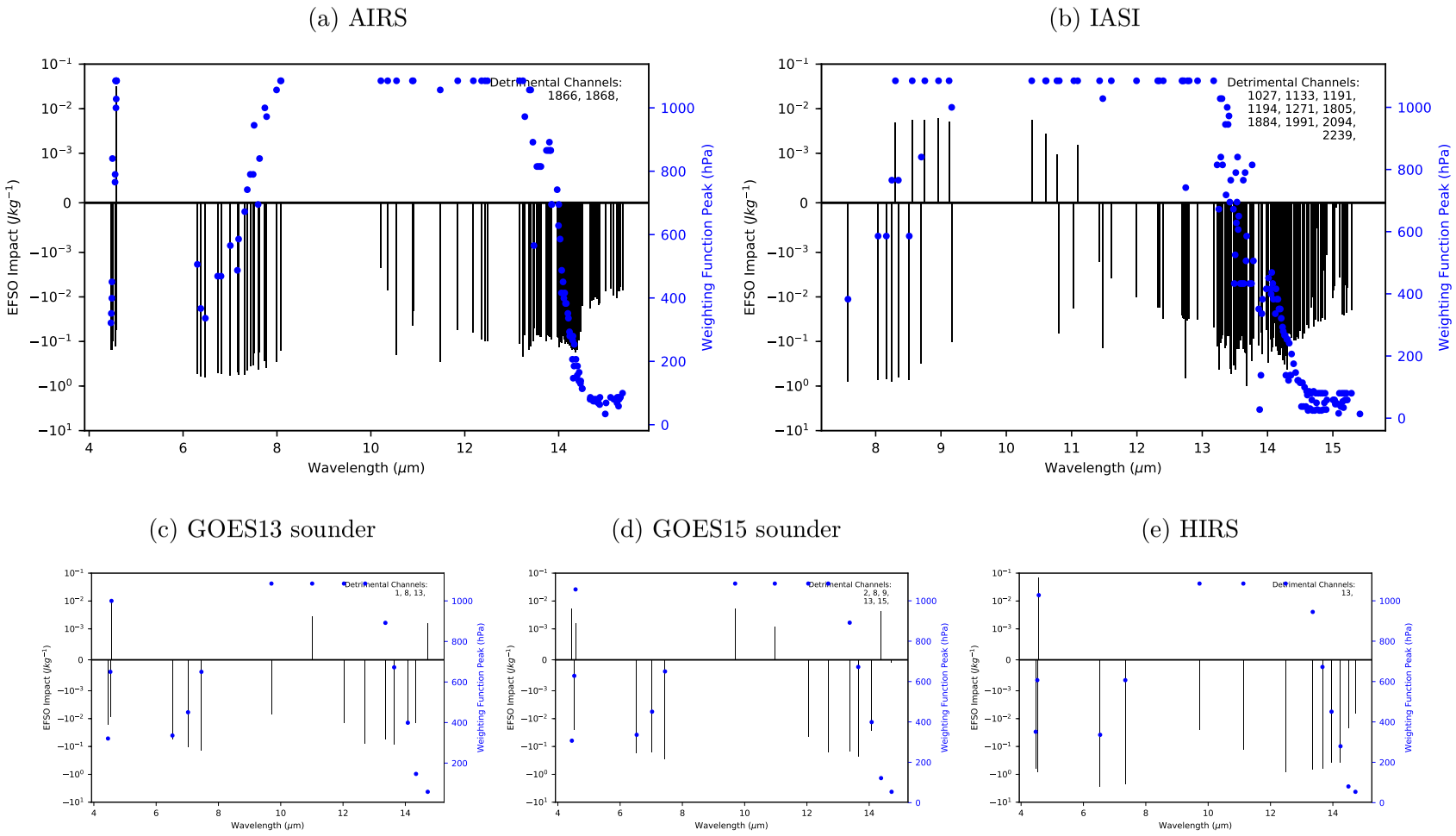


Figure 5.8: Channel wise 24hr-EFSO impact evaluation of the 2012 dataset for AIRS, IASI, GOES13 sounder, GOES15 sounder, and HIRS. Each black line represents an assimilated channel and shows the EFSO impact [Jkg^{-1}]. The blue dots indicate the weighting function peak pressure level [hPa] of the channels. Net detrimental channels are listed on the upper-right corner.

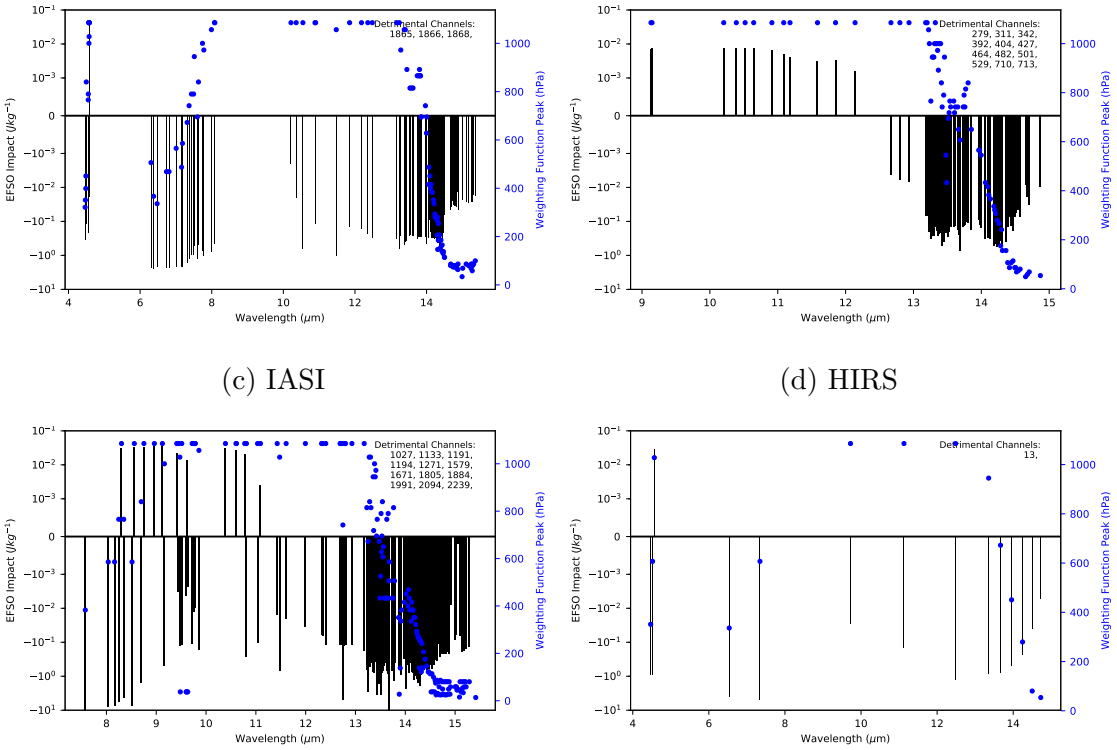
5.4.2.3 Channel evaluation for 2017 dataset

The channel evaluation for similar instruments in 2017 is presented here in Figure 5.9. CrIS, a counterpart of IASI, is a new instrument of particular interests in addition to the ones already assimilated in 2012. For AIRS, GOES sounder (not shown), HIRS, and IASI, the detrimental channels from the dataset of 2017 are almost the same as from that of 2012. This similarity shows again how EFSO statistics can provide occurring conditions for detrimental observations, which is unavailable otherwise, and minimize the number of OSEs required in the determination of assimilating observation set. A major noticeable difference is that there are few tens of the channels added to the assimilation list between $9 - 10\mu m$ and the majority of them are beneficial except for channel # 1579 and # 1671. There is one additional detrimental channel (# 1865) from AIRS whose wavelength is also around $4.58\mu m$, confirming again the existence of a common issue that affects all instruments. For CrIS measurements, we observe that all channels with wavelength longer than $12.5\mu m$ show strong beneficial impact as in IASI and AIRS. However, it is quite surprising to see channels with a wavelength shorter than $12.5\mu m$ are all detrimental.

It is clear that the EFSO identified detrimental channels are robust and consistent across different years with all the changes in the DA and QC system.

(a) AIRS

(b) CrIS



(c) IASI

(d) HIRS

Figure 5.9: Same as 5.8, but for 6hr-EFSO impact of AIRS, CrIS, IASI, and HIRS in the 2017 dataset

5.4.3 Forecast verification

To verify the EFSO-identified detrimental channels, we perform a data-denial experiment that rejects the identified channels to see if the forecast is improved. The rejected channels and the corresponding wavelength of each instrument are listed in Table 5.2. Recall the main detrimental impact is mostly from the Tropics, the relative forecast error reduction is visualized in Figure 5.10. The forecasts for winds are generally improved throughout 6-7 day forecasts at different levels with maximum improvement centering around 700 hPa on day-2. The relative humidity shows similar improvements as the winds. The temperature improvement lasts for 3 days, shorter compared to other variables. Overall, the forecasts improve on the order of 1% for each variable.

We show that the forecasts can be improved as much as 1% by merely rejecting 16 EFSO-identified detrimental channels out of hundreds. The significance of these denial experiments is a simple demonstration of the usefulness and accuracy of the EFSO impact evaluation. For data selection process, EFSO can be of greater use not by rejecting channels, but by adding channels back. Since EFSO depends on the quality of the analysis, we should not dump all the channels back into the system at once. Instead, the channels can be assimilated and evaluated with EFSO one small group at a time, and, in each iteration, we keep the channels with most substantial beneficial impact and compile a list of detrimental channels as in Rodgers (1996) but using FSO instead of DFS. This way, we might be able to come up with a better channel selection than just using the DFS introduced in Chapter 2. Limited by the

Table 5.2: A list of rejected channels in the forecast verification experiment

Instrument	Rejected Channels	Wavelength [μm]
AIRS	1866, 1868	4.58, 4.58
GOES15 SNDR	13	4.57
GOES13 SNDR	8, 13	11.03, 4.57
HIRS	13	4.57
IASI	81, 1133, 1191, 1194, 1271, 1805, 1884, 1991, 2094, 2239	15.04, 10.78, 10.61, 10.60, 10.39, 9.12, 8.96, 8.75, 8.56, 8.30

resource required to include the task in this study, we demonstrate in Figure 5.10 the subtraction of data can already provide a considerable amount of improvement.

5.4.4 Efficient EFSO Browsing Tool

By now, we have shown that FSO impact offers great flexibility to identify potential detrimental subset from the entire dataset by aggregating the impacts from any perspectives. It is, however, quite troublesome having to code-up every visualization scripts when needed or to output tens of thousands of figures for thorough analysis. To expedite the process of analyzing our EFSO dataset, we developed (and continue to improve) an efficient browsing tool built with Python and its free and open source libraries. The tool is designed with the following principles:

- Free of any dependencies on commercial software.

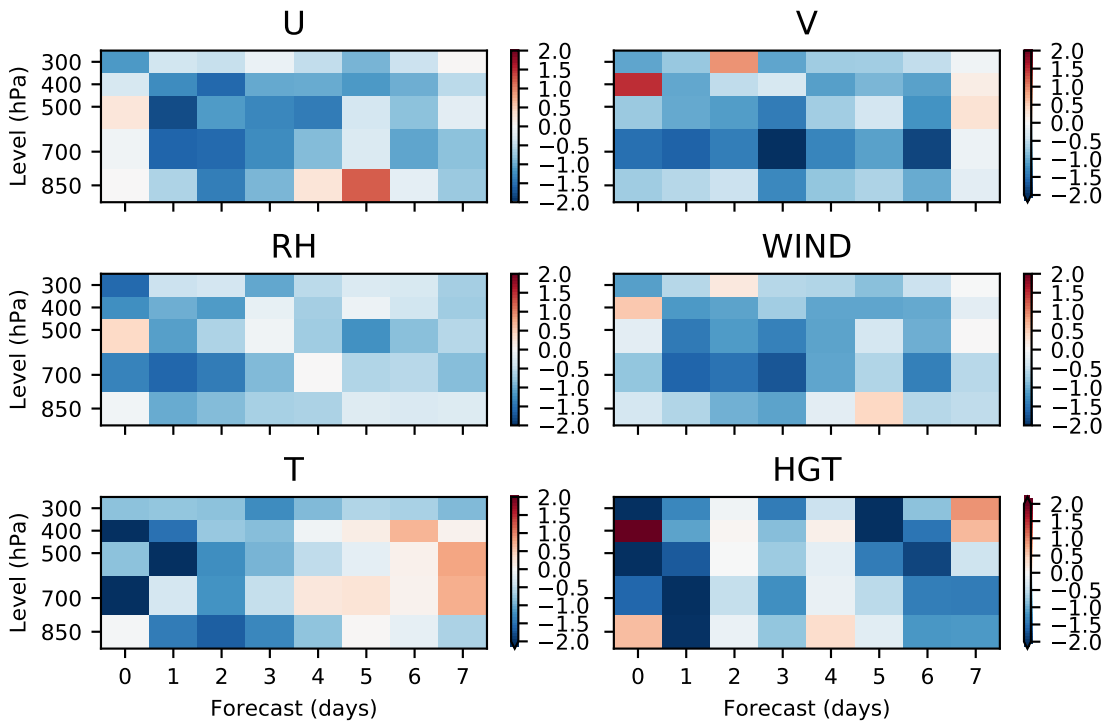


Figure 5.10: Relative 7-day forecast error reduction in the tropics by rejecting detrimental channels identified by EFSO. Variables including u- and v-component wind, relative humidity, vector wind, temperature, and geopotential height for various pressure levels.

