ABSTRACT

Title of Dissertation: A New Quantitative Framework for Application of Ensemble Forecast Sensitivity to Observations in NWP

David Neil Groff
Doctor of Philosophy, 2021

Dissertation Directed by: Eugenia Kalnay
Department of Atmospheric and Oceanic Science

Current global operational Numerical Weather Prediction (NWP) systems (e.g. the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS)) generally assimilate on the order of 10 million observations every 6 hours. Furthermore, there is substantial diversity in the sampling characteristics and associated error characteristics of the observation types assimilated. In this context, it is not feasible to obtain sufficiently detailed information for determining which available observations or observation types should be assimilated or rejected in NWP systems using traditional Observing System Experiment (OSE) approaches. Forecast sensitivity to observation impact (FSOI) based estimation techniques (Langland and Baker 2004) enable efficient estimation of forecast impacts due to assimilation of individual observations, and as such, represent a solution to this problem.

The ensemble forecast sensitivity to observations (EFSO) (Kalnay et al. 2012) impact estimation technique uses ensembles of forecasts to perform linear mapping of innovations to forecast error changes. This mapping involves application of Kalman

gain matrices consistent with the complete sets of observations assimilated during data assimilation cycles. As with the other forecast-sensitivity based observation impact estimation techniques there are two prominent "contextual" limitations for application of EFSO in NWP systems: i) the observation impacts are estimated with respect to simultaneously assimilating all other observations that contributed to an analysis, ii) EFSO calculations are relative to a background that includes information from all previously assimilated observations. To mitigate these "contextual" limitations in application of forecast-sensitivity based observation impact information, a new quantitative framework we call "EFSO-components" is developed by decomposing EFSO employed forecast errors and innovations into random and systematic components. Lorenz '96 simple model experiments indicate that application of "EFSO-components" provides potentially significant advantages in detection of specific observation flaws, and in further advancing the utility of EFSO-based PQC (Ota et al. 2013, Hotta et al. 2017a, Chen and Kalnay 2019, Chen and Kalnay 2020). As such, we explore how "EFSO-components" fundamentally addresses the aforementioned contextual limitations of forecast-sensitivity based observation impact estimation in a manner that explains the potential application advantages according to Lorenz '96 simple model experiments.

Additionally, a new technique we call predicted EFSO (PEFSO), which is a straightforward extension to EFSO, is introduced in this study. PEFSO represents a potential capability for estimating the hypothetical forecast impacts of unassimilated observations. We explore the potential application of PEFSO as a convenient low computational cost approach for comparing the efficiencies of observing systems in reducing forecast error using Lorenz '96 simple model experiments.

A New Quantitative Framework for Application of Ensemble Forecast Sensitivity to Observations in NWP

by

David Neil Groff

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
2021

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Dr. Tse-Chun Chen, Co-Advisor
Professor Brian Hunt
Professor Jonathan Poterjoy
Professor James Carton
Professor Alfredo Ruiz-Barradas

Dedication

To my incredibly supportive family.

Acknowledgments

I am sincerely thankful for all the support that my advisors Professor Eugenia Kalnay and Dr. Tse-Chun Chen have provided for this dissertation work. I am thankful for Tse-Chun's generosity and thoughtful recommendations in contributing to this dissertation work. It has been a pleasure of mine to explore and contemplate scientific questions with Tse-Chun. I am also thankful for Eugenia's open mindedness to newly proposed explanations and for not giving up on me throughout this process.

I would like to thank my committee members, Professor Alfredo Ruiz-Barradas, Professor Brian Hunt, Professor James Carton and Professor Jonathan Poterjoy in providing helpful suggestions for this dissertation work. Furthermore, I am thankful for their generosity and willingness to serve on my committee.

I would also like to thank colleagues and friends from my time working at the WWB and NCWCP buildings for their support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMSU-A	Advanced Microwave Sounding Unit-A
CrIS	Cross-track Infrared Sounder
CFL	Courant-Friedrichs-Lewy
CWB	Central Weather Bureau
DA	Data Assimilation
EFSO	Ensemble Forecast Sensitivity to Observation
EFSR	Ensemble Forecast Senstivity to **R**
ENIAC	Electronic Numerical Integrator and Computer
EnKF	Ensemble Kalman Filter
EnSRF	Ensemble Square Root Filter
EnVar	Ensemble-Variational
ETKF	Ensemble Transform Kalman Filter
FSOI	Forecast Sensitivity to Observation Impact
GFS	Global Forecast System
IASI	Infrared Atmospheric Sounding Interferometer
JCSDA	Joint Center for Satellite Data Assimilation
NCEP	National Centers for Environmental Prediction
NWP	Numerical Weather Prediction
OSE	Observing System Experiment
OSSE	Observing System Simulation Experiment
PEFSO	Predicted Ensemble Forecast Sensivitivity to Observations
PQC	Proactive Quality Control
QC	Quality Control
RADAR	Radio Detection and Ranging
RMSE	Root Mean Square Error
ROPA	**R**andom **O**bservation **P**rojection onto Forecast Error **A**nomalies
ROPS	**R**andom **O**bservation **P**rojection onto **S**ystematic Forecast Error
SOPA	**S**ystematic **O**bservation **P**rojection onto Forecast Error **A**nomalies
SOPS	**S**ystematic **O**bservation **P**rojection onto **S**ystematic Forecast Error
EoRS	**E**FSO or (**R**OPA and **S**OPA) based rejection
EoR	**E**FSO or **R**OPA based rejection
EoS	**E**FSO or **S**OPA based rejection
RS	**R**OPA and **S**OPA based rejection

Chapter 1: Introduction

1.1 The Human Impact of Extreme Weather Events; an Impetus for Development of the "Weather Machine"

On September 7th, 1900 the chief forecaster at the United States Weather Bureau's Gulf Coast regional office in Galveston, Isaac Cline, noted unusually large swells and offshore winds. "Such high water with opposing winds never observed previously" (Larson 1999). However, limited to in-situ observations of barometric pressure and wind, he had no indication that a Hurricane would inundate the island of Galveston in the next 24 hours. The next day, approximately 8,000 people on Galveston Island died, and the city of Galveston's status as the future "New York on the Gulf" (Larson 1999) had expired. With the availability of modern numerical weather prediction (NWP) and observing system capabilities (e.g. satellite imagery), this tragedy would have been largely averted. Similarly, tragedies caused by other kinds of extreme weather events were realized due to an unawareness of impending extreme conditions. For example, the "Children's Blizzard", which claimed the lives of more than 200 people, was essentially not forecasted (Laskin 2004). Extreme weather events have no doubt been a major impetus for international efforts to develop and improve NWP, or as recently coined in

popular science literature, the "weather machine" (Blum 2018).

1.2 A Summary Account of NWP History

In the 19th century, observational and theoretical advances in atmospheric science provided the background for an organized scientific framework of the NWP problem. This unified scientific framework for NWP was realized by Finish meteorologist Wilhelm Bjerknes. Bjerknes was able to confirm the essence of his circulation theorem using a special 3-dimensional kite based data set provided by Cleveland Abbe [1] , and in so doing, understood there was now a paradigm for meteorology to become an established scientific field. In a short article titled "The Problem of Weather Forecasting as Seen From the Viewpoints of Mechanics and Physics" Bjerknes 1904; Bjerknes outlined the theoretical and observational aspects of the NWP problem.

> "This was the first explicitly, coherent recognition that the future state of the atmosphere is, in principle, completely determined by its detailed initial state and known boundary conditions, together with Newton's equation of motion, the Boyle-Charles-Dalton equation of state, the equation of mass continuity, and the thermodynamic energy equation" (Kalnay 2002).

Inspired by Bjerknes scientific framework, Lewis Fry Richardson attempted the first widely recognized numerical forecast. During WW1, Richardson retrospectively performed an integration of the atmosphere for 20 May, 1910. "Richardson worked to reconstruct the weather on a spring morning when balloons had drifted over the

[1] In 1871 was appointed as first chief scientist of U.S. Weather Bureau. Advocated for a Physics, as opposed to empirical, based atmospheric forecasting approach.

then-peaceful European countryside, treating the motions of the atmosphere as nature's solution to a system of differential equations that linked the conditions in adjacent cells from one time step to the next" (Dyson 2012). The simulation of the atmosphere performed by Richardson was unrealistic, but it highlighted that a successful numerical forecast necessitated the filtering of fast moving gravity and sound waves. In 1948, Jules Charney and Arnt Eliassen applied quasi-geostrophic approximations to the equations of motion that allowed for a balanced evolution of the forecast, and made it possible to avoid violation of the Courant-Friedrichs-Lewy (CFL) condition (Kalnay 2002). As this progress in understanding how one would theoretically perform a numerical forecast matured, the advent of the electronic computing age was unfolding. This technological advancement allowed the simulated atmosphere to proceed in time at nearly the pace of reality, as such, what Richardson had previously imagined no longer seemed fantastical, "perhaps someday in the dim future, it will be possible to advance the computations faster than weather advances, and at a cost less than the saving to mankind due to the information gained" (Dyson 2012). In 1950, using the Electronic Numerical Integrator and Computer (ENIAC) and applying a 1-layer model of the quasi-geostrophic equations of motion, Jules Charney, George Platzman, Ragnar Fjortoft, John Freeman, Joseph Smagorinsky, Klari von Neumann and John von Neumann ran a numerical forecast that to a large extent successfully captured the placement of undulations in observed pressure fields. Furthermore, in a subsequent experiment, by increasing the vertical resolution to 3 layers, Charney was able to catch the cyclogenesis that occurred over the central and eastern United States in association with an intense storm that occurred in November of 1950. As Charney later recalled,

> "The two-and-a-half-dimensional model did not catch the cyclogenesis, [although] there was some vague indication of something going on, and so we went to a three-level model, that is, a two-and-two-thirds dimensional model, and we did catch the cyclogenesis. It wasn't terribly accurate, but there was no question that [we did]. And I always thought that this was a terribly important thing....I wanted the world to know about that!" (Dyson 2012)

Not long after this success, Charney was aware that the barotropic primitive equations could be applied by setting the mass divergence to zero at the initial time of an integration of the atmosphere. When applying the barotropic primitive equations, in this manner, Charney and Freeman observed that the winds adjusted to the mass fields that are defined by the geopotential heights of pressure surfaces. As such, the primitive equations would become the basis for NWP.

1.3 Initial Conditions for NWP: A Brief Historical account

To extend the useful forecast period to 3 days and beyond, Charney also understood the need to apply physical parameterizations to account for turbulent and diabatic processes. However, it was Edward Lorenz (Lorenz 1963) that characterized and explained the limits of weather predictability imposed by the accuracy of initial conditions. Lorenz accidentally discovered "The Butterfly Effect" or "sensitive dependence on initial conditions", when starting a numerical integration of a toy model from the printout of initial conditions that were realized midway through a previous

integration. “In the computer’s memory, six decimal places were stored: .506127. On the printout, to save space, just three appeared: .506” (Gleick 1988). Lorenz assumed the difference in precision between the printout and the computer’s stored precision would be inconsequential. This turned out to not be the case, as the new simulation starting from the printout rapidly diverged from the previous integration. “Yet as he stared at the new printout, Lorenz saw his weather diverging so rapidly from the pattern of the last run that, within just a few months, all resemblance had disappeared” (Gleick 1988). To explain the “sensitive dependence on initial conditions”, Lorenz integrated an unrealistically simple set of equations for convection. As with his toy model of the atmosphere, small perturbations in the initial conditions caused the integrations of his convective scheme to diverge. Lorenz understood the cause of this “sensitive dependence on initial conditions” was the presence of nonlinear terms, “There are some nonlinear terms in them, but you think there must be a way to get around them. But you just can’t” (Gleick 1988). For this reason, the extent to which a numerical weather forecast is useful depends significantly on the accuracy of the initial conditions for the integration.

The ability to achieve useful initial conditions for the NWP problem depends on the availability of a sufficiently supportive and coordinated observing system. In the 19th century, the invention of the telegraph enabled coordination of weather observations at distant locations. When the headquarters of the Washington, D.C. Smithsonian was opened in 1855, a giant map of the United States with pinned weather conditions was displayed in the lobby (Blum 2018). Following a tragic 1859 shipwreck in Wales; Robert Fitzroy, the captain of Charles Darwin’s Beagle, organized a predictive system based on coordination of surface observations at distant locations. Fitzroy introduced

the concept of synoptic charts which were constructed by organizing observations at distant locations, and his Meteorological office developed a system whereby observations were received and forecasts were returned using telegraphy. This predictive system was useful in "protecting the burgeoning steamship traffic in the Victorian era" (Blum 2018), and was based on the experience that near future weather at distant locations could be qualitatively inferred from synoptic charts with some confidence. During the 20th century, the problem of determining initial conditions for operational NWP systems has led to the establishment of atmospheric data assimilation as a scientific field. The early NWP experiments led by Charney utilized hand interpolation of available observations as the basis for assigning initial conditions. This process was time consuming, and motivated the development and utilization of automated interpolation methods. More importantly, the number of available degrees of freedom in model space easily exceeds what observations are able to support in modern NWP. As such, the four dimensional data assimilation (4DDA) NWP paradigm emerged in the late 20th century, whereby short range forecasts are used as background for the initialization. The background in this context provides useful information pertaining to the "weather of the day" in observation sparse regions by propagating information content from observations across intermittent data assimilation cycles.

1.4 An Overview of Forecast-Sensitivity Based Observation Impact Estimation in NWP

Enabled by technological developments, the global observing system underwent a significant transformation from what was largely a limited 2 dimensional account of the atmosphere to a far more comprehensive 3 dimensional account during the 20th century. In particular, the introduction and proliferation of radiosonde, aircraft, radio detection and ranging (RADAR) and satellite observations facilitated this transformation. Current operational global NWP systems generally assimilate on the order of 10 million observations every 6 hours. There is substantial diversity in the sampling characteristics and associated error characteristics of the observation types that contribute to the modern global observing system.

In addition to the characteristics of the instrumentation associated with observations, such as signal to noise, the observational impacts on forecasts depend on how the observations are assimilated, the rate of model state forecast error growth coincident with the observations, the accuracy of so called observer codes in constructing the observation associated innovations and the extent of forecast model biases, among other considerations. Given the nature of the global observing system and the established sensitivity of NWP forecast error to initial conditions; the development of methodologies to quantify, organize and explain the impacts of assimilating observations in NWP systems has been of considerable interest in recent decades. Traditionally, Observing System Experiments (OSEs) (Bauer 2009, Gelaro and Zhu 2009, Lord et al. 2016) have

been applied with the intent to obtain the cycled forecast system error changes that can be attributed solely to observing system decisions. In the OSE approach, a data assimilation system is executed to produce analyses for any arbitrary observing system configuration. The cycled forecast errors realized for different observing system configurations can then be compared to quantify the forecast impact of observing system decisions. Although the OSE approach can be useful in providing a measure of net observation forecast error impact with respect to the choice of verification, it is relatively inefficient, since it requires multiple executions of analysis systems and forecast models to merely obtain gross forecast error impact statistics (Gelaro and Zhu 2009, Kalnay et al. 2012, Langland and Baker 2004).

In what is generally considered a seminal paper on the topic of forecast-sensitivity based techniques, (Langland and Baker 2004), developed an adjoint-based Forecast Sensitivity to Observation Impact (FSOI) formulation. As with the corresponding adjoint-based FSOI formulation, the ensemble-based Ensemble Forecast Sensitivity to Observation (EFSO) (Kalnay et al. 2012, Liu and Kalnay 2008) impact estimation approach effectively enables a simultaneous computation of estimated forecast impacts and sensitivities for any and all individual observations assimilated in a numerical weather prediction (NWP) system. More recently, Buehner et al. 2018 introduced approaches to forecast-sensitivity based observation impact estimation that are applicable to ensemble-variational data assimilation systems. Chen 2018 pointed out that these approaches to observation impact estimation collectively represent a generic set of FSO methodologies. The generic FSO methodologies are common in that they map analysis increment contributions associated with assimilation of individual observations to forecast error

changes using tangent-linear approximations to forecast models. The analysis increment, $\mathbf{x}_a - \mathbf{x}_b = \mathbf{K}\delta\mathbf{y}$, is the difference between analysis ($\mathbf{x}_a$) and background forecast ($\mathbf{x}_b$) states, and is computed according to the product of the Kalman gain matrix $\mathbf{K}$ and innovation vector $\delta\mathbf{y}$. As such, generic FSO methodologies fundamentally provide innovation impact estimation (Hotta et al. 2017a) as opposed to observation impact estimation, since in practice it is not realistic to assume an absence of observation operator error contributions to the innovation vector, among other considerations.

We limit the results and associated discussion in this dissertation to EFSO, but note that the concepts developed in this study are applicable to any of the existing generic FSO methodologies. The EFSO formulation has been applied in the National Centers of Environmental Protection (NCEP) Global Forecast System (GFS). For the GFS, this approach to observation impact estimation requires Ensemble Square Root Filter (EnSRF) (Whitaker and Hamill 2002) data assimilation products as input, and it has been implemented within the source code that provides EnSRF functionality at NCEP (Ota et al. 2013). The EFSO formulation incorporates the relationship between Kalman gain and analysis-error covariance to construct observational increments that can be projected forward in time with a forecast model, enabling an estimate of the forecast impact due to assimilating individual observations. The ensemble of analyses resulting from the applicable EnSRF update can be used in the representation of analysis-error covariance, and accordingly the Kalman gain.

1.5 Objectives

In comparing adjoint FSOI and OSE impact methodologies Gelaro and Zhu 2009 provided a general outline of considerations for interpreting adjoint-based FSOI. In particular, they noted that FSOI quantities represent impact estimation with "respect to all other observations assimilated simultaneously", and perhaps more importantly to their study, FSOI quantities do not account for the cumulative impact of observation types over successive data assimilation cycles in NWP systems. In the context of exploring how to best apply EFSO information to improve NWP systems, Chen and Kalnay 2019, quantitatively demonstrated that EFSO-based Proactive Quality Control (PQC) (Chen and Kalnay 2019, 2020, Hotta et al. 2017a, Ota et al. 2013) algorithms involving relatively small (or no) modifications to the Kalman gain matrix were consistently more effective at reducing forecast and analysis RMSE compared to PQC algorithms involving relatively large modifications to the Kalman gain matrix using a Lorenz '96 (Lorenz and Emanuel 1998) based EFSO application testing apparatus. As such, from an NWP application perspective, they demonstrated that aspects of generic FSO estimation relating to direct interactions between the realized set or sets of observations assimilated and the data assimilation approach are a potentially significant consideration. Collectively, these considerations are indicative of diagnostic contributions to generic FSO that are relevant to a realized set or sets of observations assimilated (i.e. "what we are doing") and associated interactions with the data assimilation employed, among other considerations. However, from an NWP application standpoint we are often most interested in information that is "predictive" (e.g. What set or sets of observations,

if rejected, would best serve to improve the skill of numerical weather forecasts? How should assimilated observations be corrected to remove systematic errors? etc.). Specifically, we take the approach that this "predictive" information is related to the nature of errors and flaws in innovations as opposed to contextual information that tends to be more "diagnostic" from an application standpoint. The primary intent of this dissertation is to introduce and demonstrate the utility, for the first time, of a quantitative framework and approach directed towards prioritizing extraction of such "predictive" information from generic FSO impact estimation quantities. Towards this objective, we introduce an "EFSO-components" expression that can be obtained by decomposing forecast error and innovation terms into random and systematic contributions. Summarizing, this "EFSO-components" expression elucidates the nature of interactions between forecast error and innovation terms, and in so doing, enables more efficient extraction of "predictive" information that is tied directly to the nature of errors and flaws in innovations. As such, we address the following questions with the intent of advancing EFSO-based NWP application concepts.

- What are the mechanisms by which EFSO-component terms manifest themselves in EFSO statistics?
- From an application standpoint, What advantages do "EFSO-components" provide in identifying the nature of flaws in innovations?
- Quantitatively, what are the advantages of "EFSO-components" in detecting innovation flaws beyond what can be achieved with EFSO?

- What new insights do “EFSO-components” provide in comparing analysis impact and EFSO impact information?
- To what extent can “EFSO-components” provide PQC application advantages beyond EFSO?
- To what extent can “EFSO-components” help in the selection of assimilated observation types and the planning of observing systems?

The secondary purpose of this dissertation is to test a straightforward extension of the EFSO formulation, we call predicted EFSO (PEFSO), which enables prediction of EFSO for unassimilated observation sets. From an NWP observing system design perspective, this technique represents a capability to compare the efficiency of multiple observing systems in reducing forecast error without having to perform additional analyses and forecasts. Furthermore, data selection procedures employed in operational NWP systems (e.g. the EnSRF update component of the 2016 4DEnVar GFS) can be a complication in obtaining representative EFSO summary statistics. In such circumstances, the ability to predict EFSO for discarded observations enables impact estimation for a more complete set of observations. As such, in this dissertation we address the following questions as it relates to PEFSO.

- What is the robustness of PEFSO in emulation of EFSO?
- To what extent can PEFSO provide the ability to quantitatively compare the efficiency of observing systems in reducing forecast error?

1.6 Outline

This dissertation is structured as follows. The "EFSO-components" technique and its associated conceptual purpose in extracting EFSO impact information directly tied to the extent of flaws in innovations is introduced and explained in chapter 2, following a brief review of the EFSO methodology. Additionally, the PEFSO technique and its intended applications are discussed in chapter 2. In chapter 3, the Lorenz '96 EFSO observing system simulation experiment (OSSE) testing apparatus (Chen and Kalnay 2019) used to explore the concepts introduced in this study is described, and a variety of flawed observation experiments are assessed to explain and quantify the fundamental advantages of "EFSO-components" in detecting innovation flaws beyond what can be achieved with EFSO. The potential NWP utility of "EFSO-components" as a basis for PQC is explored in chapter 4 based on a variety of experimental Lorenz '96 testing configurations. In chapter 5, the robustness of PEFSO in emulating EFSO is explored using the Lorenz '96 OSSE testing apparatus, and sample PEFSO calculations from a pseudo operational GFS are provided. Finally, a summary of the dissertation and future directions for this work are presented in chapter 6.

Chapter 2: EFSO-Components and PEFSO

The purpose of this study is to introduce and test the utility of a new quantitative framework that relates forecast-sensitivity based information directly to the nature of errors and flaws in observations. We limit the explanations and results in this dissertation to the EFSO approach, but there is no reason the new quantitative framework would not be applicable to the other generic FSO methodologies. In this chapter, a brief review of the EFSO formulation is followed by an introduction to the new quantitative framework we call "EFSO-components". We closely follow the labeling and nomenclature conventions used in Ota et al. 2013 in providing a review of the EFSO formulation (Kalnay et al. 2012). A complementary technique to the concept of "EFSO-components" we call predicted-EFSO (PEFSO) is also introduced in this chapter. The explanations of these topics are designed to serve as context for interpreting the results chapters of this dissertation.

2.1 The EFSO Formulation

The forecast errors applied in the EFSO formulation are;

$$\mathbf{e}_t^f = \overline{\mathbf{x}}_t^f - \mathbf{x}_t^v \tag{2.1}$$

$$\mathbf{e}_t^g = \overline{\mathbf{x}}_t^g - \mathbf{x}_t^v \tag{2.2}$$

where $\overline{\mathbf{x}}_t^f$ is the ensemble mean forecast state vector after assimilating observations, $\overline{\mathbf{x}}_t^g$ is the ensemble mean forecast state vector before assimilating observations and $\mathbf{x}_t^v$ is the choice of verification at an evaluation forecast time, t. A practical choice for verification is often times the analysis state vector. Given these forecast errors, the quadratic forecast error reduction from assimilating all observations in a data assimilation cycle is defined as;

$$\Delta(\mathbf{e}^2)^{[f-g]} = \mathbf{e}_t^{f\mathbf{T}}\mathbf{e}_t^f - \mathbf{e}_t^{g\mathbf{T}}\mathbf{e}_t^g = (\mathbf{e}_t^f - \mathbf{e}_t^g)^{\mathbf{T}}\mathbf{C}(\mathbf{e}_t^f + \mathbf{e}_t^g), \tag{2.3}$$

where $\mathbf{C}$ is the norm operator that transforms the state vector into scalar quantities, for operational NWP systems the dry or moist total energy norms (Ehrendorfer et al. 1999) are often times used. The forecast error difference term in the quadratic forecast error reduction equation can be estimated by projecting analysis increments associated with assimilation of observations to an evaluation forecast time using a tangent-linear approximation to the forecast model. For EFSO, ensembles of analyses are used in the construction of the relevant Kalman gain matrix

$$\mathbf{K} = \mathbf{A}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1} = \frac{1}{K-1}\mathbf{X}_0^a\mathbf{X}_0^{a\mathbf{T}}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}, \tag{2.4}$$

where $\mathbf{A} = \frac{1}{K-1}\mathbf{X}_0^a\mathbf{X}_0^{a\mathbf{T}}$ is the analysis error covariance, K represents the number of ensembles, $\mathbf{X}_0^a$ are the associated K member ensemble of analysis perturbations, $\mathbf{H}$ is the linearized forward operator and $\mathbf{R}$ is the observational error covariance. Accordingly, the

forecast error difference is estimated with the following expression

$$\mathbf{e}_t^f - \mathbf{e}_t^g = \mathcal{M}(\overline{\mathbf{x}}_0^a) - \mathcal{M}(\overline{\mathbf{x}}_0^b) \approx \mathbf{M}(\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_0^b) = \mathbf{MK}\delta\mathbf{y} = \frac{1}{K-1}[\mathbf{X}_t^f \mathbf{H}\mathbf{X}_0^{a\mathbf{T}}\mathbf{R}^{-1}\delta\mathbf{y}], \tag{2.5}$$

where $\mathcal{M}$ is the non-linear forecast model, $\mathbf{M}$ is the tangent-linear approximation to the forecast model, $\overline{\mathbf{x}}_0^a$ is the analysis at the initial time, $\overline{\mathbf{x}}_0^b$ is the background forecast at the initial time, $\delta\mathbf{y}$ is the innovation vector and $\mathbf{X}_t^f$ represents the K perturbations of the forecast ensemble initialized from the ensemble of analyses and valid at the evaluation forecast time. Substituting the estimation of forecast error difference into 2.3 we arrive at the general EFSO formulation

$$\Delta(\mathbf{e}^2)^{[f-g]} \approx \delta\mathbf{y}^{\mathbf{T}} \frac{1}{K-1}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}\mathbf{C}(\mathbf{e}_t^f + \mathbf{e}_t^g). \tag{2.6}$$

The EFSO estimated impacts for individual observations are the elements of the inner product between sensitivity to innovation $[\frac{1}{K-1}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}\mathbf{C}(\mathbf{e}_t^f + \mathbf{e}_t^g)]_l$ and the innovation vector $\delta\mathbf{y}$, such that the EFSO estimated quadratic forecast error reduction for the l_{th} observation is $\delta y_l[\frac{1}{K-1}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_0^{f\mathbf{T}}\mathbf{C}(\mathbf{e}_t^f + \mathbf{e}_t^g)]_\mathbf{l}$. Negative impact estimation quantities imply that observations have reduced the magnitudes of forecast errors, and positive impact estimation quantities imply that observations have increased the magnitudes of forecast errors. On this basis, we will hereafter refer to negative impact estimation quantities as being beneficial and positive impact estimation quantities as being detrimental.