- It should be easy to port to any machine that has access to Python and the open source libraries.
- The input format should be flexible to switch from one to another.
- The visualization should be done fast enough to allow exploring EFSO interactively.

It should be noted that the raw EFSO dataset for a month is about 75 Giga-

bytes, which does not fit into the memory. After some processing, we can shrink the size down to 9 Gigabytes by trimming off redundant information. The data of this size may be more feasible to read into the memory, but it can still take considerable time to load. Besides, it requires searching through the entire dataset to complete the aggregation process. There are two options:

1. Uncompressed data with Dask, a scalable parallel analytic library, to perform all data aggregation in real-time.

2. Pre-aggregate data into a limited choice of dimensions to avoid I/O overhead.

The two options are in fact choosing between the aggregation efficiency and the granularity of the data. In this implementation, we prefer the second option since the response time should be as short as possible in interactive mode and not all the subtle features are worth preserving. Now that the compressed data fits into machine memory, we utilized the data frame feature in Pandas, an open-source data analytics library in Python, for its advanced and simple-to-use data slicing and grouping functionality.

The interactive control panel makes use of the Jupyter interactive computing framework to query, aggregate, and visualize the EFSO data. We demonstrate a snapshot of the interface and visualization area of the tool in Figure 5.11.

5.5 Summary

In this chapter, we explored the data monitoring and selection applications. EFSO has shown to be a valuable online observation monitoring tool that provides

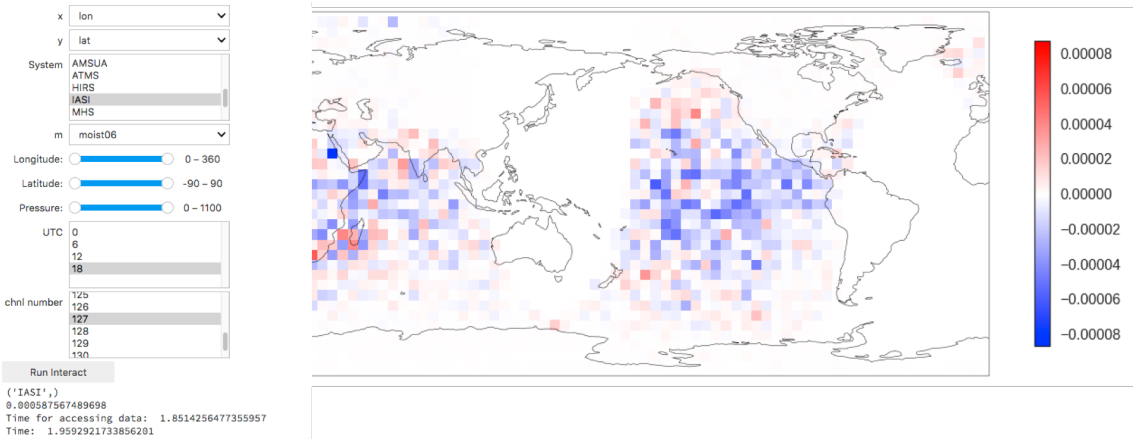


Figure 5.11: Screen shot of the python-based EFSO browsing tool. The control panel includes drop-down menus to select x and y axis of the plot, observing system, EFSO verifying lead-time, time of the day, channel numbers for EFSO impact aggregation. The sliders control the spatial region for display.

real-time or near real-time information of the impact of observations on forecasts and identify detrimental episodes potentially degrading future forecasts. Using the satellite winds and the radiance channels as examples, we demonstrate the data selection based on EFSO efficiently identifies detrimental subsets and the forecasts quality can be improved just merely by rejecting a tiny fraction of the assimilated channels identified by the aggregation of EFSO data. However, It is noteworthy that the full potential of the generic FSO data selection is not yet revealed by iteratively adding beneficial but rejected channels back. Also, to expedite the exploration of EFSO data, we developed a Python-based prototype of the efficient EFSO browsing tool.

In the next chapter, we will continue the study of PQC from Chapter 4, but

using the realistic GFS model.

Chapter 6

Realistic-model Experiments with NCEP GFS Model II: Proactive Quality Control

6.1 Introduction

The idea of PQC based on immediate EFSO impact was pioneered by Ota et al. (2013); Hotta et al. (2017a) using the same NCEP GFS-GSI 3D-Var based En-Var hybrid DA system introduced in Chapter 5, but both studies focused only on many selected cases of non-cycling PQC, meaning that the improved analyses were not utilized in the following DA system. In Chapter 4, we have explored cycling PQC using Lorenz (1996) system, and we use a close-to-operation system for demonstrating the usefulness of EFSO in data monitoring and selection. In this chapter, we will explore cycling PQC using realistic high-dimensional GFS model to see if we can obtain similar results as in Chapter 4. To reduce the complexity of the DA system and speed up the experiment, we switch to a system of low-resolution GFS model coupling with a pure (not hybrid) LETKF scheme. Also, in preparation for the implementation of PQC into operations, the shortcuts proposed in Chapter 3 for lowering the resource requirements of PQC in NCEP operation will also be tested.

The chapter is structured as follows: We first introduce the GFS-LETKF sys-

tem and the experimental setup in Section 6.2, followed by a description of results in Section 6.3 and the justification of PQC shortcuts for operational implementation under NCEP framework in Section 6.4. Lastly, a summary and discussion are provided.

6.2 Experimental setup

In this chapter, we utilize the GFS-LETKF system developed by Lien (2014) rather than the NCEP operational hybrid system to reduce the complexity and expedite the experiments. A brief review of the GFS-LETKF system is provided in this section.

The underlying philosophy behind the design of the system is to have a simple configuration of DA system coupled with the realistic GFS model to allow fast experiments to explore innovative data assimilation techniques. The DA system is the generic and simple LETKF core code developed and maintained by Takemasa Miyoshi (public Google Code platform: http://code.google.com/p/miyoshi/). It preserves the flexibility of switching GFS resolutions from T62 to T1534 (current resolution in operation) and the choice of observation operators using built-in conventional data operator (simple spatial interpolation) or the GSI from NCEP to ingest sophisticated data such as satellite radiances. For computational efficiency, we choose to perform the experiment with T62 resolution and use the built-in simple observation operator. Only 32 ensemble members are required with this low-resolution configuration.

The flow chart of the GFS-LETKF system is shown in Figure 6.1 (from Lien, 2014). The GFS forecast integration is carried using the native sigma/surface file formats. The 4D-LETKF analysis is performed in gridded file format. We only assimilate the conventional observations from the PREPBUFR dataset provided by NCEP. The error statistics used in the LETKF analysis is extracted directly from the PREPBUFR data. Adjusting to the low-resolution of our system, the observations are superobed/thinned to at most only one observation per model grid point for each data type and variable in one assimilation window which reduces the data density to one-third of the original. Some prognostic variables not available from the DA system, such as ozone concentration and surface temperature are obtained from the NCEP Climate Forecast System Reanalysis (CFSR; Saha et al., 2010). The CFSR dataset, which has a much higher resolution (T382) and assimilates many more observations (including satellite radiances) also serves as the verification truth for measuring the improvement obtained from PQC correction. The ensemble is initialized with the operational GDAS analysis ensemble of the same date from a different year. A mixture of adaptive multiplicative inflation (Miyoshi, 2011) and the relaxation to prior perturbation (RTPP; Zhang et al., 2004) is added to account for model error. A fixed horizontal length scale of $500km$ and vertical length scale of 0.4 scale height is chosen for the localization (R localization in Greybush et al., 2011) accounting for insufficient ensemble size. The experimental period spans the period 00Z Jan 01, 2008 to 00Z Feb 06, 2008 with the first five days used as DA spin-up period. This period is chosen simply based on the availability of an existing database from our group on the cluster.

EFSO computation is also implemented in the GFS-LETKF system. The same simple localization advection method and the moist total energy norm in Ota et al. (2013); Hotta et al. (2017a) is used. The PQC algorithm is the same as described in Chapter 3. With PQC using 6 hours as verifying lead-time (or 6-hr PQC) as an example, we first conduct standard DA cycles to obtain the verifying analysis 6 hours after the desired (current) PQC cycle. The EFSO impact of observations at the current cycle is evaluated using the verifying analysis valid 6 hours later. We then obtain the PQC correction by rejecting the observations based on the EFSO impact. Here in this chapter, we show only the results from cycling PQC-H analysis update method and the non-cycling case results of PQC-K update method. For the data-denial strategy, we follow the same threshold method used in Chapter 4 for the Lorenz (1996) system where a month of EFSO impact data is obtained from a control experiment, and a range of thresholds are determined by the top 10%-40% detrimental percentile of EFSO impact. A list of all the experiments performed and their rejecting threshold and other configurations are provided in Table 6.1.

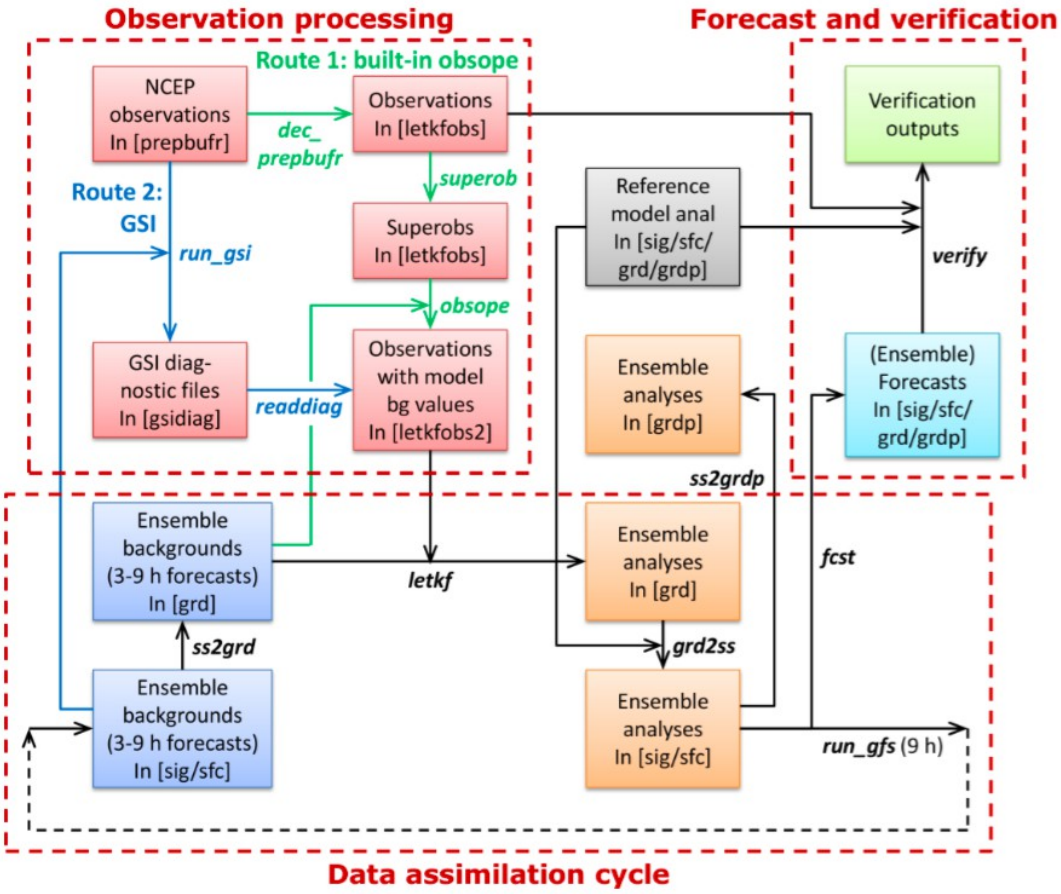


Figure 6.1: The flow chart of GFS-LETKF system (Lien, 2014).

Table 6.1: A list of experiments performed with the GFS-LETKF system with the PQC configurations

Exp. Name	Cycling PQC	EFSO lead-time	Threshold $[Jkg^{-1}]$	Rejecting Percentage	Note
exp00	No	6 and 24 hours	N/A	N/A	Control
exp01	No	6 hours	1.67×10^{-5}	10%	Non-cycling PQC
exp02	Yes	6 hours	1.67×10^{-5}	10%	Only accumulated PQC
exp03	Yes	6 hours	1.67×10^{-5}	10%	Full cycling PQC
exp04	Yes	6 hours	4.94×10^{-6}	20%	Full cycling PQC
exp05	Yes	6 hours	1.41×10^{-6}	30%	Full cycling PQC
exp06	Yes	6 hours	6.76×10^{-8}	40%	Full cycling PQC
exp07	Yes	6 hours	0	50%	Full cycling PQC
exp08	Yes	24 hours	3.65×10^{-5}	10%	Full cycling PQC

6.3 Results

6.3.1 PQC improvement in the analysis

We first examine the monthly-averaged cycling PQC corrections on the analysis from exp05 as an example with a map view. In Figure 6.2, we show the u-component wind, temperature, and humidity analysis corrections at 500 and 850 hPa. It is clear that the analysis error for u-component wind and temperature are reduced all over the globe for the two pressure level while the error reduction for specific humidity mainly distributed in lower latitudes over the ocean. A noticeable feature is that the u-component wind and temperature corrections are largest over the Southern Ocean.