To understand how the EFSO formulation assigns detrimental and beneficial

directions for innovations, it is helpful to consider the properties of the EFSO formulation that relate to directionality. The cross-covariance between analysis perturbations in observation space and forecast perturbations in state space, $[\frac{1}{K-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_0^{f\mathbf{T}}]$, effectively specifies forecast perturbation directions in state space for given perturbation directions in observation space, and the sum of forecast errors, $[\mathbf{e}_t^f + \mathbf{e}_t^g]$, defines the helpful (harmful) forecast perturbation directions (i.e. the directions in which forecast errors decrease (increase)) for the segment of state space relevant to the assimilation of an individual observation. Note that in the context of generic FSO the innovations effectively include and specify the observation perturbation directions in observation space. Accordingly, the combined information from the cross-covariance and sum of forecast errors as employed in the EFSO formulation assigns detrimental and beneficial innovation directions. The implication of this EFSO directionality explanation is that the ensemble cross-covariance based projection employed in the EFSO formulation serves to provide information as to "**relative**" observation positions, according to their associated innovations, w.r.t. the **forecasts from background** valid at the evaluation forecast time, t. To help visualize this, consider the "EFSO relative position map" diagrams in figure 2.1, which are representative of the four possible sign combinations of the cross-covariance and sum of forecast errors employed in EFSO. In the figure 2.1 diagrams, forecast error is plotted on the abscissa and quadratic forecast error is plotted on the ordinate. As such, negative forecast perturbations are directed to the left and positive forecast perturbations are directed to the right. The cyan boxes imply a cross-covariance sign such that the innovation sign (i.e. positive or negative) is the same as the forecast perturbation sign when the cross-covariance is positive, and the innovation sign is opposite the forecast

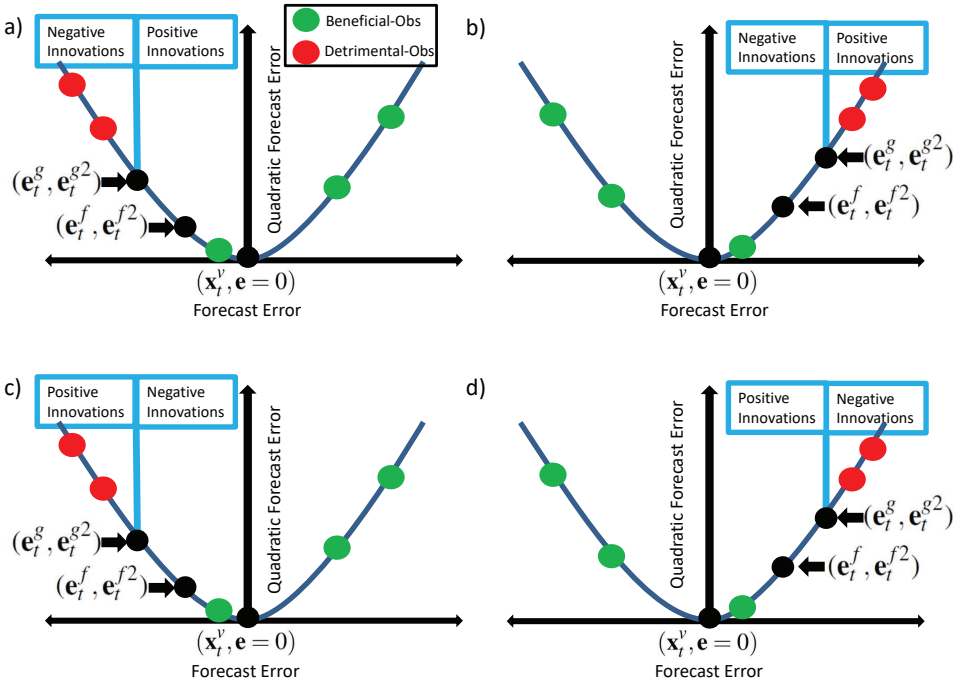

Figure 2.1: Figure diagrams representing the four possible circumstances according to EFSO directionality properties, cyan boxes imply a cross-covariance sign such that the innovation sign (i.e. positive or negative) is the same as the forecast perturbation sign when the cross-covariance is positive, and the innovation sign is opposite the forecast perturbation sign when the cross-covariance is negative.: a) cross-covariances are positive quantities and sum of forecast errors are negative quantities, b) cross-covariances and sum of forecast errors are positive quantities, c) cross-covariances and sum of forecast errors are negative quantities, d) cross-covariances are negative quantities and sum of forecast errors are positive quantities.

perturbation sign when the cross-covariance is negative. According to this framework for explaining EFSO directionality, observations are beneficial (detrimental) as indicated by green (red) dots for circumstances where their "relative" position is in the same (opposite) direction as verification w.r.t. the sum of forecast errors.

2.2 EFSO-Components

For simplification, the sum of forecast errors used in EFSO can be consolidated into one term $\mathbf{e}_t^{f+g} = \mathbf{e}_t^f + \mathbf{e}_t^g$. Rearranging the scalar innovation term so that it is adjacent to the sum of forecast errors term and consolidating all relevant covariances with the Kalman gain, the EFSO formulation for a subset of l individual observation impact estimates can be expressed as:

$$\Delta(\mathbf{e}^2)^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_t^{f+g}\delta\mathbf{y})])_l. \tag{2.7}$$

The sum of forecast errors and innovation terms can then be decomposed into random and systematic components,

$$\delta y = \delta y_r + \delta y_s \tag{2.8}$$

$$\mathbf{e}_t^{f+g} = \mathbf{e}_{[t,a]}^{f+g} + \mathbf{e}_{[t,s]}^{f+g}, \tag{2.9}$$

where $\mathbf{e}_{[t,a]}^{f+g}$ is the forecast error anomaly, $\mathbf{e}_{[t,s]}^{f+g}$ is the systematic contribution to forecast errors, δy_r is the random contribution to innovations and δy_s is the systematic contribution to innovations. Substituting 2.8 and 2.9 into 2.7 and expanding the expression leads to a decomposed form of the EFSO formulation for a subset of l observations.

$$\Delta(\mathbf{e}^2)^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,a]}^{f+g} + \mathbf{e}_{[t,s]}^{f+g})(\delta y_r + \delta y_s)])_l. \tag{2.10}$$

$$\Delta(\mathbf{e}^2)^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,a]}^{f+g}\delta y_r + \mathbf{e}_{[t,a]}^{f+g}\delta y_s + \mathbf{e}_{[t,s]}^{f+g}\delta y_r + \mathbf{e}_{[t,s]}^{f+g}\delta y_s)])_l. \tag{2.11}$$

From left to right, the four component contributions to EFSO in the "EFSO-components" expression are:

1. products of forecast error anomalies and random innovation contributions, $\mathbf{e}_{[t,a]}^{f+g}\delta y_r$. We hereafter refer to this EFSO contribution as **random observation projection** onto forecast error **anomalies** (ROPA).

$$\Delta(\mathbf{e}^2)_{ROPA}^{[f-g]} = \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,a]}^{f+g}\delta y_r)])_l. \tag{2.12}$$

2. products of forecast error anomalies and systematic innovation offsets, $\mathbf{e}_{[t,a]}^{f+g}\delta y_s$. We hereafter refer to this contribution as **systematic observation projection** onto forecast error **anomalies** (SOPA).

$$\Delta(\mathbf{e}^2)_{SOPA}^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,a]}^{f+g}\delta y_s)])_l. \tag{2.13}$$

3. products of systematic forecast error and random innovation offsets, $\mathbf{e}_{[t,s]}^{f+g}\delta y_r$. We hereafter refer to this contribution as **random observation projection** onto

systematic forecast error (ROPS).

$$\Delta(\mathbf{e}^2)^{[f-g]}_{ROPS} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}^{f+g}_{[t,s]}\delta y_r)])_l. \tag{2.14}$$

4. products of systematic forecast error and systematic innovation offsets, $\mathbf{e}^{f+g}_{[t,s]}\delta y_s$. We hereafter refer to this contribution as **systematic observation projection** onto **systematic** forecast error (SOPS).

$$\Delta(\mathbf{e}^2)^{[f-g]}_{SOPS} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}^{f+g}_{[t,s]}\delta y_s)])_l. \tag{2.15}$$

According to Reynolds averaging, the SOPA and ROPS EFSO components will not provide aggregate impact in the absence of correlations with the projection of the Kalman gain matrix to the evaluation forecast time ($\frac{1}{K-1}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}^a_0\mathbf{X}^{f\mathbf{T}}$), since these components represent products of zero mean anomalies and systematic offsets.

For the purpose of explaining the intended advantages of this EFSO decomposition, we can recognize that the innovation includes both corrective and error contributions in practice,

$$\delta y_r = \delta y_{rc} + \delta y_{re} \tag{2.16}$$

$$\delta y_s = \delta y_{sc} + \delta y_{se}, \tag{2.17}$$

where δy_{rc} and δy_{sc} are the innovation contributions that relate to "corrective" information for systematic and random model-error represented in observation space, and δy_{re} and

δy_{se} are the random and systematic observation "error" contributions to innovations. Substituting 2.12 and 2.13 into 2.11, the decomposition of EFSO can be expanded further to recognize that there are corrective and error contributions for the defined EFSO components.

$$\begin{aligned}\Delta(\mathbf{e}^2)^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{MK}]^{\mathbf{T}}\mathbf{C}[((\mathbf{e}_{[t,a]}^{f+g}\delta y_{rc} + \mathbf{e}_{[t,a]}^{f+g}\delta y_{re}) + \\ (\mathbf{e}_{[t,a]}^{f+g}\delta y_{sc} + \mathbf{e}_{[t,a]}^{f+g}\delta y_{se}) + \\ (\mathbf{e}_{[t,s]}^{f+g}\delta y_{rc} + \mathbf{e}_{[t,s]}^{f+g}\delta y_{re}) + \\ (\mathbf{e}_{[t,s]}^{f+g}\delta y_{sc} + \mathbf{e}_{[t,s]}^{f+g}\delta y_{se}))])_l,\end{aligned} \tag{2.18}$$

Contemplating the implications of this expression, we expect net detriment from components where the error contributions to the innovation are large relative to the corrective contributions.

From a visual perspective, a pared down version of the "relative position map" concept is helpful in identifying "corrective" and "error" contributions to innovations as would be represented in state space. In figure 2.2, the differences between background forecasts and verification effectively correspond to "corrective" contributions in innovations, and the differences between projected observation positions in state space and verification effectively correspond to "error" contributions in innovations. While it is not the case for all possible "relative position" configurations, innovations will tend to be beneficial (detrimental) when the "error" contribution is smaller (larger) than the "corrective" contribution. The purpose of developing "EFSO-components" as

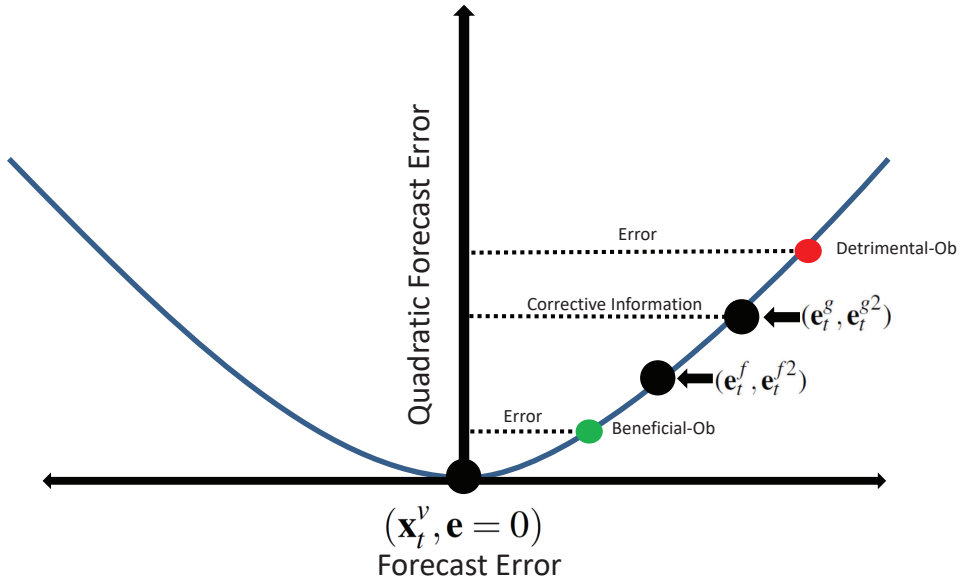

Figure 2.2: A visual account of "corrective" and "error" innovation contributions.

a new framework that can advance application of EFSO and generic FSO is twofold. First, by distributing "corrective" innovation contributions it is intended that "EFSO-components" can improve detection of specific innovation errors and flaws beyond what can be achieved with EFSO. In chapter 3, we explore and explain this fundamental advantage using Lorenz '96 flawed observation experiments. Secondly, in the context of PQC applications, (Chen and Kalnay 2019, 2020, Hotta et al. 2017a, Ota et al. 2013), we expect observations which have no compensating benefit (i.e. all substantive "EFSO-component" contributions are detrimental) can be rejected more aggressively while avoiding the "unstable error growth" explained according to a thought experiment introduced in Chen and Kalnay 2019. In chapter 4, we demonstrate and explore this advantage based on PQC experiments for the Lorenz '96 simple model (e.g. figures 4.5 and 4.9) and again return to "relative position map" concepts in explaining the "unstable error growth" that "EFSO-components" based rejection more effectively avoids (i.e. figure 4.3).

2.3 PEFSO and PEFSO Components

From an application standpoint, there is potential utility in being able to predict EFSO or "contextual" aspects of the EFSO calculation for unassimilated observation sets. For example, in the context of observing system design efforts, such a capability would make it possible to quantitatively compare the overall efficiency of observing systems in reducing forecast error without having to run multiple cycling experiments. Furthermore, a difficulty in achieving representative EFSO summary statistics for operational NWP

systems is the application of data selection procedures (e.g. the EnSRF update component of the 2016 4DEnVar GFS). In such circumstances, the ability to predict EFSO for discarded observations enables impact estimation for a more complete set of observations. As such, we propose a new technique we call predicted EFSO (PEFSO), with the purpose of adequately predicting EFSO information for unassimilated observation sets.

The PEFSO methodology is a straightforward extension of the EFSO formulation, and carries out the EFSO calculation with the exception that observation impact estimates are for observation sets that did not contribute to the employed cross-covariances and forecasts from analyses, such that

$$\begin{aligned}\Delta(\mathbf{e}^2)^{[f-g]} \approx \delta\mathbf{y}^{\mathbf{T}}\frac{1}{K-1}\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}\mathbf{C}(\mathbf{e}_t^f+\mathbf{e}_t^g) \approx \\ \delta\mathbf{y}^{\mathbf{T}}\frac{1}{K-1}\mathbf{R}^{-1}[\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}]_{\mathbf{u}}\mathbf{C}(\mathbf{e}_{t,\mathbf{u}}^f+\mathbf{e}_t^g)\end{aligned} \tag{2.19}$$

, where $[\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}]_{\mathbf{u}}$ and $\mathbf{e}_{t,\mathbf{u}}^f$ are the cross-covariance and forecast error quantities for the unassimilated PEFSO observation sets. Similarly, the PEFSO components methodology can be expressed as,

$$\begin{aligned}\Delta(\mathbf{e}^2)^{[f-g]} \approx \sum_{n=1}^{l}([\mathbf{M}\mathbf{K}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,r]}^{f+g}\delta y_r+\mathbf{e}_{[t,r]}^{f+g}\delta y_s+\mathbf{e}_{[t,s]}^{f+g}\delta y_r+\mathbf{e}_{[t,s]}^{f+g}\delta y_s)])_l \approx \\ \sum_{n=1}^{l}([\mathbf{M}\mathbf{K}_{\mathbf{u}}]^{\mathbf{T}}\mathbf{C}[(\mathbf{e}_{[t,r,\mathbf{u}]}^{f+g}\delta y_r+\mathbf{e}_{[t,r,\mathbf{u}]}^{f+g}\delta y_s+\mathbf{e}_{[t,s,\mathbf{u}]}^{f+g}\delta y_r+\mathbf{e}_{[t,s,\mathbf{u}]}^{f+g}\delta y_s)])_l\end{aligned} \tag{2.20}$$

In general, it is expected that useful predictions of EFSO and "EFSO-components" can be achieved where the unassimilated observation sets contribute a relatively small change,

had they been assimilated, to the cross-covariances and forecasts from analyses (i.e. $[\mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}]_{\mathbf{u}} \approx \mathbf{H}\mathbf{X}_0^a\mathbf{X}_t^{f\mathbf{T}}$ and $\mathbf{e}_{t,\mathbf{u}}^f \approx \mathbf{e}_t^f$).

By using a common background in all cases, the PEFSO methodology provides a means to control for the "corrective" contribution to innovation that was defined in 2.18. As such, the PEFSO technique provides utility in comparing the efficiency with which observing systems reduce forecast error. In chapter 5 of this dissertation, experiments designed to demonstrate the potential utility this aspect of PEFSO provides for observing system design efforts are examined.

2.4 Summary

In this chapter, a brief review of the EFSO formulation was provided, and a new quantitative framework called "EFSO-components" was introduced. The concept of "EFSO relative position maps" was introduced with the purpose of explaining EFSO directionality. The "EFSO-components" expression was expanded into corrective and error contributions for explanatory purposes. This expanded "EFSO-components" expression implies that EFSO detriment will occur where the magnitude of the error contribution to innovation is large relative to the magnitude of the corrective innovation contribution, and in a statistical context represents an explanation to EFSO directionality that is complementary to "EFSO relative position maps". In the next chapter, we use Lorenz 96' observing system simulation experiments (OSSEs) to explain how "EFSO-components" can extend the utility of EFSO by extracting information directly related to the nature and extent of observation flaws.

Additionally, a new technique called PEFSO was introduced in this chapter. This technique is intended to provide a convenient apparatus for comparing the utility of observing systems in operational NWP systems, and to allow for more comprehensive impact estimation in operational ensemble-based data assimilation settings where data selection procedures are utilized to mitigate computational costs. In chapter 5, Lorenz 96' OSSEs are performed to demonstrate the utility of PEFSO for the aforementioned applications.

Chapter 3: "EFSO-Components": Explanatory Lorenz '96 Model Observing System Simulation Experiments

3.1 Introduction

To explain and quantify the potential utility of "EFSO-components" in better extracting "predictive" EFSO information that is tied more directly to the nature of flaws in the observations as opposed to diagnostic information, it is advantageous to use a testing environment where all errors and flaws can be exactly quantified. Furthermore, to demonstrate potential applicability to the NWP problem, testing requires a model that can simulate some of the large scale atmospheric behavior and error growth. As such, in this dissertation we use the Lorenz '96 (Lorenz and Emanuel 1998) EFSO OSSE testing apparatus explained and introduced in (Chen and Kalnay 2019).

As with the other forecast sensitivity based observation impact estimation methodologies, EFSO calculations provide impact at each analysis time separately (Gelaro and Zhu 2009). As such, EFSO does not account for the cumulative impacts from all previously assimilated observations that contribute to the background. We hereafter refer to this as the "cumulative" contextual aspect of EFSO. Furthermore, in practice, the Kalman gain associated cross-covariances and forecasts from analyses

used in EFSO correspond to the realization of all observations assimilated at each analysis time. Therefore, as with other forecast sensitivity based impact estimation methodologies, EFSO estimated impacts *are with respect to all other observations assimilated simultaneously* (Gelaro and Zhu 2009). We hereafter refer to this as the "simultaneous" contextual aspect of EFSO. In this chapter and chapter 4 of this dissertation, results from Lorenz '96 observing system simulation experiments (OSSEs: Hoffman and Atlas 2016, Zeng et al. 2020), where varying observation flaws are introduced, are used to explain the advantages of "EFSO-components" in limiting the extent to which these "cumulative" and "simultaneous" aspects of EFSO inhibit extraction of information that is tied more closely to the nature of errors and flaws in observations. Note that although the results and discussion pertaining to the "cumulative" and "simultaneous" aspects are limited to EFSO and "EFSO-components" in this dissertation, these concepts are general to the topic of forecast-sensitivity based observation impact estimation.

In operational NWP applications self analyses are often a practical choice of verification for forecast sensitivity based observation impact estimation. The extent to which this practical choice of verification may limit the utility of EFSO estimation depends on the evaluation forecast lead time and the extent of error in assimilated observations, among other considerations. As it relates to this consideration, in this chapter we also explore the extent to which "EFSO-components" provides information in identifying observation flaws that is unavailable from analysis impact approaches.

3.2 The "EFSO-components" Testing Methodology and Experimental Settings

Throughout this study we use the Lorenz 96' model EFSO OSSE testing apparatus introduced and explained in (Chen and Kalnay 2019) to test the potential utility of "EFSO-components". The flow diagram in figure 3.1 identifies the elements of this Lorenz '96 model EFSO OSSE testing apparatus.

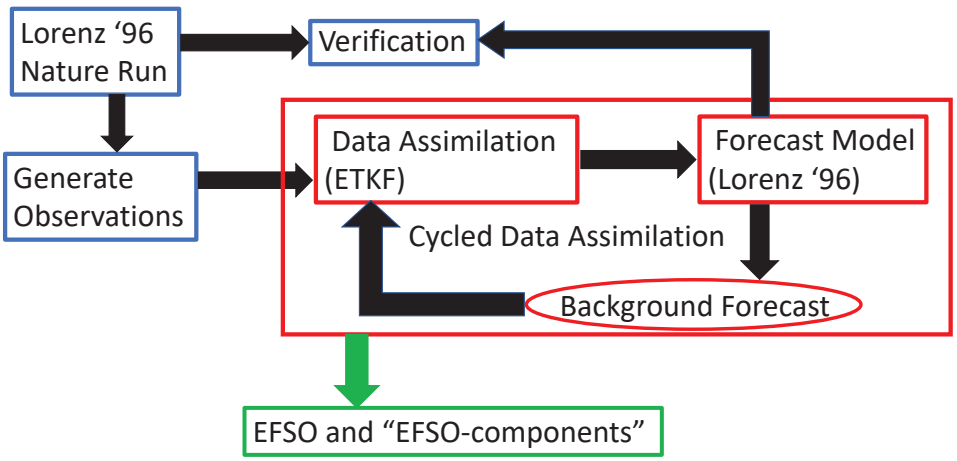

Figure 3.1: Flow diagram explaining the elements of the Lorenz '96 model EFSO OSSE testing apparatus (Chen and Kalnay 2019)

As indicated in figure 3.1, the Lorenz 96' model is used in this OSSE approach

to provide both the nature run (i.e. the "truth") and the background forecasts within the relevant cycled data assimilation. Unless otherwise stated the model configuration for all experimental simulations in this study are identical to those applied for the nature run. The Lorenz '96 model consist of N equations and N variables,

$$\frac{dx_n}{dt} = x_{n-1}(x_{n+1} - x_{n-2}) - x_n + F, \tag{3.1}$$

where on the right-hand side of the equation the first two terms simulate advection, the third term simulates dissipation and the fourth term represents external forcing. As discussed in Lorenz and Emanuel 1998, chaos ensues when the forcing parameter (F) exceeds 4.0. Consistent with that applied in Lorenz 1996, in this study we set the constant forcing term for the nature run and cycled DA system to 8.0 unless otherwise stated. Given this model forcing, the error doubling rate according to the leading Lyapunov exponent λ_1 is 0.42 model time units (Lorenz and Emanuel 1998). Following the testing approach applied in Lorenz and Emanuel 1998 we use a model dimension of $N = 40$, and a fourth order Runge-Kutta time integration scheme with a unit time step of $\Delta t = 0.05$.

Ensemble Kalman filtering (EnKF) is performed in this OSSE testing apparatus to provide the data assimilation. The general approach of EnKF is to use K background ensemble forecasts started from K perturbed initial conditions as a basis for constructing background error covariances,

$$\mathbf{B} = \frac{1}{K-1}\mathbf{X}_0^g\mathbf{X}_0^{g\mathbf{T}} \tag{3.2}$$

where $\mathbf{X}_0^g$ represents the background forecast perturbation matrix, at the analysis time, with columns that account for K differences between background forecast members and the ensemble mean. The Kalman gain matrix $\mathbf{K}$ and analysis equation can then be established by combining $\mathbf{B}$ and the observational error covariance $\mathbf{R}$,

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1} \tag{3.3}$$

$$\mathbf{x}^a = \mathbf{x}^g + \mathbf{K}\delta\mathbf{y} \tag{3.4}$$

where $\mathbf{x}^a$ represents the analysis state and $\mathbf{x}^g$ represents the background state. As indicated in figure 3.1 we use the ETKF, Bishop et al. 2001, implementation of this analysis equation to perform the data assimilation within this OSSE testing apparatus.

The ensemble size for all experiments matches the dimensions (i.e. N=40) of the Lorenz '96 system applied, and as such localization is unnecessary. Furthermore, inflation is not applied within the cycled data assimilation for these "perfect model" experiments. For simplicity, the observation operator applied in all cases is the identity matrix, and contains no error in this controlled testing environment. To allow the ensemble to converge to the attractor, we use a spinup period of 500 unit time steps. After spinup, all experiments are run for 5000 model unit time steps, and the nature run provides verification for all 5000 experimental unit time steps. Following the general OSSE testing methodology, the hypothetical observations assimilated are generated by adding known error characteristics to the Lorenz '96 nature run (i.e. the "truth" run).

The ensembles of analyses and background forecasts from the Lorenz '96 / ETKF cycled data assimilation are then used to perform the EFSO calculations according to

equation 2.6 for all observations assimilated. The EFSO employed sum of forecast errors and innovations are averaged over the 5000 experimental time steps. The random innovation contributions and forecast error anomalies are then calculated by subtracting out the average quantities at each time step. Using these calculated average (i.e. systematic) and random contributions to the innovations and EFSO employed forecast errors, we perform the EFSO component calculations according to equations 2.12, 2.13, 2.14 and 2.15.

3.3 Explanatory "EFSO-components" Experiments

3.3.1 Experimental Overview

The primary purpose of using the experimental settings and testing methodology described in section 3.2 is to assess the extent to which "EFSO-components" provides fundamental advantages in identifying observation flaws and in better extracting the "predictive" information, that was discussed in chapter 1, beyond what can be achieved with EFSO. For all experiments in this chapter the observations are assigned by adding noise from the normal distribution, $\mathcal{N}(\mu, \mathbf{R})$, to the truth run. We assess "EFSO-components" calculations for Lorenz '96 OSSEs where random observation errors and flaws, and systematic observation flaws are introduced separately. In particular, we use these experiments to explain basic quantitative properties of "EFSO-components", and to explain the potential utility of "EFSO-components" in addressing "cumulative" aspects of EFSO that are general to the topic of forecast-sensitivity based observation impact estimation. Furthermore, we discuss fundamental advantages of "EFSO-components" in

extracting information about observation flaws that is unavailable from analysis impact approaches.

With the exceptions explicitly stated in the associated discussion for each experiment, the following specifications apply:

- the prescribed observation error covariance is 0.01 (standard deviation of 0.1)
- the actual assimilated observation error covariance is 0.01 (standard deviation of 0.1)
- there are no systematic errors
- self analyses are used as verification for the EFSO and "EFSO-components" calculations
- the data assimilation window is one time-step in all cases

3.3.2 Random Error Experiments

The 11 time-step realized quadratic forecast error reduction from assimilating observations according to equation 2.3, total cycle EFSO estimated impacts, and the total cycle sums of the four individual EFSO component estimates are plotted in figure 3.2a for an "ideal baseline" OSSE where the magnitudes of the prescribed and actual assimilated random observation errors are the same. Overall there is a reduction in forecast error as would be expected, but note that positive quantities indicate net detriment and negative quantities indicate net benefit from assimilating observations in any particular cycle. We plot the sums of EFSO component contributions as confirmation that the EFSO

component calculations are as intended, and as such, the summation of EFSO component contributions is consistent with the overall EFSO calculation. The negligible impacts from SOPA, ROPS and SOPS according to figure 3.2b are as expected given the absence of systematic errors in the associated experiment. EFSO consistently tracks the true (i.e. realized) error reduction impact from the observations, and as expected, ROPA accounts for all substantive EFSO impact estimation in this configuration.