6.3.2 PQC improvement in the forecast

We continue to examine the PQC corrections on the forecasts. Figure 6.3 shows the same as in Figure 6.2, but for improvement in forecast valid at 24-hour lead-time. We can see that the general distribution in the forecast corrections is the same as in the analysis corrections with a slight decrease in magnitude, which indicates the improvement starts decaying gradually after 24 hours, but as we will show later, the improvement persists even after 5 days.

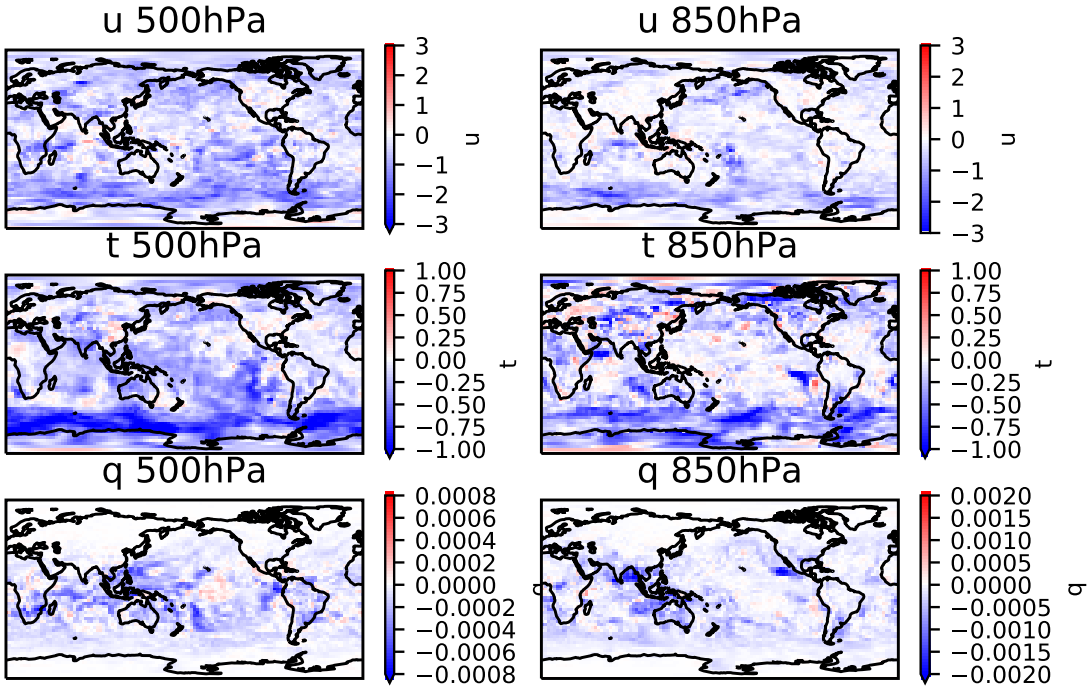


Figure 6.2: Monthly mean analysis error (RMSE) reduction for u-component wind [u; ms^{-1}], temperature [t; K], and specific humidity [q; $kgkg^{-1}$] at 500 and 850 hPa from exp05 in Table 6.1.

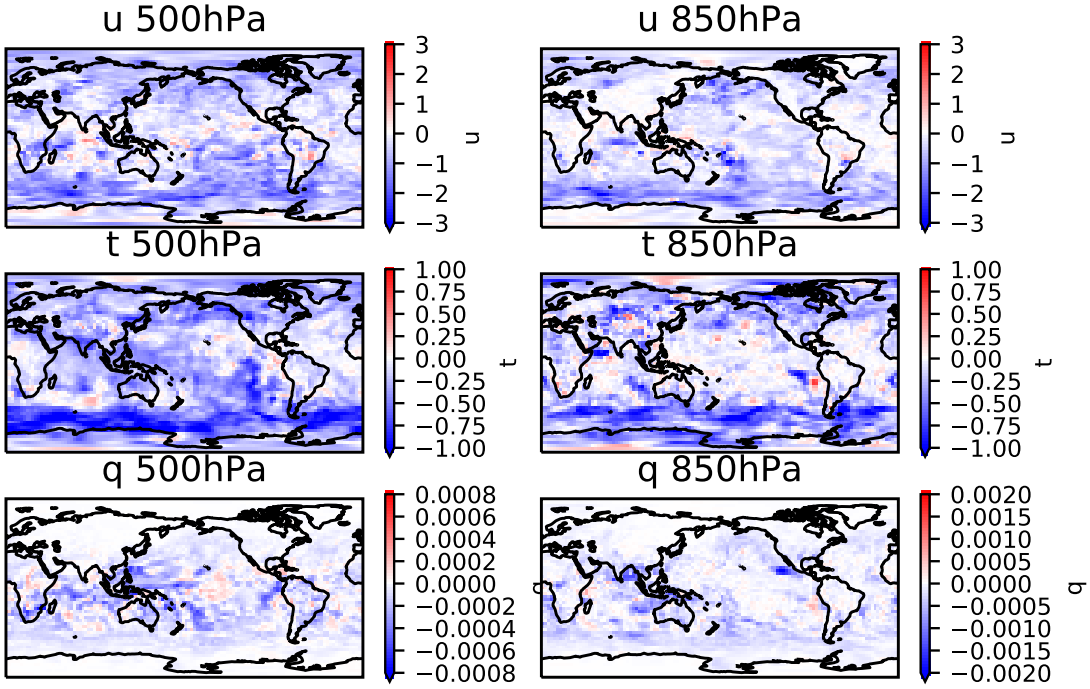


Figure 6.3: Same as 6.2, but for 24-hr forecast error reduction.

6.3.2.1 Sensitivity to the number of rejecting observations

In Hotta et al. (2017a), a significant focus was on designing the data-denial strategy, due to the speculation that rejecting all detrimental observations could lead to forecast degradation. However, as shown in Chapter 4 with Lorenz (1996) system, rejecting 10% of the most detrimental observations contributes to the majority of the improvement and the forecast even improves further when rejecting more observations with smaller detrimental impact, but the additional improvement becomes smaller as we reject more observations up to around 50% of the total number of the observations. Here we like to check if this observation is still valid for the high-dimensional and realistic GFS model.

In Figure 6.4, we show the relative forecast error reduction valid from $0-$

120 hours and compare the improvements from rejecting 10% to 40% of the most detrimental observations. It shows consistent results with Figure 6.2 and 6.3 that the relative improvement is the largest within 24 hours ranging from 10% to 20% (for all regions and all listed variables) and decays afterwards, but even after 5 days the improvement tends to saturate at about $5 - 10\%$ (not 0%!). In addition, we can observe that the majority of improvement comes from rejecting the top 10% most detrimental observations, and the improvement saturates around rejecting 40 % of the most detrimental observations which is consistent with the previous results that the forecast is improved rather than degraded by rejecting almost all of the detrimental observations even in GFS model.

In Figure 6.5 and 6.6, the correction for u-component wind, temperature and specific humidity at each pressure level is shown for the analysis and the 24-hr forecast. It is clear that u-component wind and temperature corrections make larger contributions to mid-levels in higher latitudes but upper-levels in the tropics. The correction for specific humidity is largest at lower level around 700 hPa in all regions.

6.3.2.2 Immediate and Accumulated impact of PQC

The cycling PQC improvement is further broken down into the immediate correction (non-cycling PQC) and the accumulated correction (improved background from cycling PQC). The immediate impact comes from the PQC update from the original analysis at current cycle, which is the same as the non-cycling PQC. The main benefit of cycling PQC is the accumulation of the improvements throughout the

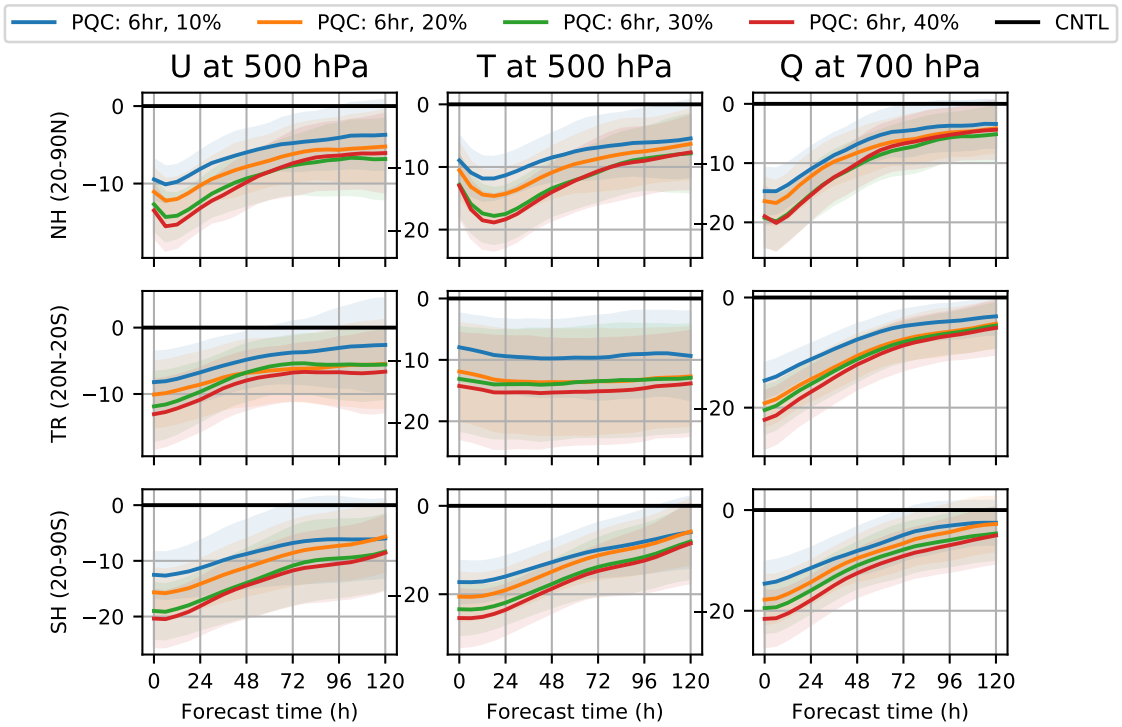


Figure 6.4: Monthly mean relative forecast error (RMSE) reduction percentage in u-component wind at 500 hPa, temperature at 500 hPa, and specific humidity at 700 hPa for the Northern Hemisphere (20N-90N), the tropics(20N-20S), and the Southern Hemisphere(20S-90S) throughout 5 days. The curves represent the improvement by rejecting overall 10%, 20%, 30%, and 40% of observations with cycling PQC with 6-hour lead-time. The shading shows the standard deviation of the forecast improvements.

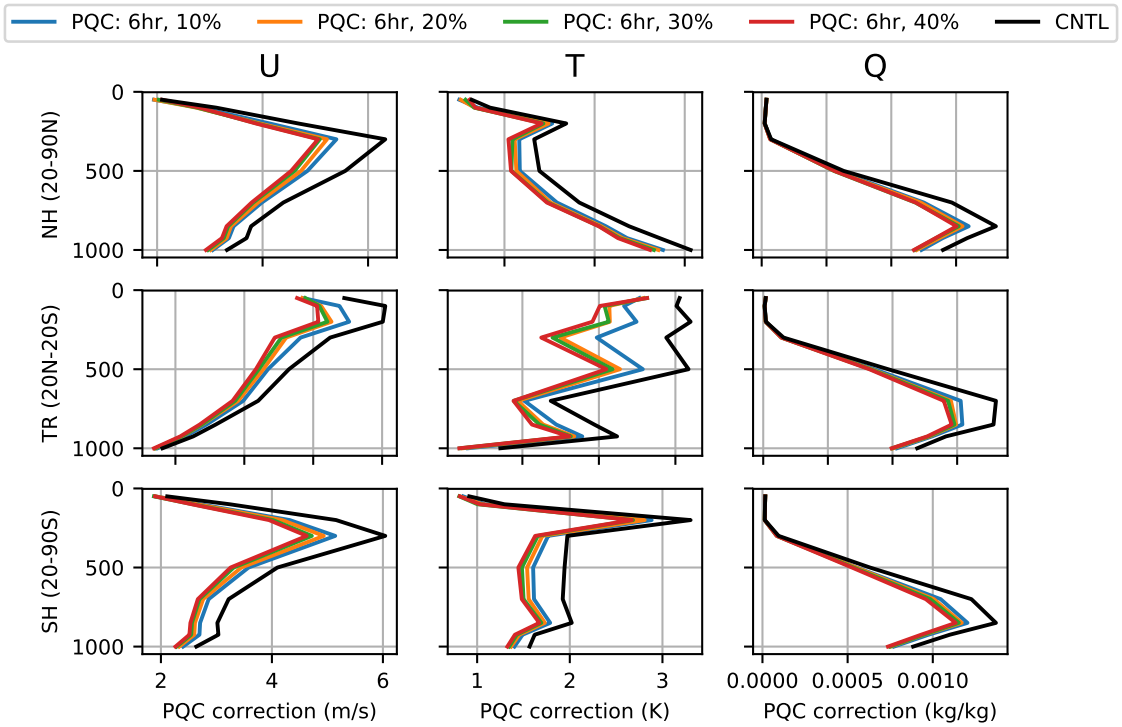


Figure 6.5: Monthly mean analysis error (RMSE) reduction by rejecting overall 10%, 20%, 30%, and 40% of observations with cycling PQC in u-component wind, temperature, and specific humidity at each pressure level for the Northern Hemisphere (20N-90N), the tropics(20N-20S), and the Southern Hemisphere(20S-90S).

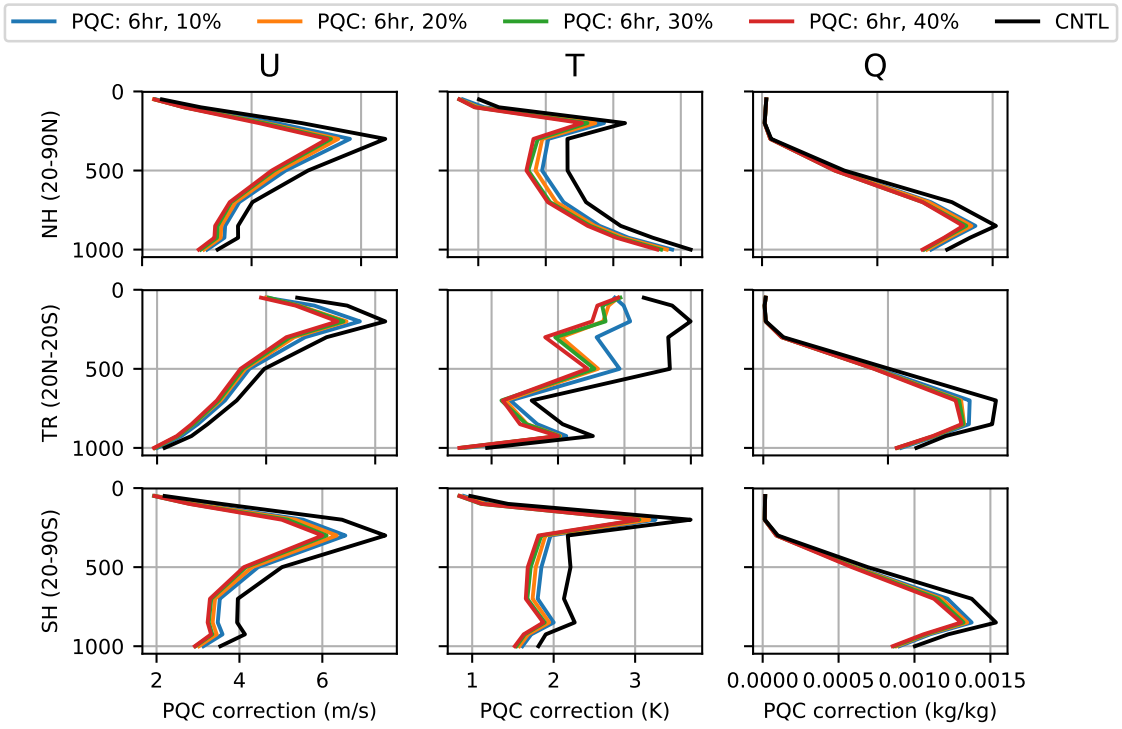


Figure 6.6: Same as Figure 6.5, but for 24 hours forecasts.

past cycles that the improved forecast initiated from the previously PQC-corrected analysis serves as a much more accurate background and further boosts the accuracy of the subsequent analysis. We separate the accumulated correction from the full impact by verifying the forecasts initiated not from the PQC corrected analysis but from the original analysis before PQC (which is still improved by previous PQC corrections). As we can see, the primary advantage of cycling PQC comes from the accumulation of past improvements whereas the independent immediate improvement is at most only 2%. It is also noticeable that the benefit from the accumulated impact has a more significant contribution to the full impact in the tropics and the Southern Hemisphere comparing to that in the Northern Hemisphere, indicating that the PQC improvement in the Northern Hemisphere has a shorter memory on

average.

The fact that the accumulation of past impacts contributes to a major portion of the full impact of cycling PQC has two important implications. One is that the PQC improvement has a long-term impact, and remains in the system even after several cycles of DA. Secondly, this supports the feasibility of implementing PQC in operational NWP. The operational centers need to initiate the forecast as soon as the analysis is completed to deliver the forecast products on time, so we can only afford to perform PQC after the current forecast is out, meaning the direct impact from PQC is not available in operations. Therefore, the huge portion corresponding to accumulated indirect impact gives a very encouraging message that even without the direct impact of the current observations we can still get a forecast improvement close to the full impact of PQC.

6.3.2.3 Verifying lead-time

After seeing such promising results, we would like to test the sensitivity of the PQC improvement to the EFSO verifying lead-time. Although we could not afford more than 6 hours in operation, it could still be useful in retrospective analysis systems. The FSO impacts are usually computed with lead-time of 24 hours to avoid potential influence from the diurnal cycle. If such influence exists, then PQC could suffer more than any other EFSO applications since it relies on immediate EFSO impact in each cycle. In other words, we would like to see if PQC using longer lead-time provide even further improvements as in Lorenz (1996) system.

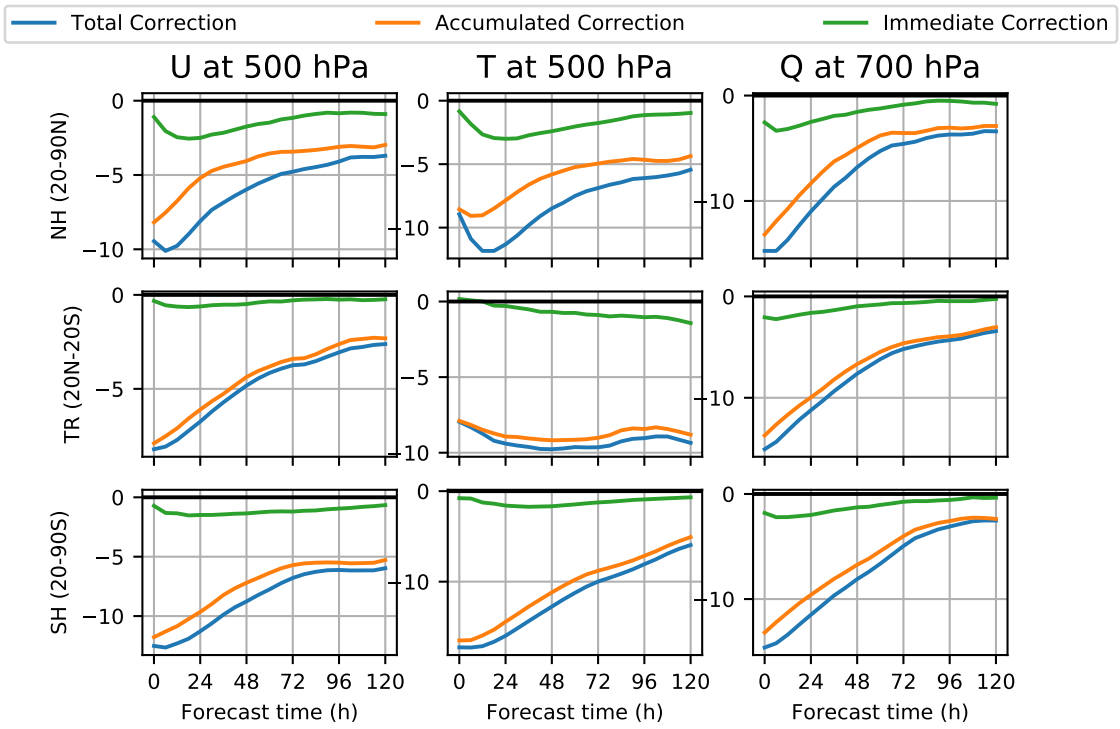


Figure 6.7: Monthly mean relative forecast error (RMSE) reduction percentage initiated from cycling PQC analysis (full correction), original analysis in cycling PQC experiment (accumulated correction), and non-cycling PQC (immediate correction) in u-component wind at 500 hPa, temperature at 500 hPa, and specific humidity at 700 hPa for the Northern Hemisphere (20N-90N), the tropics(20N-20S), and the Southern Hemisphere(20S-90S) throughout 5 days.

Intuitively, PQC with longer lead-time may better capture the long-term evolution of the flow and lead to longer and better improvement. Longer lead-time also poses technical challenges, including the evolution of localization and the nonlinearity of flow evolution. In Figure 6.8, we compare the PQC relative improvements to the forecast using 6 and 24 hours for verifying lead-time. As expected, it is clear that 24-hr PQC reaches its maximum improvement around 24 hours later compared to 6-hr PQC across all variables and regions. Surprisingly, the maximum improvement of 24-hr PQC is only as good as that of 6-hr PQC at best. In the tropics, the 24-hr PQC performs even worse than the 6-hr PQC. We observe the same for specific humidity for all regions, which can be related to the intrinsic time scale of the dominant process and the nonlinear regime for error growth of the region and the variable. The tropics and the humidity field are generally considered to be short-memory in weather time scale. Also, the analysis of 24-hr PQC is always less accurate compared to that of 6-hr PQC as we seen in Lorenz (1996) system that the analysis is not necessarily improved as much with the forecast improvement. Moreover, there are no clear advantages in the forecast improvement of using 24 hours as verifying lead-time over 6-hr PQC. A probable explanation is that the accuracy of the localization advection method to address the actual nonlinear evolution of the initial LETKF localization function with the background flow. The inaccuracy of the advection method increases with the verifying lead-time. This source of error for 24-hr PQC may compensate the benefit of EFSO impact evaluation for longer lead-time.

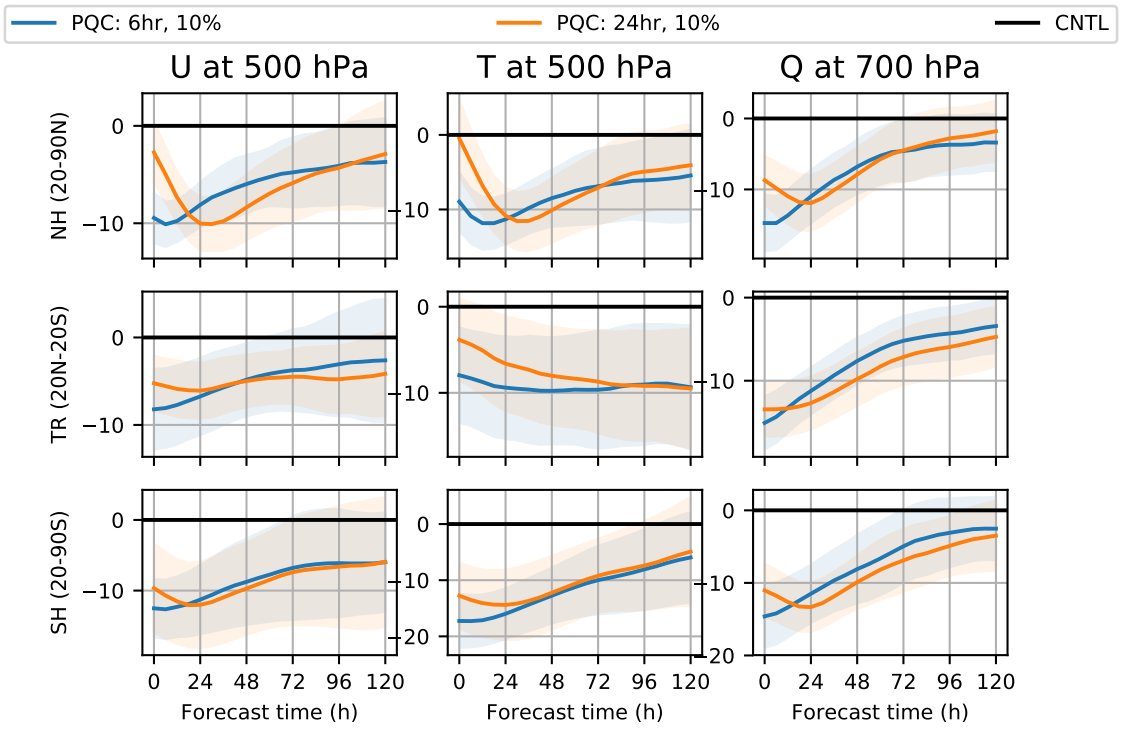


Figure 6.8: Same as Figure 6.4, but for 6-hr and 24-hr PQC rejecting 10% of the observations.

6.4 Towards operational implementation

To accommodate PQC into NCEP operational framework, several shortcuts were proposed in Chapter 3. In this section, we will validate all those shortcuts: (1) Use GFS (instead of GDAS) analysis as the verifying truth, (2) Reuse the original ensemble forecasts, and (3) Reuse the original Kalman gain (PQC-K update method).

6.4.1 Using GFS analysis as verifying truth

One of the significant time-delaying sources of PQC is having to wait for next available verifying analysis. In standard PQC, the verifying analysis is the GDAS final analysis from the following DA cycle. It is, however, proposed (Hotta et al., 2017a) that we can shorten the wait time by 2.5 hours by taking advantage of the GFS early analysis. In figure 3.2, the NCEP dual track framework and how PQC take advantage of GFS early analysis is illustrated. The main reason for the dual track design is for mitigating the forecast delay from observation transition time. However, there are already 70-80% of the data ingested into GDAS final analysis available to GFS early analysis in our experimental period (in 2012). This ratio has been increasing due to upgraded data transmission efficiency. Hence, the difference between GFS and GDAS analysis is even smaller nowadays.