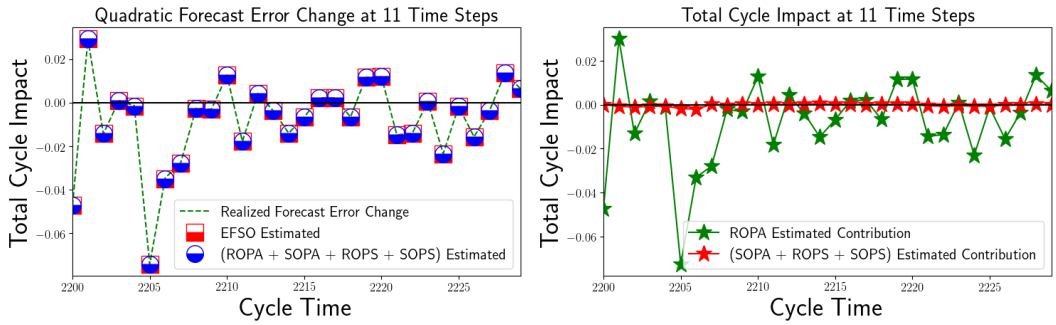

Figure 3.2: Total cycle impact plots for a lead-time of 11 time steps. Positive quantities indicate net detriment and negative quantities indicate net benefit from assimilating observations in any particular cycle. a) Plots of realized forecast error reduction, EFSO estimated net impact and the sums of EFSO estimated component contributions, b) Plots of the ROPA EFSO component, and the summation of the SOPA, ROPS and SOPS EFSO estimated component contributions.

In operational NWP systems the prescribed observational error covariance $\mathbf{R}$ is never exactly or truly known (Hollingsworth and LÖnnberg 1986, Desroziers et al. 2005,Satterfield et al. 2017). As such, it is of interest to examine the utility of EFSO and "EFSO-components" in identifying where the actual assimilated observation error covariance is inconsistent with prescribed observation error covariance. To do so, we assess EFSO when random observation flaws are introduced in spike (e.g. Hotta et al. 2017b), staggered (e.g. Kalnay et al. 2012), adjacent and ubiquitous configurations. For each configuration, we ran experiments where the actual assimilated random observation

error standard deviation is uniformly assigned to be 0.13, 0.16 and 0.19. A detailed description as to the locations of the uniformly assigned random observation flaws is as follows:

a. Spike Configuration: The observation flaws are introduced at the 11th grid-point

b. Staggered Configuration: The observation flaws are introduced at all odd numbered grid-points

c. Adjacent Configuration: The observation flaws are introduced at adjacent grid-points 8 through 13

d. Ubiquitous Configuration: The observation flaws are introduced at all grid-points

The per observation EFSO impacts, for the lead-time of 11 time steps (i.e. the EFSO evaluation time), are plotted in figure 3.3 for the aforementioned experiments. Following the conventions defined in chapter 1, positive quantities indicate per observation detriment and negative quantities indicate per observation benefit. It is only in the spike configuration that EFSO suggest overall detriment where flaws are introduced. Furthermore, it is overwhelmingly the case that the magnitude of EFSO suggested benefit from observations increases with the magnitude of random observation flaws in the staggered, adjacent and ubiquitous configurations. This result does not reflect error in the EFSO calculation, but is instead contextual. Note that the correlation coefficients between actual forecast error changes and the EFSO estimated forecast error changes were >0.99 for all 11 time-step flawed experimental configurations. We propose here that these results are tied to the "cumulative" contextual aspect of the EFSO ROPA component,

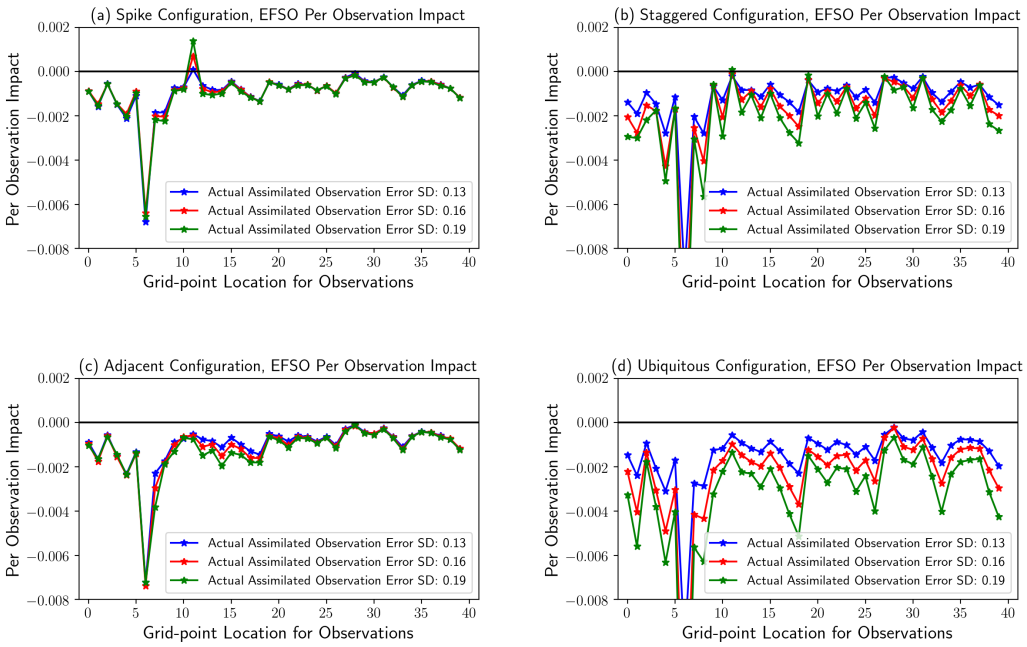

Figure 3.3: Per Observation Impact from observations at the 40 Lorenz '96 grid-points for a lead-time of 11 time steps. Positive quantities indicate per observation detriment and negative quantities indicate per observation benefit. a) Spike configuration experiments: Per observation EFSO where the actual assimilated standard deviation of observation error is 0.13, 0.16 and 0.19 at the 11th grid-point, b) same as (a), but the flaws are introduced in a staggered configuration at all odd grid-points, c) same as (a), but the flaws are introduced at adjacent grid-points 8 through 13, d) same as (a), but the flaws introduced are ubiquitous.

such that the directionality (i.e. detrimental or beneficial) of ROPA contributions, are largely determined by the relative magnitudes of "corrective" and "error" contributions to innovation defined in equation 2.18, and explained according figure diagram 2.2. To demonstrate this is the case, we perform EFSO calculations and analyses with the same set of observations as those represented in figure 3.3, but in all cases used background from the "ideal baseline" configuration (i.e. no observation flaws experiment where the standard deviation of actual observation error is 0.1). The result of removing "cumulative" contributions from observation flaws assimilated in antecedent cycles is

plotted in figure 3.4, and as would be expected, the EFSO estimated observation impacts are overwhelmingly more detrimental or less beneficial as the magnitude of the random error flaws in the observations being evaluated are increased.

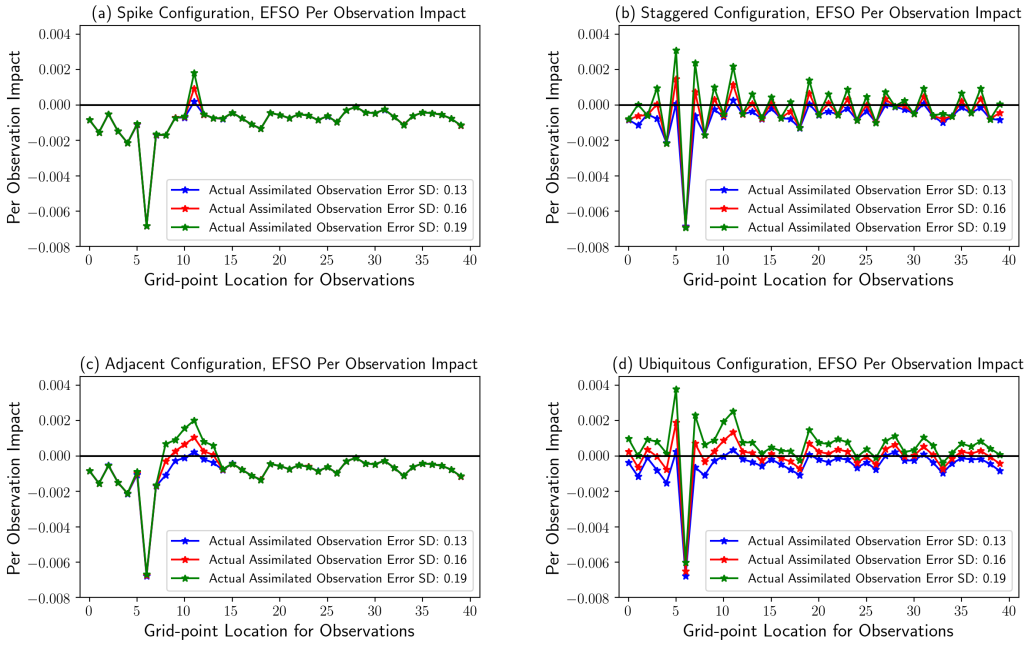

Figure 3.4: Per observation impact, same as 3.3 except the background used in all cases is from the baseline ideal observation configuration. That is, the contributions from assimilating observation flaws in antecedent cycles are removed. Positive quantities indicate per observation detriment and negative quantities indicate per observation benefit. a) Spike configuration experiments: Per observation EFSO where actual assimilated observation error covariance is 0.13, 0.16 and 0.19 at the 11th grid-point, b) same as (a), but observation flaws are introduced in a staggered configuration at all odd grid-points, c) same as (a), but observation flaws are introduced at adjacent grid-points 8 through 13, d) same as (a), but the flaws introduced are ubiquitous.

In "real world" NWP applications, background forecasts that reflect assimilation of an ideal set of observations, similar to the "ideal baseline" experiment in this section, are unavailable. As such, the "cumulative" aspect of EFSO is generally unavoidable in practice. The useful information from EFSO in this context is comparative, and the possible prominence of "cumulative" contributions in any specific circumstance such as

those exhibited in 3.3, should be a consideration when interpreting or applying EFSO in practice. As expected, the "cumulative" aspect of EFSO is more problematic from an application standpoint in the "adjacent" and "ubiquitous" configurations where the truly flawed observations are effectively compared to one another through the background. As such, from a macroscopic perspective, the available comparative information from EFSO and ROPA is not helpful in identifying observation flaws in the "ubiquitous" and "adjacent" configurations. However, for the staggered configuration where the truly flawed observations at the odd numbered grid-points are effectively being compared to the ideal observations (i.e. no flaws) at the even numbered grid points through the background, EFSO calculations provide information that is useful in identifying which observations are more or less flawed. Note that in the staggered configuration the flawed observations located at the odd numbered grid-points are consistently less beneficial than the ideal (i.e. no flaws) observations located at the even numbered grid-points in 3.3b. This demonstrates that even when "cumulative" aspect considerations are significant, EFSO and ROPA provide useful information in comparing the extent of flaws in observations, and accordingly, the relative efficiencies with which observation types or subsets of observations reduce forecast error. For the spike configuration results, plotted in figure 3.3a, EFSO successfully identifies observation flaws at the 11th grid-point. Following the explanation for the identification of random flaws in the staggered experiments, this demonstrates that the "cumulative" aspect of EFSO may be less problematic in identifying where assimilated observations are particularly flawed relative to other nearby assimilated observations. This macroscopic perspective provided in figures 3.3 and 3.4 has relevance in explaining the utility of EFSO and ROPA

directionality in identifying individual observations that have large errors relative to other nearby observations in space and time.

To further investigate the potential significance of the "cumulative" aspect of EFSO and ROPA impact estimates, we assess related explanatory statistics and extend the ubiquitous flaws experiments represented by figures 3.3a and 3.4a to include circumstances where the actual assimilated observation error standard deviation is smaller than that prescribed. As such, summary EFSO, "EFSO-components" and related explanatory statistics are plotted in figure 3.5 for 2 sets of 7 experiments associated with 7 different corresponding standard deviations of observation error assimilated.

The red lines and symbols in figure 3.5 indicate summary results from a set of 7 experiments where observation errors characterized on the abscissa contributed to background forecasts through assimilation in antecedent cycles (i.e. "cycling" experiments). The 11 time-step percent beneficial EFSO statistic for these experiments is notably invariant. This reflects a proportional adjustment of "corrective" contributions to observation errors assimilated over successive cycles. As such, the suggested increase in magnitude of per observation benefit with the errors in observations assimilated is better explained by the increase in forecast RMSE plotted in figure 3.5b.

The blue lines and symbols in figure 3.5 indicate summary results from a set of 7 experiments where the background forecasts applied in all cases is from the "ideal baseline" experiment where prescribed and actual assimilated observation error standard deviation is the same (i.e. blue lines and symbols correspond to "non-cycling" experiments). The 11 time-step percent beneficial EFSO statistic for these experiments is sensitive to the magnitude of observation error assimilated. This reflects

that the "error" contribution to innovation is prominent in EFSO directionality since the "corrective" contribution is static across these experiments. As such, for the "non-cycling" experiments, changes in per observation impact with increasing observation error reflect increasing "error" contributions to innovations relative to the static "corrective" contributions.