To justify the replacement of GDAS with GFS analysis for EFSO verifying truth, here we compare the EFSO impact verified with GDAS final analysis and with GFS early analysis. In figure 6.9, the scatter plot comparing the EFSO impact

with GDAS analysis and that with GFS analysis of the same set of observations. The correlation between the two is high, 0.95 and all the observations fall closely to the diagonal line. The high correlation demonstrates the statistical similarity between the two choices of the verifying truth and clearly shows that it is legitimate to replace GDAS analysis with GFS analysis with only slight differences in the impact evaluation.

6.4.2 Reuse the original ensemble forecasts

In NCEP operational Hybrid EnVar system, the ensemble component provides only the error covariances for the construction of cost function in the variational component. We argue that the primary benefit of PQC is the corrections on the mean trajectory rather than the error covariances, so it is not necessary to repeat the computationally expensive ensemble forecasts.

Due to the difficulties in visualizing the error covariances of a high dimensional model, we show here in Figure 6.10 the comparison of the original analysis error covariance and the covariance difference for PQC in Lorenz (1996) system. All detrimental observations were rejected with 6-step PQC-K update method to obtain the maximum possible corrections. It is clear that the corrections on $\boldsymbol{A}$ is negligible due to its small magnitude. To show that this is not just a particular case, we demonstrate the time evolution of the maximum value of the PQC correction on $\boldsymbol{A}$ in Figure 6.11. Throughout the entire 5000-step experimental period, the most significant difference comes to around $2 * 10^{-17}$. We confirmed the hypothesis that

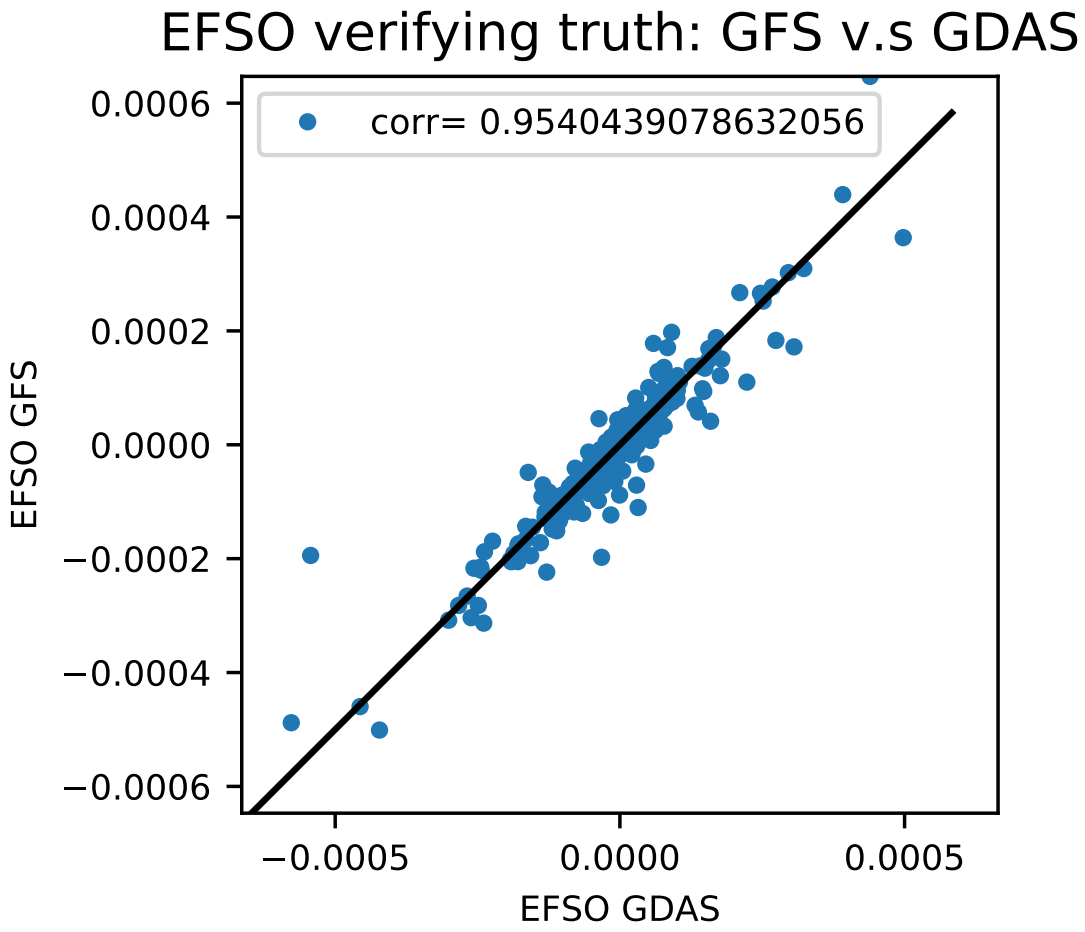


Figure 6.9: Scatter plot for samples of EFSO impact of typical DA cycle (18Z Feb 06, 2012) verified by GDAS final analysis and GFS early analysis. The correlation between the two choices is 0.95.

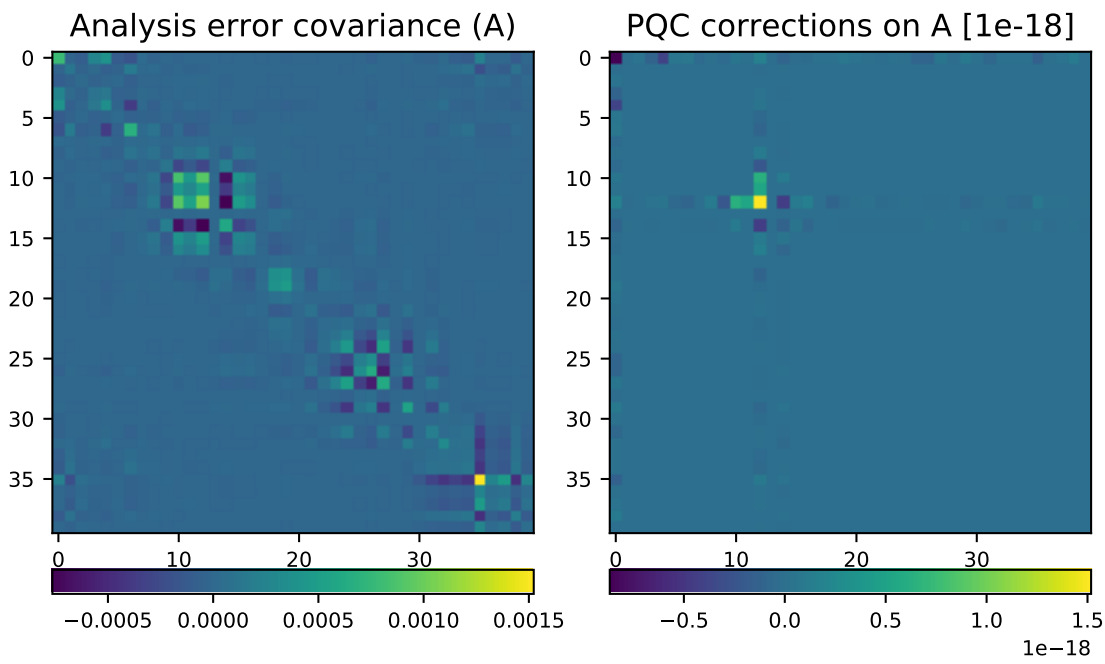


Figure 6.10: Analysis error covariance ($\boldsymbol{A}$; left) and the PQC correction on $\boldsymbol{A}$ (right) of the cycle #2002.

PQC corrects mainly the mean trajectory and leaves the error covariance almost the same as before.

To summarize, we found evidence showing the PQC corrections on the covariances is negligible in our experiments, suggesting that we may not need to re-run the ensemble forecasts in EnVar system, but a careful investigation should be carried out using a close-to-operation configuration to provide a definite conclusion.

6.4.3 Reuse the original Kalman gain (PQC-K update method)

As we have discussed earlier in Chapter 3 and 4, PQC-K computes the PQC correction with the original gain matrix estimated by original analysis perturbation. PQC-K not only avoids repeating the analysis again but also provides better im-

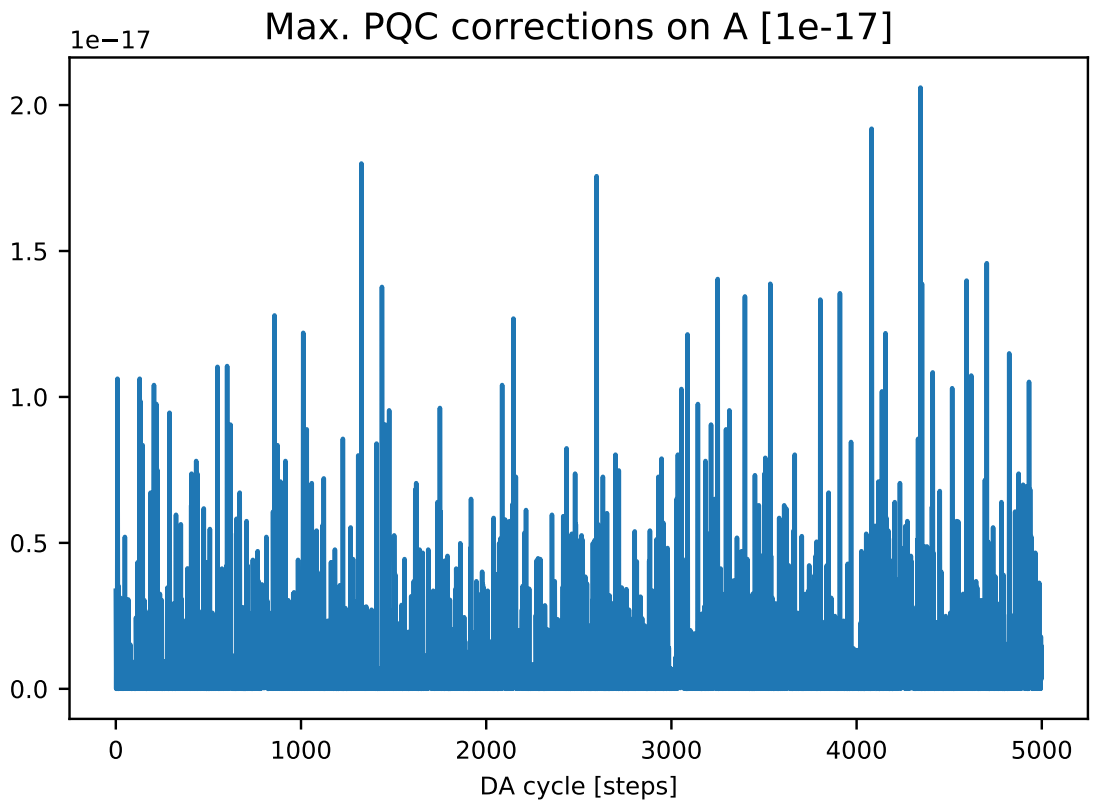


Figure 6.11: Time evolution of the maximum magnitude of PQC corrections on A.

provement compared to standard data-denial methods (PQC-H or PQC-R) since it is more consistent with the EFSO configuration and preserves the original ensemble spread that avoids unintentional inflation of the error covariance.

Here we compare the results of the standard rejection (PQC-H) and that of reusing the original Kalman gain (PQC-K) with GFS model in one of the data denial case (18Z Feb 06, 2012) in Hotta et al. (2017a). In figure 6.12-6.14, we compare the differences between the regional PQC-H and PQC-K corrections in analysis with map view respectively for u-component wind, temperature, specific humidity, and geopotential height. The first thing to notice is that the PQC-K corrections are generally larger than those from PQC-H, but the overall patterns of the corrections are very similar. From the map view, we can see this is true especially for the u-component wind, temperature, and geopotential height at both 500 and 850 hPa pressure levels. For specific humidity, the general pattern of PQC-H and PQC-K corrections are not as similar as in other variables.

The differences between the two PQC update choices in magnitude and the similarities in the pattern can be summarized in scatter plots in Figure 6.15-6.16. Consistent with what we observed before, all variables except specific humidity and relative humidity show high correlations (with coefficient around 0.9) between the two PQC corrections across pressure levels from 700 to 300 hPa. Also, we found the magnitudes of PQC-K corrections is generally 1-2 times larger than that of PQC-H by comparing with the two reference line with slope=1 and 2.

The difference in magnitude can be explained by the relative weightings to the observations in the two update methods. Consistent with EFSO, PQC-K computes

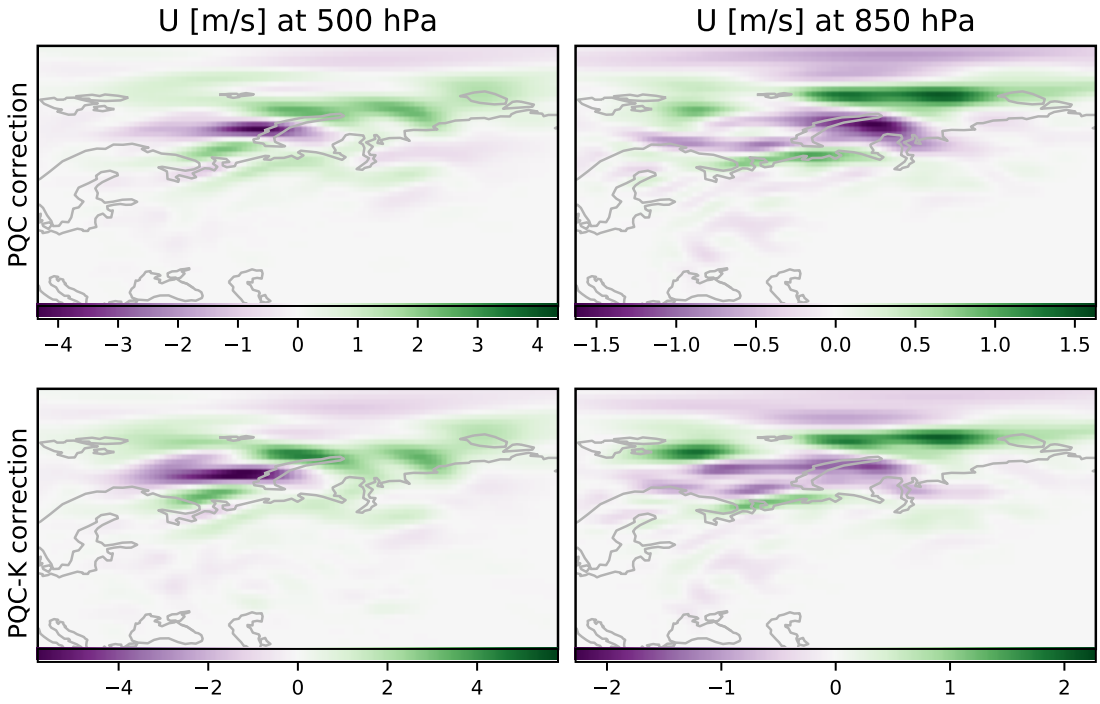


Figure 6.12: Comparisons of PQC-H (upper panels) and PQC-K (lower panels) corrections for U-component wind field at 500 (left column) and 850 (right column) hPa for the 18Z Feb 06, 2012 case in Hotta et al. (2017a).

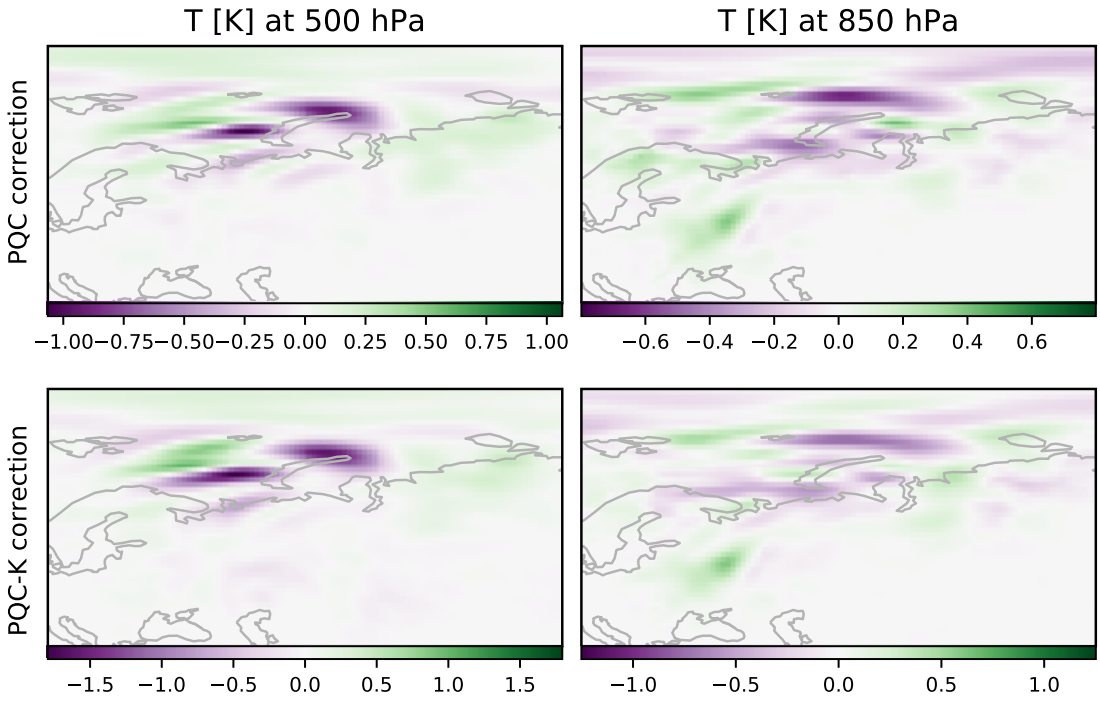


Figure 6.13: Same as Figure 6.12, but for temperature field $[K]$.

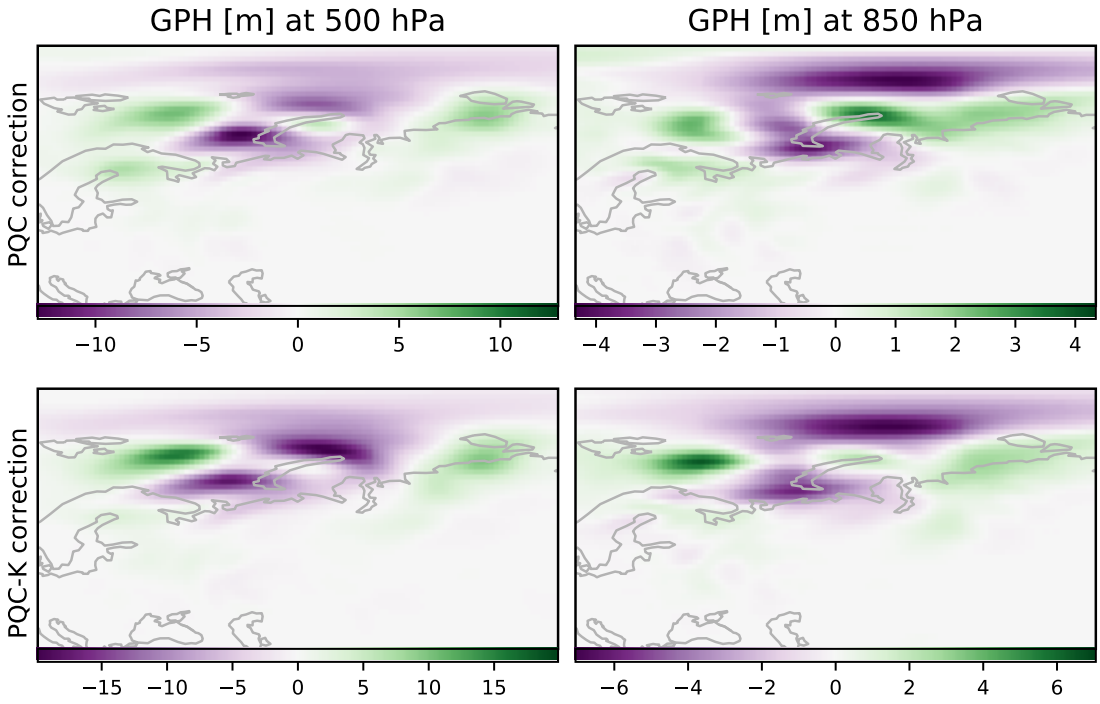


Figure 6.14: Same as Figure 6.12, but for geopotential height field $[m]$.

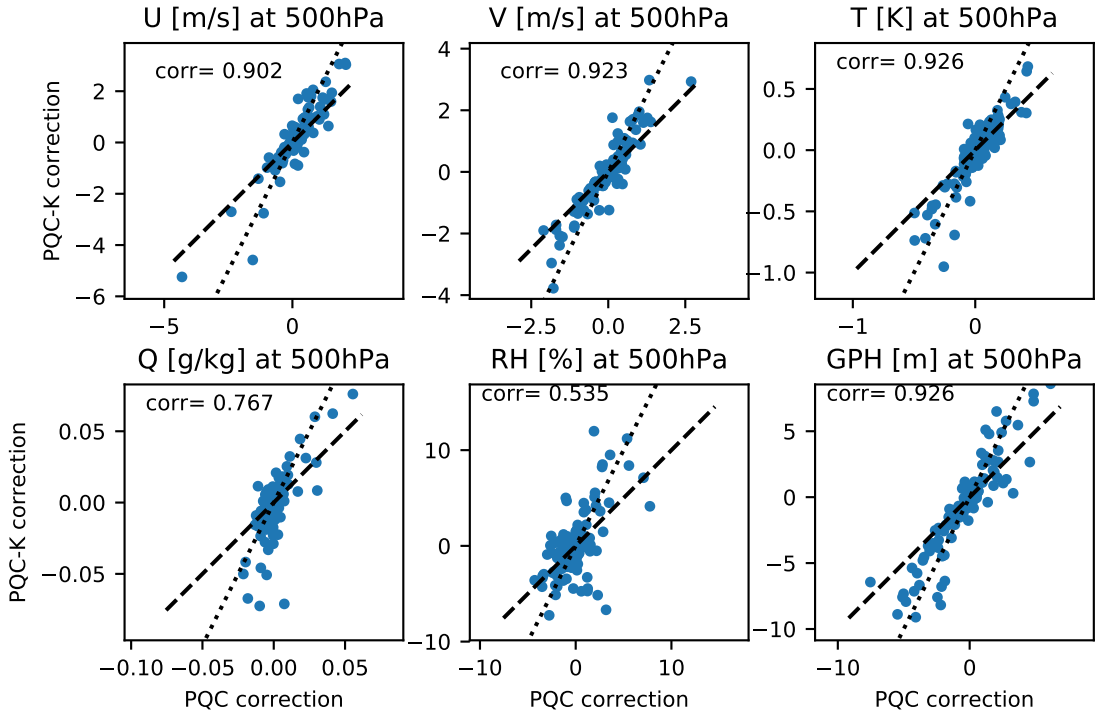


Figure 6.15: Scatter plots and correlation coefficients of PQC-H and PQC-K corrections for U-component wind, V-component wind, Temperature, Specific Humidity, Relative Humidity, and Geopotential height at 500 hPa. For reference, a dashed line with slope=1 and a dotted line with slope=2 is added.

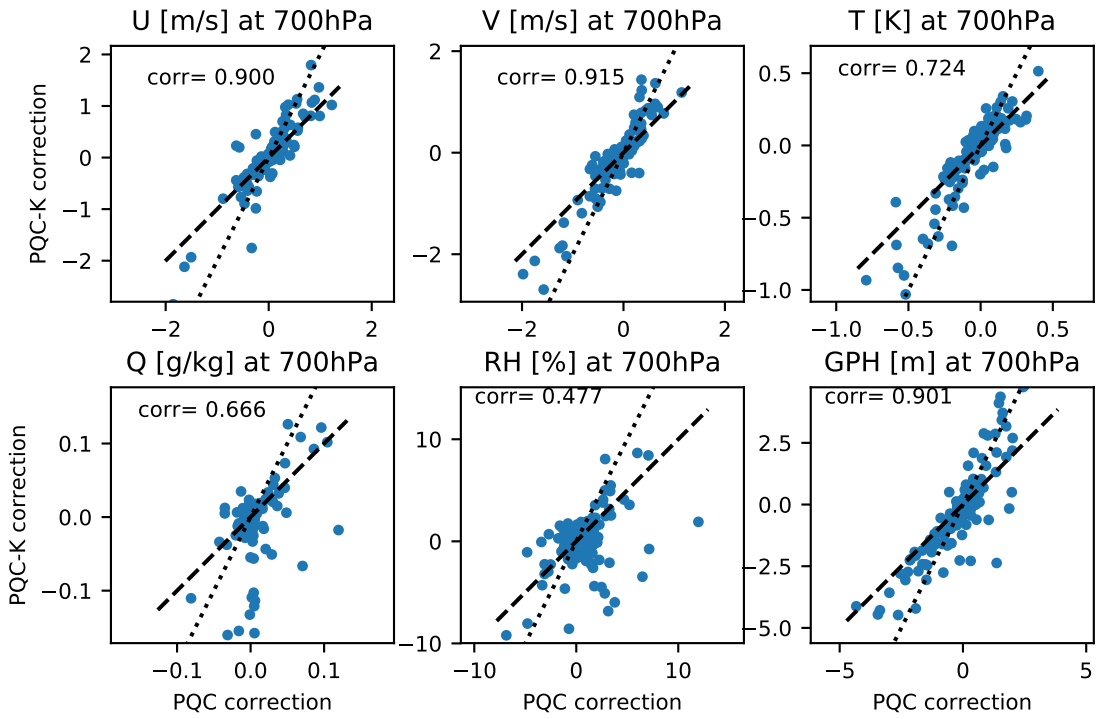


Figure 6.16: Same as figure 6.15, but for 700 hPa.

the correction as if all observations are assimilated. By contrast, the PQC-H correction is obtained by the actual rejection of the data, which naturally increases the weighting of other observations in the vicinity and result in decreasing the apparent weighting of the rejected observations. As we discussed in Chapter 4, the PQC-H method slightly changes the analysis increments associated with EFSO identified detrimental observations while PQC-K method rejects precisely the desired analysis increment. Here we merely offer a verification that the PQC-K update method also works in realistic GFS model and the correction is only slightly different from PQC-H correction. This slight difference could lead to considerable performance changes after several accumulations as seen in the Lorenz (1996) system where the performance of PQC-K was superior.