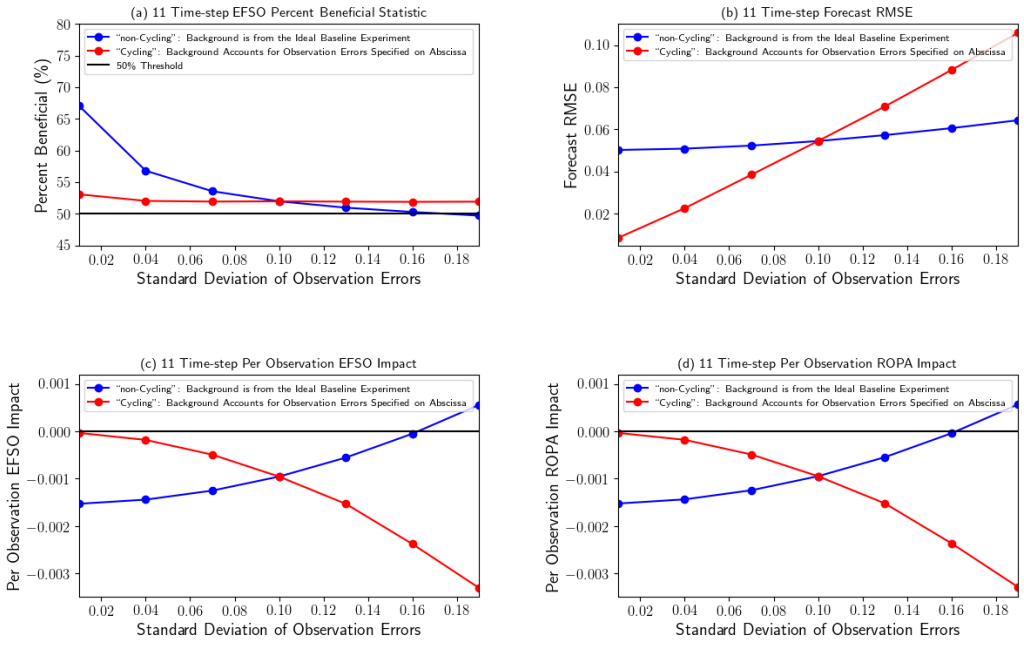

Figure 3.5: Summary results from Lorenz '96 ubiquitous random observation error experiments. The red plots and symbols indicate summary results for a set of 7 experiments where observation errors characterized on the abscissa contributed to background forecasts through assimilation in antecedent cycles (i.e. "cycling" experiments). The blue plots and symbols indicate summary results for experiments where the background forecasts applied in all cases is from the "ideal baseline" experiment where prescribed and actual assimilated observation error standard deviation is 0.1 in all cases (i.e. "non-cycling" experiments). a) The 11 time-step EFSO percent beneficial statistic, b) 11 time-step forecast RMSE (i.e. the forecasts associated with the EFSO and "EFSO-components" estimation), c) Per observation EFSO estimated impacts, d) Per observation ROPA estimated impacts.

In an expected manner, figures 3.5c and 3.5d reflect that ROPA explains all substantive contributions to EFSO in these random error experiments. As such, for

circumstances where there are no systematic errors in the observations or forecast model, which is unrealistic in practice, there is no advantage to using "EFSO-components" for this Lorenz '96 testing system since EFSO and ROPA are interchangeable. This is not to say that the ROPA contribution to EFSO has no utility. In chapter 4 we demonstrate the importance of retaining ROPA information to help mitigate the "simultaneous" contextual aspect for PQC applications.

The key explanatory results from the random observation error experiments presented in this subsection are as follows:

- ROPA directionality reflects comparative magnitudes of "corrective" and "error" contributions in innovations.
- ROPA magnitude in "cycling" experiments is in part explained according to changes in forecast errors at the EFSO (or "EFSO-components") evaluation time
- The "cumulative" contextual aspect of EFSO or ROPA involves adjustments to "corrective" innovation contributions that are proportional to the magnitudes of observation errors assimilated in antecedent cycles.
- The spike, staggered, ubiquitous and adjacent configuration experiments in this subsection represents a macroscopic view as to the application utility of comparative EFSO information in identifying individual observations with large errors relative to nearby observations in space and time.

3.3.3 Systematic Observation Flaws Experiments

The 11 time-step realized quadratic forecast error reduction defined by equation 2.3, total cycle EFSO estimated impacts, and the total cycle sums of the four individual EFSO component estimates are plotted in figure 3.6a for an experiment where we have ubiquitously added systematic flaws of 0.05 to the observations. We will hereafter refer to this as the "baseline positive systematic" experimental configuration. As expected, the sums of the total cycle EFSO component estimates match the total cycle EFSO calculations, confirming the EFSO component calculations are as intended when systematic observation errors are introduced, and EFSO estimation consistently tracks the true (i.e. realized) error reduction associated with assimilation of observations. Figure 3.6b reflects that all substantive contributions to EFSO are from the ROPA and SOPA components. Relative to the "ideal baseline" configuration examined in 3.3.2, where

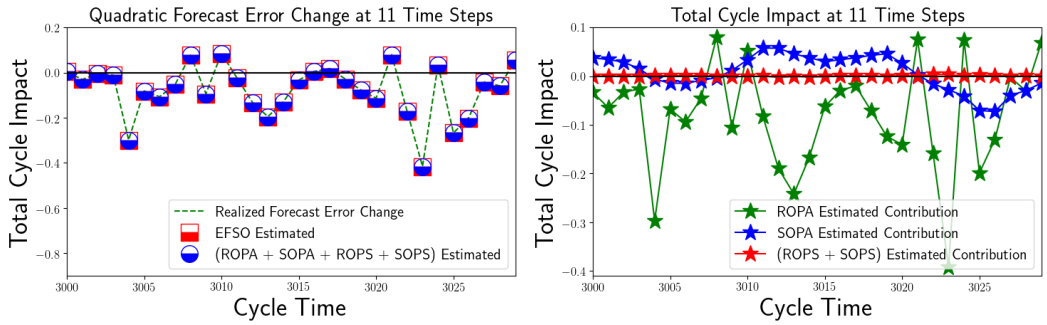

Figure 3.6: Total cycle observation impact plots for a lead-time of 11 time steps. Positive quantities indicate net detriment (i.e. increase in forecast error) and negative quantities indicate net benefit (i.e. decrease in forecast error). Results correspond to an experiment where systematic observation errors of 0.05 were ubiquitously introduced. a) Plots of realized foreast error reduction, EFSO estimated net impact and the sums of EFSO estimated component contributions. This serves as verfication that EFSO and the sum of EFSO component contributions are interchangeable, and that EFSO tracks actual forecast error reduction. b) Plots of the ROPA EFSO component, SOPA EFSO component, and the summation of the ROPS and SOPS estimated component contributions.

no observation flaws were introduced, the increase in the 11 time-step forecast error standard deviation was approximately an order of magnitude larger than the increase in systematic forecast error (i.e. forecast bias). As such, assimilation of the observations with systematic flaws are primarily projecting onto forecast error anomalies as opposed to systematic forecast error in this Lorenz '96 testing system. This result is consistent with there being substantive impact from SOPA, as opposed to ROPS and SOPS, since SOPA represents mechanisms by which assimilated systematic innovation contributions (i.e. biases) project onto forecast error anomalies. Furthermore, for this 30 cycle sample, figure 3.6 indicates that SOPA tended to be detrimental whereas both EFSO and ROPA were beneficial overall, suggesting that the SOPA component provides utility in identifying the systematic observation flaws introduced.

In operational NWP systems the extent and variation of systematic errors across assimilated observation types (or observing systems) and observables is never truly known. As such, significant efforts have been made to quantify and correct systematic observation errors assimilated in operational NWP systems and other "real world" data assimilation applications. The most common approaches use predictors that are dependent on instrument type and observable, where the associated predictor coefficients are then based on the use of nearby observations as predictands (e.g. Dee et al. 2008, Zhu et al. 2015, Sun et al. 2013). However, the utility of these methodologies relies on the extent to which accurate nearby "anchoring" observations are available in both space and time. As such, it is worth exploring the potential utility of EFSO and "EFSO-components" in providing independent information that can be used to quantify and correct for systematic errors in observations. To do so, we assess EFSO and

"EFSO-components" calculations for experiments where systematic flaws of 0.05 were introduced in the ubiquitous (i.e. "baseline positive systematic"), staggered, adjacent and spike manner outlined in section 3.3.2. The per observation impacts for the aforementioned systematic observation error experimental configurations are plotted in figure 3.7. It is apparent that the SOPA component consistently identifies systematic

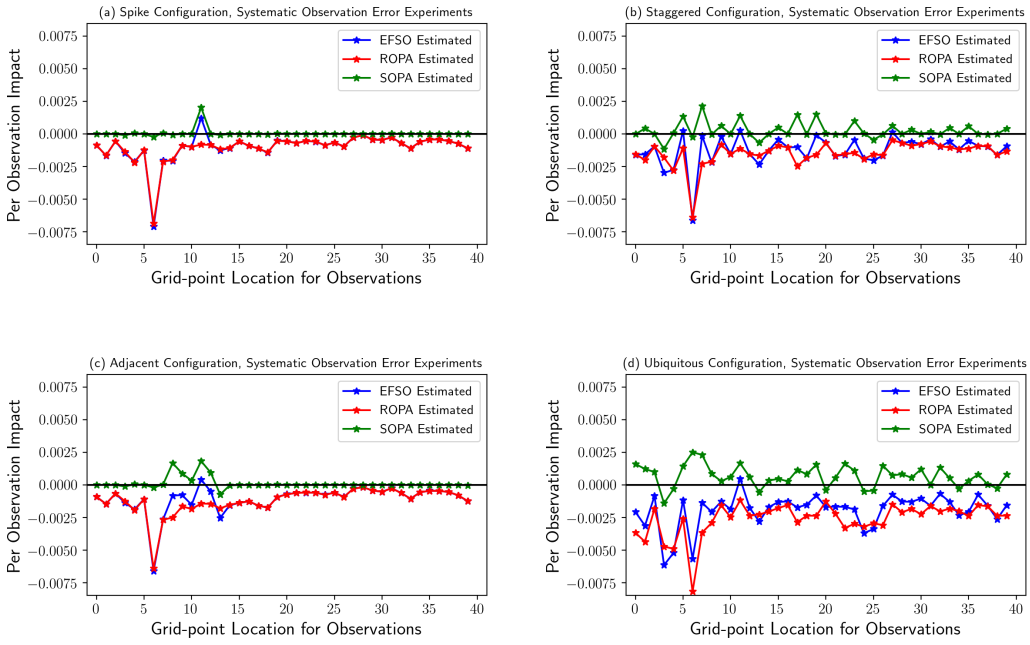

Figure 3.7: EFSO, ROPA and SOPA per observation impact estimates for a forecast lead times of 11 time-steps. Positive quantities indicate per observation detriment and negative quantities indicate per observation benefit. a) Systematic observation errors of 0.05 are introduced in the spike configuration, b) same as (a), but the systematic flaws are introduced in the staggered configuration, c) same as (a), but the systematic flaws are introduced in the adjacent configuration, d) same as (a), but the systematic flaws are introduced in the ubiquitous configuration.

flaws in all experimental configurations, and that the beneficial contributions from ROPA effectively hide these systematic flaws when access to only EFSO quantities are available for the ubiquitous and adjacent configurations. As such, in a manner that is unlike the ROPA calculations from the section 3.3.2 random flaw experiments, "cumulative"

aspects from assimilating observations with systematic flaws in antecedent cycles do not inhibit SOPA from detecting the presence of systematic observation flaws. To further explore this aspect of SOPA, we perform EFSO and EFSO component calculations for the same set of observations as those represented in figure 3.7, but remove "cumulative" aspects by using background from the "ideal baseline" configuration defined in 3.3.2. The result of removing "cumulative" impacts in this manner is plotted in figure 3.8, and it is overwhelmingly the case that SOPA is less detrimental and ROPA is less beneficial when removing "cumulative" contributions. We propose here that for ROPA the "cumulative"

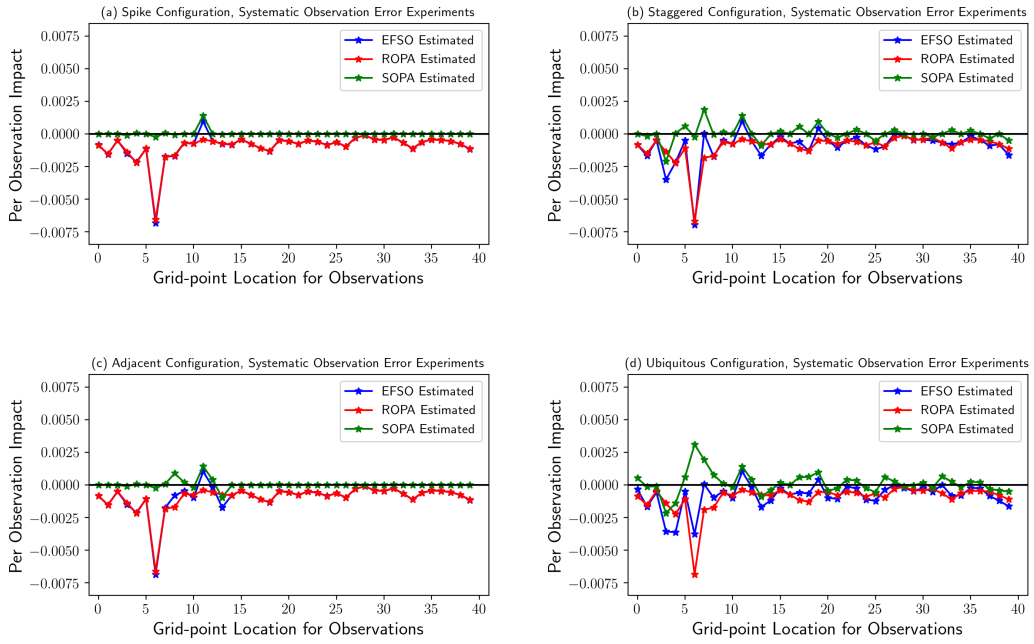

Figure 3.8: Same as figure 3.7 except the "ideal baseline" experiment described in 3.3.2 is used as the background forecasts for all calculations. EFSO, ROPA and SOPA per observation impact estimates for a forecast lead times of 11 time-steps. Positive quantities indicate per observation detriment and negative quantities indicate per observation benefit. a) Systematic observation errors of 0.05 are introduced in the spike configuration, b) same as (a), but the systematic flaws are introduced in the staggered configuration, c) same as (a), but the systematic flaws are introduced in the adjacent configuration, d) same as (a), but the systematic flaws are introduced in the ubiquitous configuration.

contextual aspect operates in a manner similar to that in section 3.3.2 in that directionality is largely determined by the relative magnitudes of "corrective" and "error" contributions in the innovations; whereas for SOPA the "cumulative" aspect arises as the assimilated systematic observation flaws tug on the EFSO employed cross-covariances such that the direction of systematic observation flaws tend to align with the forecast error anomalies.