6.5 Summary and discussion

In this chapter, we have explored the cycling PQC using low-resolution GFS model coupling to a simple LETKF DA system. We showed that rejecting most of the detrimental observations gives the largest forecast improvement. PQC improves the forecast all over the globe for most of the variables. Furthermore, we demonstrated that the PQC based on EFSO with verifying lead-time longer than 6 hours does not gain any clear advantages. More importantly, the accumulation of past PQC corrections was shown to be the major contributor to the cycling PQC correction, indicating that we do not need PQC in real-time before the long GFS forecast initiation and still get most of the benefit of cycling PQC. Even if we perform PQC just for GDAS final analysis, the accumulation of benefits will still improve the subsequent forecasts.

The shortcuts proposed in Chapter 3 for operational implementation to NCEP were also validated. We showed that EFSO verified with GFS early analysis is statistically similar to that verified with GDAS final analysis. Thus, it is acceptable to replace GDAS with GFS analysis and gain extra 2.5 hours to perform EFSO and PQC. Also, we show evidence supporting that PQC corrections on the error covariance are negligible, suggesting that it may not be necessary to repeat the ensemble forecasts, but a careful investigation should be carried out to provide a definite conclusion. We further demonstrated that repeating of analysis step in PQC-H method can be avoided with the PQC-K update method by reusing the original gain matrix for computing the corrections on the analysis.

These promising results suggest that PQC is not only beneficial but is also affordable for the operational system. We would like to point out that there still exist possible challenges when applying cycling PQC to real NCEP operational system. The first is the observation density. PQC relies on the luxury of having a large number of observations in the atmosphere. With the low-resolution GFS model, the observation density is still high even after data thinning. It is possible that the observation density is lower in the operational resolution. Besides, a model with higher resolution resolves more small-scale features with higher nonlinearity that may reduce the PQC improvement. Further, the high quality of analysis in operational system assimilates much more observations with careful quality control and leaves smaller room for improvement. Lastly, we would like to point out that it can be more desirable to compute the FSO impact with the approach consistent with the DA system. Hence, the HFSO using En-Var approach may be more suitable for PQC in NCEP operational system, mainly because that only a fraction of the assimilated observations in high-resolution variational component goes into the low-resolution ensemble component. This problem may be overcome by estimating the EFSO impact of the observations as if they are assimilated into the ensemble component (David Groff 2017, personal communication).

Chapter 7

Summary and Future Directions

7.1 Summary

In this section, a summary of the dissertation is provided. We proposed two applications of FSO impact, namely the data monitoring and selection based on FSO and proactive quality control (PQC).

FSO allows online monitoring of any subset of the observations. We demonstrated that grouping the EFSO impacts based on observation types indeed reveals occasional detrimental-episodes from the overall beneficial impact of all observations. Implementing this tool informs the forecasters ahead of time about forth-coming degraded forecast by specific detrimental observation subset. Besides, the accumulated impact data can form the basis of frequent update of the blacklist QC. In the data selection application, we first show as a proof of concept that EFSO is capable of identifying observations with artificially added biases and random errors from a group of ordinary observations in Lorenz (1996) system. It is then shown in a GFS LETKF(and EnSRF)/3DVAR Hybrid GSI system with a close-to-operation configuration that FSO identifies detrimental radiance channels that are robust across different lead-times from specific satellite instruments, including the hyperspectral AIRS, CrIS, and IASI, as well as the multi-channel GOES sounders and HIRS. Additionally, we found a particular detrimental spectral-band that shows up in almost

every instrument covering the same spectrum. This detrimental band in common is a supportive evidence that some of the identified channels are genuinely suffering from the same cause. A channel-denial experiment demonstrates the forecast can be improved merely by rejecting just 16 out of hundreds of the assimilated channels, indicating that FSO is a reliable tool permitting data selection based on the actual observational impact on the forecasts instead of indirect metric such as DFS.

PQC, a fully flow-dependent QC scheme utilizing immediate EFSO impact in each cycle, was first tested in cycling fashion in this study. For the Lorenz (1996)-ETKF system, we demonstrate that the optimal data-denial strategy is, in fact, rejecting most of the detrimental observations with PQC-K update method that reuses the original Kalman gain. A surprising result we found is that even in a perfectly idealized system (no bad observations), PQC still improves the system. It is also found that the improvement is most significant when rejecting the most detrimental 10% observations and becomes insensitive to rejecting more observations. This change in sensitivity is entirely consistent with EFSO impact evaluation that only small amount of observations associated with growing mode have substantial impacts while the rest provide impacts orders of magnitude smaller. Also, we found that PQC robustly improves the performance of the system even with suboptimal configuration. In some border cases, PQC even helps stabilize the filter from divergence. A significant finding is that PQC improvement is insensitive to the error in model forcing, indicating that PQC is not correcting the model errors which is a common question encountered.

We further examine PQC in a high-dimensional GFS-LETKF system using

real PREPBUFR observations. It is confirmed that even in a realistic and complex model as such, PQC still improves the analysis by 10-20% relatively and the forecast up to 5 days by 5%. Surprisingly, the same dependence on the number of rejecting observations in the Lorenz (1996) system is also found in this GFS system. The primary improvement comes from denying the most detrimental 10% of observations. More improvement is provided by rejecting more detrimental observations, but the additional improvement saturates around rejecting 40% of the data. It is also shown that PQC using longer EFSO verifying lead-time does not have clear advantages over that using just 6 hours, which is more feasible in operation. Additionally, the accumulated correction of cycling PQC is shown to be the major contributor of the total correction. The accumulation of corrections is a critical finding supportive to operational implementation since the immediate PQC correction of each cycle is not affordable in operation for releasing forecast products on time.

Finally, we verified several shortcuts proposed to lower the computational burden of PQC in NCEP operation setup. We show that EFSO verified with GFS analysis is statistically similar to that verified with GDAS analysis but saves 2.5 hours from waiting for the next analysis for verification. Additionally, we show that PQC does not correct the error covariance but mainly on the mean trajectory. Hence, it is not necessary to repeat the ensemble forecasts (only used to produce the error covariance for En-Var analysis) after PQC correction. Lastly, as also shown with the Lorenz (1996) system, PQC-K provides more accurate corrections with lower computational cost compared to the standard data-denial update method (PQC-H).

7.2 Future directions

First, we would like to increase the GFS-LETKF resolution and examine how sensitive is PQC to the resolution. As mentioned in Chapter 6, the observation density may be an important factor determining PQC improvement. As the model resolution increases, the relative observation density decreases. Also, more small-scale features are resolved with the increase of the resolution that could lead to an increase of nonlinearity of the system.

Second, it should be explored the impact of the vast amount of satellite radiance data on PQC improvement. A possible scenario may be that the increased accuracy of the analysis may reduce the room for improvement. However, the quality of the radiance data is generally considered not as good as the conventional observations in PREPBUFR data which could be leveraged by PQC.

Furthermore, we should extend the experimentation of cycling PQC with the En-Var system as in operation. A potential challenge may be that the observations assimilated in the variational component are not the same as in the ensemble component as discussed in Chapter 2. However, non-cycling PQC experiments by Ota et al. (2013); Hotta et al. (2017a) have shown encouraging results with the same GFS LETKF/3DVAR GSI Hybrid system (close to operational setup) we used in Chapter 5. It is noteworthy here that we are collaborating with Global Forecast Dropout Prediction Tool (GFDPT) team to alleviate the skill dropout problem in NCEP using PQC.

For the theoretical understanding of PQC, the smoother aspect of PQC men-

tioned in Chapter 3 should be explored. This understanding could provide a way to fit PQC into the fundamental framework of data assimilation.

For the longer-term goals, we would like to extend the data selection application further to the actual processing of channel selection from scratch. By the proposed iterative method in Chapter 5, we can scan through the channels with a small number at a time to come up with an optimal selection that is EFSO-based. In fact, we have begun a collaborative effort with Quantitative Observing System Assessment Program (QOSAP) team to complement OSSE with EFSO diagnostic to provide better impact evaluation for data selection.

Lastly, earth system modeling has become mature enough recently for coupling data assimilation, where the observations from one component could be assimilated into and benefit other components in the earth system. Currently one of the primary efforts is to determine the localization of the observation influence across the boundaries of different components. EFSO could play an essential role in the identification of the observations in one component (e.g., atmosphere) with a beneficial impact on the forecast of another (e.g., ocean).

Bibliography

Anderson, E., and H. Järvinen, 1999: Variational quality control. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 697–722, doi:10.1002/qj.49712555416, URL http://dx.doi.org/10.1002/qj.49712555416http://doi.wiley.com/10.1002/qj.49712555416.

Bauer, P., 2009: Observing System Experiments (OSE) to estimate the impact of observations in NWP. **(June)**, 15–17.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, doi:10.1038/nature14956, URL http://www.nature.com/doifinder/10.1038/nature14956.

Bi, L., J. A. Jung, M. C. Morgan, and J. F. Le Marshall, 2011: Assessment of Assimilating ASCAT Surface Wind Retrievals in the NCEP Global Data Assimilation System. *Monthly Weather Review*, **139 (11)**, 3405–3421, doi:10.1175/2011MWR3391.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2011MWR3391.1.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129 (3)**, 420–436, doi:10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282001%29129%3C0420%3AASWTET%3E2.0.CO%3B2.

Bormann, N., and J.-N. Thépaut, 2004: Impact of MODIS Polar Winds in ECMWF's 4DVAR Data Assimilation System. *Monthly Weather Review*, **132 (4)**, 929–940, doi:10.1175/1520-0493(2004)132⟨0929:IOMPWI⟩2.0.CO;2.

Buehner, M., P. Du, and J. Bédard, 2018: A New Approach for Estimating the Observation Impact in Ensemble–Variational Data Assimilation. *Monthly Weather Review*, **146 (2)**, 447–465, doi:10.1175/MWR-D-17-0252.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-17-0252.1.

Cardinali, C., 2009: Forecast sensitivity to observation (FSO) as a diagnostic tool. *ECMWF report*, **(October)**.

Cardinali, C., 2018: Forecast Sensitivity Observation Impact with an observation-only based objective function. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3305, URL http://doi.wiley.com/10.1002/qj.3305.

Cardinali, C., S. Pezzulli, and E. Andersson, 2004: Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **130 (603 PART B)**, 2767–2786, doi:10.1256/qj.03.205, URL http://doi.wiley.com/10.1256/qj.03.205.

Charney, J. G., R. FjÖrtoft, and J. V. Neumann, 1950: Numerical Integration of the Barotropic Vorticity Equation. *Tellus*, **2 (4)**, 237–254, doi:10.3402/tellusa.v2i4.8607, URL https://www.tandfonline.com/doi/full/10.3402/tellusa.v2i4.8607.

Daescu, D. N., and R. Todling, 2010: Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quarterly Journal of the Royal Meteorological Society*, **136 (653)**, 2000–2012, doi:10.1002/qj.693.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597, URL http://centaur.reading.ac.uk/24937/.

Dumelow, R., 2002: Overview of Observing System Experiments. *Office*, 97–124.

Ehrendorfer, M., R. M. Errico, and K. D. Raeder, 1999: Singular-Vector Perturbation Growth in a Primitive Equation Model with Moist Physics. *Journal of the Atmospheric Sciences*, **56 (11)**, 1627–1648, doi:10.1175/1520-0469(1999)056⟨1627:SVPGIA⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281999%29056%3C1627%3ASVPGIA%3E2.0.CO%3B2.

European Centre for Medium-Range Weather Forecast, 2016: Ifs Documentation - Part I : Observations. *IFS Documentation CY43R3*, I, 1–82.

Gambacorta, A., and C. D. Barnet, 2013: Methodology and Information Content of the NOAA NESDIS Operational Channel Selection for the Cross-Track Infrared Sounder (CrIS). *IEEE Transactions on Geoscience and Remote Sensing*, **51 (6)**, 3207–3216, doi:10.1109/TGRS.2012.2220369, URL http://ieeexplore.ieee.org/document/6375806/.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757, doi:10.1256/smsqj.55416.

Gasperoni, N. A., and X. Wang, 2015: Adaptive Localization for the Ensemble-Based Observation Impact Estimate Using Regression Confidence Factors. *Monthly Weather Review*, **143 (6)**, 1981–2000, doi:10.1175/MWR-D-14-00272.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00272.1.

Geer, A. J., 2016: Significance of changes in medium-range forecast scores. *Tellus, Series A: Dynamic Meteorology and Oceanography*, **68 (1)**, 1–21, doi:10.3402/tellusa.v68.30229.

Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX Observation Impact Intercomparison Experiment. *Monthly Weather Review*, **138 (11)**, 4009–4025, doi:10.1175/2010MWR3393.1.

Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou, and M. Divakarla, 2003: AIRS near-real-time products and algorithms in support of operational numerical weather prediction. *IEEE Transactions on Geoscience and Remote Sensing*, **41 (2 PART 1)**, 379–388, doi:10.1109/TGRS.2002.808307.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Monthly Weather Review*, **139 (2)**, 511–522, doi:10.1175/2010MWR3328.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2010MWR3328.1.

Groff, D., K. Ide, R. Mahajan, and Y. Zhu, 2017: Assessment of Ensemble Forecast Sensitivity to Observation (EFSO) Quantities for Satellite Radiances Assimilated in the 4DEnVar GFS. *28th Conf WA&F / 24th Conf NWP, 4E, 25JAN2017 WA*.

Hotta, D., 2014: Proactive Quality Control Based on Ensemble Forecast Sensitivity to Observations. Ph.D. thesis, Ph. D. dissertation, University of Maryland, URL http://drum.lib.umd.edu/handle/1903/15794.

Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017a: Proactive QC: A Fully Flow-Dependent Quality Control Scheme Based on EFSO. *Monthly Weather Review*, **145 (8)**, 3331–3354, doi:10.1175/MWR-D-16-0290.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-16-0290.1.

Hotta, D., E. Kalnay, Y. Ota, and T. Miyoshi, 2017b: EFSR: Ensemble Forecast Sensitivity to Observation Error Covariance. *Monthly Weather Review*, 17–0122, doi:10.1175/MWR-D-17-0122.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-17-0122.1.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230 (1-2)**, 112–126, doi:10.1016/j.physd.2006.11.008, URL http://linkinghub.elsevier.com/retrieve/pii/S0167278906004647.

Ingleby, N. B., and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Q. ˜J. ˜R. Meteorol. Soc.*, **119**, 1195–1225.

Isaksen, L., M. Fisher, E. Andersson, and J. Barkmeijer, 2005: The structure and realism of sensitivity perturbations and their interpretation as 'Key Analysis Errors'. **131**, 3053–3078, doi:10.1256/qj.04.99.

Janisko, M., and C. Cardinali, 2016: On the impact of the diabatic component in the Forecast Sensitivity Observation Impact diagnostics. Tech. Rep. August. URL http://www.ecmwf.int/en/research/publications.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, **64**, 1–9, doi:10.3402/tellusa.v64i0.18462,

URL http://www.tellusa.net/index.php/tellusa/article/view/18462/xmlhttp://www.tellusa.net/index.php/tellusa/article/view/18462.

Kazumori, M., and Y. Nakamura, 2004: Modis polar winds assimilation experiments at jma. *Proc. Seventh Int. Winds Workshop, 1-17 June 2004, Helsinki*, 265–72, URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.520.56&rep=rep1&type=pdf.

Kleist, D., 2012: AN EVALUATION OF HYBRID VARIATIONAL-ENSEMBLE DATA ASSIMILATION FOR THE NCEP GFS. Ph.D. thesis, Ph. D. dissertation, University of Maryland, 149 pp., URL http://drum.lib.umd.edu/handle/1903/13135.

Kleist, D. T., and K. Ide, 2015a: An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part I: System Description and 3D-Hybrid Results. *Monthly Weather Review*, **143 (2)**, 433–451, doi:10.1175/MWR-D-13-00351.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-13-00351.1.

Kleist, D. T., and K. Ide, 2015b: An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. *Monthly Weather Review*, **143 (2)**, 452–470, doi:10.1175/MWR-D-13-00350.1, URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-13-00350.1http://journals.ametsoc.org/doi/10.1175/MWR-D-13-00350.1.

Klinker, E., F. Rabier, and R. Gelaro, 1998: Estimation of key analysis errors using the adjoint technique. *Quarterly Journal of the Royal Meteorological Society*, **124**, 1909–1933, doi:10.1002/qj.49712455007.

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. *Monthly Weather Review*, **145 (12)**, 4977–4995, doi:10.1175/MWR-D-17-0094.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-17-0094.1.

Kotsuki, S., K. Kurosawa, and T. Miyoshi, 2018: Ensemble Forecast Sensitivity to Observations Verified with Multiple References. *The 8th EnKF Data Assimilation Workshop*.

Kumar, V. K., A. Eichmann, J. Alpert, and S. Boukabara, 2017: Global Forecast Dropout Prediction Tool in Support of the NCEP Model Evaluation Group (MEG)—Collaboration Project Between JCSDA/NESDIS & NWS. *JCSDA Quarterly*, **(55)**, 5–13, doi:10.7289/V5V98648.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, **56 (3)**, 189–201, doi:10.1111/j.1600-0870.2004.00056.

x, URL http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0870.2004.00056.x/fullhttp://tellusa.net/index.php/tellusa/article/view/14413.

Lawson, W. G., and J. a. Hansen, 2005: Alignment Error Models and Ensemble-Based Data Assimilation. *Monthly Weather Review*, **133 (6)**, 1687–1709, doi:10.1175/MWR2945.1.

Le Marshall, J., J. Jung, T. Zapotocny, C. Redder, M. Dunn, J. Daniels, and L. P. Riishojgaard, 2008: Impact of MODIS atmospheric motion vectors on a global NWP system. *Australian Meteorological Magazine*, **57 (1)**, 45–51, URL http://www.scopus.com/inward/record.url?eid=2-s2.0-46449084592&partnerID=tZOtx3y1.

Li, H., J. Liu, and E. Kalnay, 2010: Correction of 'Estimating observation impact without adjoint model in an ensemble Kalman filter'. *Quarterly Journal of the Royal Meteorological Society*, **136 (651)**, 1652–1654, doi:10.1002/qj.658, URL http://doi.wiley.com/10.1002/qj.658.

Lien, G.-y., 2014: Assimilation of Satellite Precipitation Analysis with the GFS-LETKF System. Ph.D. thesis, Ph. D. Prospectus, University of Maryland.

Lien, G.-Y., D. Hotta, E. Kalnay, T. Miyoshi, and T.-C. Chen, 2018: Accelerating assimilation development for new observing systems using EFSO. *Nonlinear Processes in Geophysics*, **25 (1)**, 129–143, doi:10.5194/npg-25-129-2018, URL https://www.nonlin-processes-geophys.net/25/129/2018/.

Lien, G.-Y., T. Miyoshi, S. Nishizawa, R. Yoshida, H. Yashiro, S. A. Adachi, T. Yamaura, and H. Tomita, 2017: The Near-Real-Time SCALE-LETKF System: A Case of the September 2015 Kanto-Tohoku Heavy Rainfall. *Sola*, **13 (0)**, 1–6, doi:10.2151/sola.2017-001, URL https://www.jstage.jst.go.jp/article/sola/13/0/13_2017-001/_article.

Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **134 (634)**, 1327–1335, doi:10.1002/qj.280, URL http://doi.wiley.com/10.1002/qj.280.

Liu, J., and E. Kalnay, 2009: Analysis sensitivity calculation in Ensemble Kalman Filter. *Quarterly Journal of the ...*, 1–30, URL http://onlinelibrary.wiley.com/doi/10.1002/qj.511/full.

Liu, J., E. Kalnay, T. Miyoshi, and C. Cardinali, 2009: Analysis sensitivity calculation in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135 (644)**, 1842–1851, doi:10.1002/qj.511, URL http://onlinelibrary.wiley.com/doi/10.1002/qj.511/abstract%5Cnpapers3://publication/uuid/5C9D122F-73D7-4B0A-A91C-2A5BE6E79B98.

Lorenc, A. C., and R. T. Marriott, 2014: Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, **140 (678)**, 209–224, doi:10.1002/qj.2122.

Lorenz, E., 1996: Predicability - a Problem partly solved. *Seminar on Predictability, Reading, United Kingdom, ECMWF*, 1–18, URL http://web.mit.edu/lorenzcenter/about/LorenzPubs/Predicability_a_Problem_2006.pdf.

Lorenz, E. N., and K. a. Emanuel, 1998: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model. *Journal of the Atmospheric Sciences*, **55 (3)**, 399–414, doi:10.1175/1520-0469(1998)055⟨0399:OSFSWO⟩2.0.CO;2.

Magnusson, L., and E. Källén, 2013: Factors Influencing Skill Improvements in the ECMWF Forecasting System. *Monthly Weather Review*, **141 (9)**, 3142–3153, doi:10.1175/MWR-D-12-00318.1, URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00318.1.

Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. *Monthly Weather Review*, **139**, 1519–1535, doi:10.1175/2010MWR3570.1.

Miyoshi, T., and Coauthors, 2016: "Big Data Assimilation" Toward Post-Petascale Severe Weather Prediction: An Overview and Progress. *Proceedings of the IEEE*, **104 (11)**, 2155–2179, doi:10.1109/JPROC.2016.2602560, URL http://ieeexplore.ieee.org/document/7576655/.

Nerger, L., 2015: On Serial Observation Processing in Localized Ensemble Kalman Filters. *Monthly Weather Review*, **143 (5)**, 1554–1567, doi:10.1175/MWR-D-14-00182.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00182.1.

Ng, G.-H. C., D. Mclaughlin, D. Entekhabi, and A. Ahanin, 2011: The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*, **63 (5)**, 958–977, doi:10.1111/j.1600-0870.2011.00539.x, URL http://tellusa.net/index.php/tellusa/article/view/15877.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus A*, **65**, 20 038, doi:10.3402/tellusa.v65i0.20038, URL http://www.tellusa.net/index.php/tellusa/article/view/20038/xml.

Penny, S. G., 2014: The Hybrid Local Ensemble Transform Kalman Filter. *Monthly Weather Review*, **142 (6)**, 2139–2149, doi:10.1175/MWR-D-13-00131.1, URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-13-00131.1.

Rabier, F., N. Fourrié, D. Chafaï, and P. Prunet, 2002: Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quarterly Journal of the Royal Meteorological Society*, **128 (581)**, 1011–1027, doi:10.1256/0035900021643638.

Rao, V., A. Sandu, M. Ng, and E. D. Nino-Ruiz, 2017: Robust Data Assimilation Using L_1 and Huber Norms. *SIAM Journal on Scientific Computing*, **39 (3)**, B548–B570, doi:10.1137/15M1045910, URL http://epubs.siam.org/doi/10.1137/15M1045910.

Riishojgaard, L., and Y. Zhu, 2004: Impact experiments on GMAO data assimilation and forecast systems with MODIS winds during MOWSAP. *Proc. Seventh Int. Winds Workshop*, URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.541.6361&rep=rep1&type=pdf.

Riishojgaard, L. P., and C. Redder, 2008: Observing System Experiments for MODIS winds in the Joint Center for Satellite Data Assimilation. 2007.

Rodgers, C. D., 1996: Information content and optimization of high-spectral-resolution measurements. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II*, P. B. Hays, and J. Wang, Eds., Vol. 2830, 136, doi:10.1117/12.256110, URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.256110.

Roh, S., M. G. Genton, M. Jun, I. Szunyogh, and I. Hoteit, 2013: Observation Quality Control with a Robust Ensemble Kalman Filter. *Monthly Weather Review*, **141 (12)**, 4414–4428, doi:10.1175/MWR-D-13-00091.1, URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-13-00091.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91 (8)**, 1015–1058, doi:10.1175/2010BAMS3001.1, URL http://journals.ametsoc.org/doi/10.1175/2010BAMS3001.1.

Sarrazin, R., and Y. Zaitseva, 2004: MODIS polar winds assimilation impact study with CMC operational NWP system. *Proc. 7th International Winds Workshop*, URL https://scholar.google.com/scholar?q=MODIS+polar+winds+assimilation+impact+study+with+CMC+operational+NWP+system&btnG=&hl=en&as_sdt=0%2C21.

Sommer, M., and M. Weissmann, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus, Series A: Dynamic Meteorology and Oceanography*, **68 (1)**, doi:10.3402/tellusa.v68.27885.

Tavolato, C., and L. Isaksen, 2009: 122. Huber norm quality control in the IFS. doi:10.21957/4k12rmv6.

Tavolato, C., and L. Isaksen, 2015: On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **141 (690)**, 1514–1527, doi:10.1002/qj.2440, URL http://doi.wiley.com/10.1002/qj.2440.

Verlaan, M., and J. Sumihar, 2016: Observation impact analysis methods for storm surge forecasting systems. *Ocean Dynamics*, doi:10.1007/s10236-015-0912-0.

Wang, X., D. F. Parrish, D. T. Kleist, J. S. Whitaker, X. Wang, D. F. Parrish, D. T. Kleist, and J. S. Whitaker, 2013: GSI 3DVar-Based Ensemble–Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Monthly Weather Review*, **141 (11)**, 4098–4117, doi:10.1175/MWR-D-12-00141.1, URL http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-12-00141.1.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*, **140 (9)**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble Data Assimilation with the NCEP Global Forecast System. *Monthly Weather Review*, **136 (2)**, 463–482, doi:10.1175/2007MWR2018.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007MWR2018.1.

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. *Monthly Weather Review*, **132 (5)**, 1238–1253, doi:10.1175/1520-0493(2004)132⟨1238:IOIEAO⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282004%29132%3C1238%3AIOIEAO%3E2.0.CO%3B2.

Zhu, Y., and R. Gelaro, 2008: Observation Sensitivity Calculations Using the Adjoint of the Gridpoint Statistical Interpolation (GSI) Analysis System. *Monthly Weather Review*, **136**, 335–351, doi:10.1175/2007MWR2046.1.