To test this proposed EFSO components directionality explanation, we performed 2 sets of 7 experiments associated with 7 different ubiquitously assimilated systematic observation errors. For both sets of experiments we perform EFSO and "EFSO-components" calculations for observations where known systematic flaws of -0.05, -0.03, -0.01, 0.00, 0.01, 0.03 and 0.05 have been assimilated ubiquitously.

In figure 3.9a and 3.9b the percent beneficial and per observation impact statistics for EFSO, ROPA and SOPA are plotted for 7 experiments where systematic observation errors indicated on the abscissa contributed to background forecasts through assimilation in antecedent cycles (i.e. "cycling" experiments). We use the percent beneficial statistic as proxy information in support of the aforementioned ROPA and SOPA directionality explanations. The ROPA percent beneficial statistic in figure 3.9a uniformly increases with the magnitude of ubiquitously assimilated systematic observation errors. This can be explained by an increase in random "corrective" innovation contributions (i.e. background error variance) associated with the increase in magnitude of systematic observation errors assimilated. The per observation SOPA impact represents mechanisms by which systematic observation errors assimilated project onto forecast error anomalies and random "corrective" innovation contributions. Given that the observation error standard deviation is held constant across these experiments, the increase in the ROPA

percent beneficial statistic with systematic observation errors assimilated is consistent with the explanation that ROPA directionality reflects the relative magnitudes of "corrective" and "error" contributions to innovation. In contrast, the SOPA percent beneficial statistic plotted in figure 3.9a decreases with the magnitude of systematic observation errors assimilated. The SOPA percent beneficial statistic reflects the extent of tugging on the cross-covariance by assimilation of systematic observation errors in antecedent cycles. As such, the SOPA component reflects an establishment of directions, over successive cycles, in which the EFSO cross-covariances tend to project onto forecast error anomalies. That is, the cross-covariances effectively maintain a memory of systematic observation error directions assimilated in antecedent cycles.

The plots in figures 3.10a and 3.10b correspond to summary results from a set of 7 experiments where the background forecasts applied in all cases are from the "baseline positive systematic" experimental configuration (i.e. "non-cycling" experiments). As expected, the ROPA percent beneficial statistic is invariant across these experiments since the background forecasts (i.e. "corrective" innovation contribution) and standard deviation of observation errors (i.e. "error" innovation contributions) are static across experiments. The SOPA percent beneficial statistic plotted in figure 3.10b for the set of "non-cycling" experimental configurations changes abruptly when the systematic observation error is increased from 0.00 to 0.01. In this case, the background reflects tugging from the assimilated observing system that includes ubiquitous systematic errors of 0.05 (i.e. the "baseline positive systematic" configuration), such that a systematic direction in which the EFSO cross-covariances tend to project onto forecast error anomalies has been established. As such, when observations assimilated have systematic

observation errors that point in this established direction, the percent beneficial statistic is consistently less than 50% and notably invariant. This again demonstrates that SOPA directionality is explained by established directions that project onto forecast error anomalies over successive antecedent cycles.

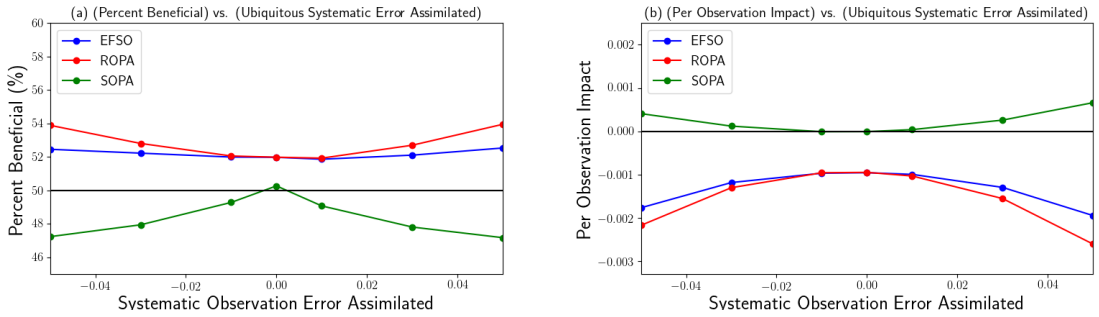

Figure 3.9: Percent beneficial and per observation impact estimate summary statistics for 7 different cycling experiments where systematic observation errors indicated on the abscissa have been assimilated ubiquitously. a) EFSO, ROPA and SOPA percent beneficial statistic for 7 "cycling" experiments according to ubiquitous assimilation of systematic errors indicated on the abscissa. b) Same as (a), but per observation impact.

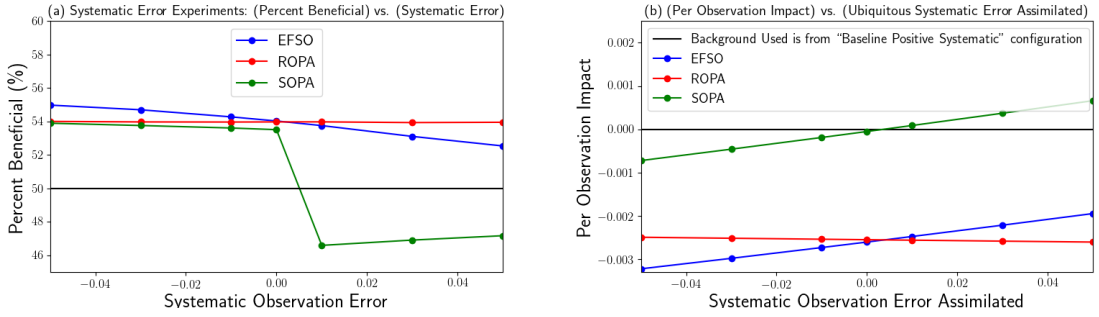

Figure 3.10: Percent beneficial and per observation impact estimate summary statistics for 7 different "non-Cycling" experiments where systematic observation errors indicated on the abscissa have been assimilated ubiquitously. The background used in all cases is from the "Baseline Positive Systematic" experimental configuration. a) EFSO, ROPA and SOPA percent beneficial statistic for 7 "cycling" experiments according to ubiquitous assimilation of systematic errors indicated on the abscissa. b) Same as (a), but per observation impact.

The key results from the explanatory systematic observation error experiments in

this subsection are as follows:

- The SOPA component reflects an establishment of directions, over successive cycles, in which the EFSO cross-covariances tend to project onto forecast error anomalies
- For a variety of configurations tested the SOPA component consistently identifies systematic observation errors assimilated, and ROPA effectively hides this information when access to only EFSO quantities is available
- For the Lorenz '96 system, assimilated systematic observation errors overwhelmingly project onto forecast error anomalies, and SOPA represents the mechanism by which this occurs
- The Lorenz '96 experiments here demonstrate general advantages in using "EFSO-components" to detect observation flaws and errors by distributing "corrective" innovation contributions
- Other factors being equal, systematic observation errors assimilated will serve to increase ROPA suggested benefit for the Lorenz '96 EFSO OSSE testing apparatus

3.3.4 "0-hour EFSO" and "0-hour EFSO-components" Assessments

Application of the the EFSO methodology at the "0-hour" verification time are plotted in figures 3.11 and 3.12. Although the results plotted in figures 3.11 and 3.12 represent instantiations of analysis impact estimation, we hereafter refer to such calculations as "0-hour EFSO" and "0-hour EFSO-components" on the basis that we are

using the general EFSO methodology at the "0-hour" verification time. The purpose of performing these 0-hour calculations is to assess the extent to which "EFSO-components" may provide additional information in identifying observation flaws and errors that is unavailable from analysis impact approaches.

Figure plot 3.11b corresponds to "0-hour EFSO" calculations where analyses are used as verification and the flawed observations indicated in the plot legend contributed to background forecasts through assimilation in antecedent cycles (i.e. "cycling" experiments), and figure plot 3.11d corresponds to "0-hour" calculations where analyses are used as verification and the background forecasts applied were from the "ideal baseline" experiment (i.e. "non-cycling" experiments). Figures 3.11b and 3.11d indicate that regardless of cumulative contributions to the background the "0-hr EFSO" calculations do not provide substantive information regarding the ubiquitous flaws in the observations when analyses are used as verification. However, when the truth is used as verification, substantive information pertaining to the flaws in observations was available only when the "ideal baseline" experiment was used for the background forecasts (i.e. there was no cumulative impact on the background). That is, regardless of the verification used, cumulative impacts on the background forecasts conceal information pertaining to the observation flaws assimilated.

For the "baseline positive systematic" configuration described in 3.3.3, "0-hour EFSO", "0-hour ROPA" and "0-hour SOPA" calculations were performed using both the "truth" and self analyses as verification. The aforementioned "0-hour EFSO" and "0-hour EFSO-components" calculations are plotted in figure 3.12. Figures 3.12a and 3.12b indicate that using the analysis as verification for "0-hour" calculations conceals SOPA

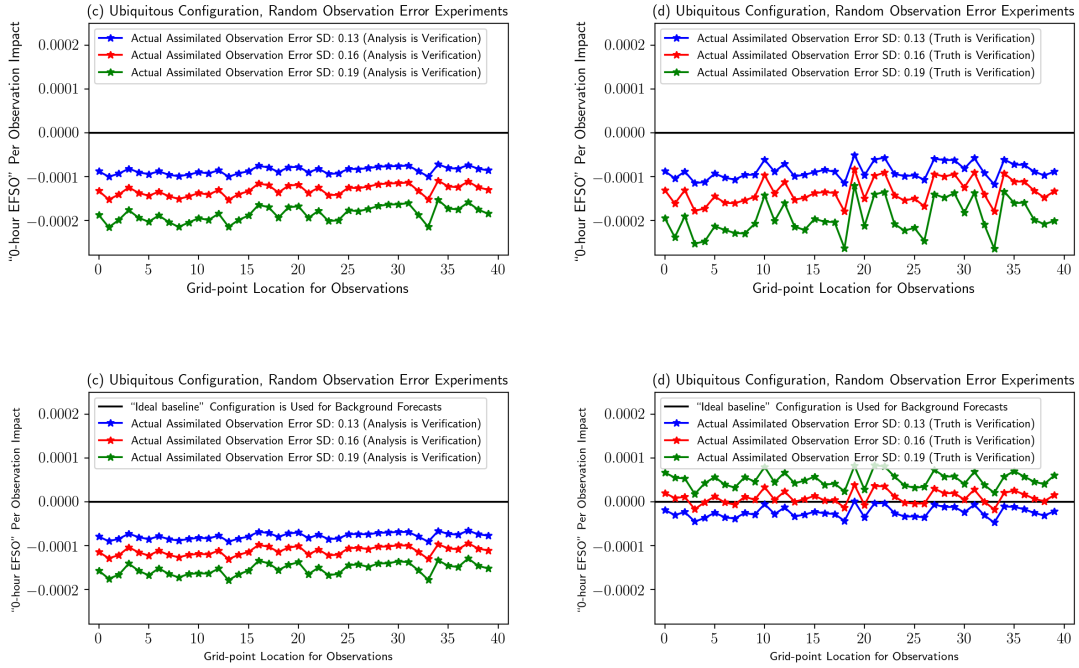

Figure 3.11: "0-hour EFSO" per observation impact for experiments where the ubiquitously assimilated standard deviation of observation error is indicated in the legend. a) "0-hour EFSO" per observation impact where the "analysis" is used as verification, and the observation error standard deviation indicated in the legend contributes to backround forecasts through assimilation in antecedent cycles. b) same as (a) except the "truth" is used as verification. c) "0-hour EFSO" per observation impact where the "analysis" is used as verification, and background forecasts are from the "ideal baseline" experiment described in section 3.3.2. d) Same as (c) except the "truth" is used as verification.

information pertaining to systematic observation errors assimilated that is otherwise available. That is, consistent with results from 3.3.3, "cumulative" considerations do not inhibit identification of systematic observation errors assimilated. Note that SOPA was able to identify these same systematic errors for lead times of 11 time-steps according to figure 3.7d when self analyses were used as verification. This demonstrates that SOPA extracts useful independent information that is unavailable when applying analysis impact estimation (e.g. "0-hour EFSO" and "0-hour EFSO-components").

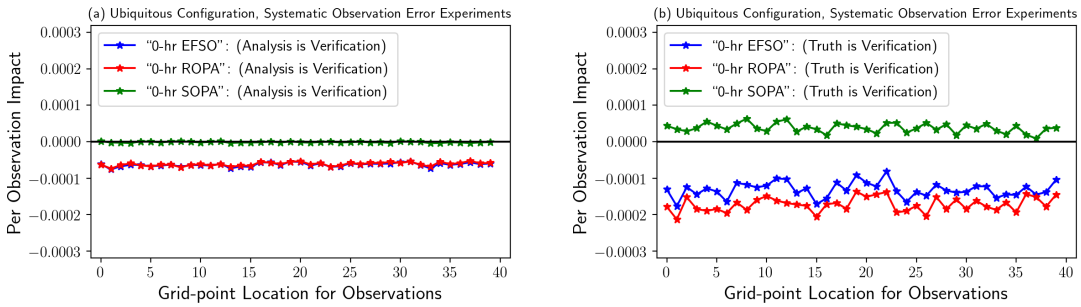

Figure 3.12: "0-hour EFSO", "0-hour ROPA" and "0-hour SOPA" per observation impact for the "Positive Baseline Experimental" OSSE configuration where systematic observation errors of 0.05 are assimilated ubiquitously. a) Analysis is used as verificaion, b) Same as (a), but truth is used as verification.

Chapter 4: Application of "EFSO-Components" in Proactive Quality Control

4.1 Overview

In this chapter the Lorenz '96 observing system simulation experiment (OSSE) testing apparatus described in chapter 3 is used again to explore application of "EFSO-components" as a basis for data rejection within proactive quality control (PQC) (Chen and Kalnay 2020,Chen and Kalnay 2019, Ota et al. 2013, Hotta et al. 2017a). An overview of observation quality control (QC) strategies applied in operational NWP systems and perspective regarding the intended purpose of applying the more recently proposed EFSO-based PQC approach is discussed in the first portion of this chapter. The expected fundamental advantage of "EFSO-components" in extracting more predictive information in the context of PQC observation rejection decisions is then explained. In the results section of this chapter, the effectiveness of "EFSO-components" based PQC rejection strategies in reducing analysis and forecast RMSE is compared to EFSO based PQC using Lorenz '96 OSSEs with the intent of exploring and confirming the expected fundamental PQC application advantages of "EFSO-components".

4.2 Observation Quality Control in Operational NWP Systems

Assimilation of observations with gross errors can have a substantially harmful impact on numerical weather forecasts, (Alpert J. C. et al. 2009), and the errors introduced by assimilating such observations can persist by propagation through background forecasts across successive analysis-forecast cycles (Hotta 2014). The process of removing observations with gross errors is known as observation quality control (QC). The overarching approach of traditional observation QC procedures applied in operational NWP systems is that observations are rejected or retained according to the magnitude of their differences from "expected values" (Kalnay 2002 section 5.8), with the idea that large residuals between observations and "expected values" are an indication of gross errors in the observations. In this context, the "expected values" often applied in operational NWP systems include climatology (DiMego et al. 1985), other nearby observations (Dee et al. 2001), background forecasts (Atkins 1984), optimal interpolation (OI) based analyses (Lönnberg P and Shaw DB 1984,Woollen 1991) and physical consistency (e.g. vertical observations are consistent with hydrostatic balance) (Collins and Gandin 1990), among others. The process of performing these consistency checks simultaneously was introduced in Gandin 1988. A limitation of these traditional QC approaches is their dependence on the "expected values" being a credible source of information for assessing the quality of observations in any given situation. For example, in observation sparse regions, observations that have large departures from both background forecasts and OI analyses will tend to be rejected with these approaches. Where these large departures have more to do with errors in background forecasts or OI

analyses useful observations can be rejected according to traditional QC procedures, and this can serve to further increase errors in background forecasts over successive cycles.

For the Japanese Meteorological Agency (JMA) NWP system, (Onogi 1998 as cited in Hotta 2014) introduced a dynamical QC approach to partially alleviate this problem by using horizontal gradients in background forecasts (i.e. the first guess) as a basis for specifying observation rejection thresholds. However, this approach does not attempt to obtain or use direct information as to the absolute or relative quality of observations, and as such, ultimately depends on statistical assumptions regarding innovations and background forecasts as a basis for assessing the quality of observations in any given situation. In this case, more lenient departure thresholds associated with larger horizontal gradients in background forecast fields could serve to retain observations with large gross errors where forecast uncertainty may tend to be higher than usual.

Where there is an absence of direct information as to the absolute or relative "effective quality" of an observation, Variational quality control (VarQC) (Anderson and Järvinen 1999, Lorenc and Hammon 1988, Purser 1984) assigns weights to observations according to their departure from preliminary analysis states within iterations of the variational minimization process. This allows observations to provide contributions to the analysis in circumstances when large deviations to "expected values" may have more to do with the quality of the background forecasts or preliminary analyses (etc.) than gross errors in the observations.

4.3 Proactive Quality Control

The essence of the PQC approach, (Chen and Kalnay 2019, 2020, Hotta et al. 2017a, Ota et al. 2013), is that EFSO provides a direct situation dependent basis for rejecting observations according to comparisons between background forecast errors (i.e. "corrective" information) and errors in individual innovations. In this context, tangent-linear approximations to forecast models enable estimation of forecast error changes associated with assimilation of individual observations, and as such, provide a basis for making decisions as to whether observations (innovations) should be rejected or assimilated. The efficacy of this approach, in part, depends on the accuracy of the tangent-linear approximations to the forecast model in any situation, and the extent to which the choice of verification is something close to the "truth" relative to the forecasts employed in the observation impact estimation. As such, in PQC applications, we take the approach that direct information relating to the relative quality of observation sets and background forecasts can be determined according to whether estimated impacts for observation sets tend to be detrimental (i.e. the observation is estimated to increase forecast error) or beneficial (i.e. the observation is estimated to decrease forecast error) (Chen and Kalnay 2019, Hotta 2014) according to EFSO. That is, as discussed in chapter 1, we expect observation sets or observation types will tend to be detrimental when the "error" contribution to innovation is large compared to the "corrective" contribution to innovation.

For a hypothetical 6-hour data assimilation window the cycling PQC methodology consists of the following steps outlined in Chen 2018 and represented in figure 4.1:

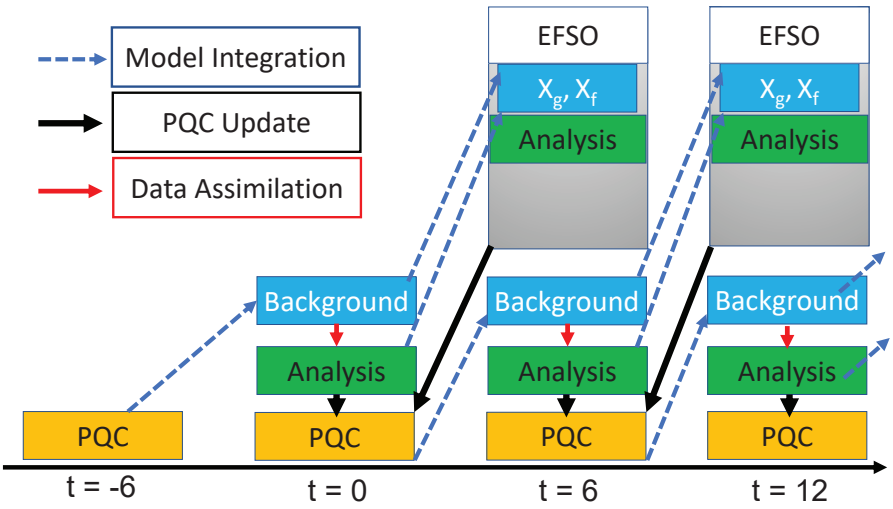

Figure 4.1: Flowchart of cycling proactive quality control (PQC) (adapted from Hotta 2014 and Chen 2018).

Paraphrase of steps outlined in Chen 2018

- Perform cycled DA from t = -6 to t = 6 to obtain a verifying analysis at t = 6.
- Obtain 12-hour and 6-hour forecasts from the original analyses at t = -6 and t = 0
- Perform 6-hour EFSO calculations
- Identify observations to be rejected based on the EFSO calculations
- Update the analysis at t = 0 with the intent to remove or significantly reduce contributions from the rejected observations

The general PQC step includes application of a strategy for using EFSO in identifying harmful observations to be rejected, and a methodology for removing

contributions from the original analyses associated with assimilating the identified harmful observations. In (Chen and Kalnay 2019, figure 7) a schematic in one-dimensional model space was used to explain limitations in the aggressiveness with which observations can be rejected based on EFSO information to reduce forecast error in general. According to this schematic, Chen and Kalnay 2019 explained that only the fast growing detrimental observations should be rejected, so as to avoid error growth in the "original beneficial" direction. As such, we take the approach that the best rejection strategies will be able to comprehensively identify, and aggressively reject, truly detrimental observations that have no compensating benefits.

4.4 "EFSO-Components" Based PQC

We propose here application of "EFSO-components" as a basis for rejecting harmful observations within cycling PQC. As such, the PQC flowchart in section 4.3 requires one additional step to compute updated systematic contributions to innovation ($\delta \mathbf{y}_s$) and forecast error quantities ($\mathbf{e}^{f+g}_{[t,s]}$) as represented in figure diagram 4.2. The EFSO component calculations can then be performed according to equation 2.11. For all "EFSO-components" based PQC experiments presented in this dissertation, a running mean approach is applied to update the systematic innovation and systematic forecast error quantities that are needed for the component computations.

To explain the problem space we intend to address using "EFSO-components" based PQC, consider the relative position maps in figure 4.3, which represent a hypothetical scenario where there are three EFSO suggested detrimental observations

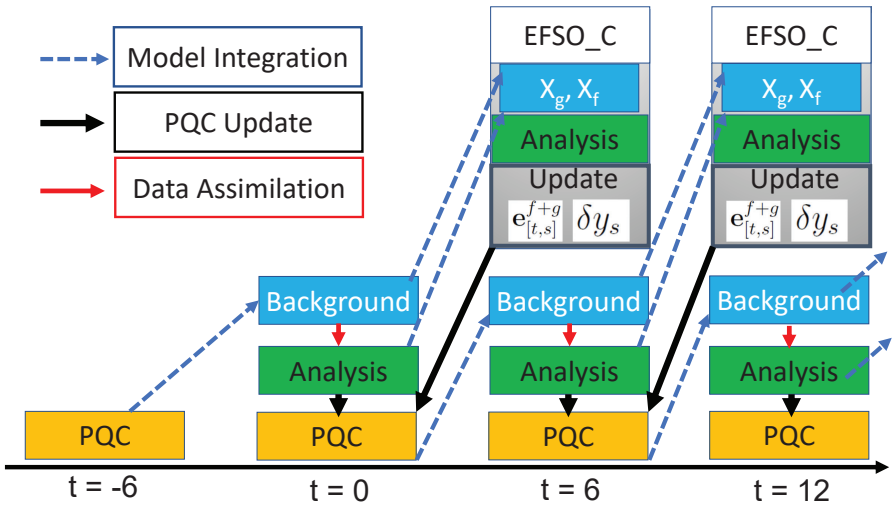

Figure 4.2: Flowchart of cycling proactive quality control (PQC) (adapted from Hotta 2014 and Chen 2018) with an additional step to compute updated systematic contributions to innovation and EFSO forecast error quantities. This additional step facilitates application of "EFSO-components" in PQC.

to be considered for rejection. In practice, the EFSO calculation available for making observation rejection decisions is directly applicable to only the initial configuration in which no observations have been rejected (i.e. figure 4.3a). As the EFSO suggested detrimental observations are removed from the system (i.e. rejected) in figures 4.3b through 4.3d with increasing aggressiveness, the errors in the hypothetical forecast from analysis (i.e. $\mathbf{e}_t^f$) change in an expected manner. After rejecting two EFSO suggested detrimental observations, figure 4.3c, the forecast error sign changes, and further rejecting EFSO suggested detrimental observations will only serve to increase the magnitude of forecast error. Note the hypothetical figure 4.3 sequence represents an approach in explaining potential application limitations of EFSO associated with the "simultaneous" aspect that was discussed briefly in chapter 3.

The intent of applying "EFSO-components" is that aggressive rejection of observations with no compensating benefit (i.e. all substantive EFSO component contributions are detrimental) will allow us to address the original forecast errors (i.e. those indicated in 4.3(a)) while mitigating the extent to which forecast error growth in the original (i.e. prior to rejecting observations) "EFSO beneficial direction" is introduced (e.g. 4.3d). To explain why we expect "EFSO-components" will be helpful in this regard, it is helpful to consider as explained in chapter 3, that the EFSO component terms operate differently in providing information regarding the extent of flaws and errors in observations. For example, the SOPA component advantages in detecting systematic observation flaws relate to the establishment of directions, over successive cycles, in which the EFSO cross-covariances tend to project onto forecast error anomalies (section 3.3.3); whereas ROPA represents an immediate comparsion between random "corrective"

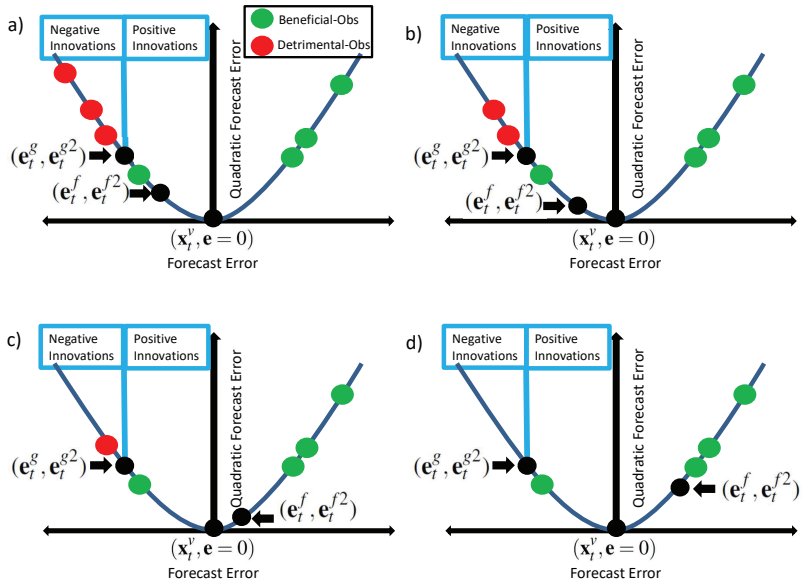

Figure 4.3: Hypothetical sequence, from figure (a) to figure (d), in which increasing observation rejection aggressiveness eventually leads to error growth in the "original beneficial direction". The intent is that "EFSO-components" based observation rejection will be more effective in removing errors in the "original detrimental direction" without introducing error in the "original beneficial direction".

and "error" innovation contributions at the cycle time (section 3.3.2 and 3.3.3), and as such, does not provide a means to account for flaws and errors in observations that have accumulated over successive cycles (i.e. the errors are random so there is no "memory" retained from one cycle to the next). For the Lorenz '96 system this contrast in the component mechanisms relates in part to the fact that both the random and systematic observation errors primarily project onto forecast error anomalies. As such, we expect that the combined directionality (i.e. beneficial or detrimental) information from the ROPA and SOPA components will be helpful in addressing the nature of errors and flaws in observations that relate to both the immediate cycle and errors that have accumulated over successive cycles. That is, for PQC applications, we expect the rejection of observations with no associated substantive beneficial component contributions will have a tendency to be more effective in addressing the original forecast errors (i.e. 4.3(a)) without introducing new errors in what was the "original beneficial direction", relative to rejection of observations that have at least one substantive beneficial component contribution. Furthermore, we expect that "EFSO-components" will allow us to directly address the component dumping mechanisms discussed in chapter 3 section 3.3.3, and this will indirectly help to mitigate the hiding of random error flaws associated with "cumulative" aspects in the primary ROPA EFSO component.

Rejection decisions based on the overall EFSO calculation in contrast will tend to depend on the magnitude of the primary component (e.g. ROPA for the Lorenz '96 system) in any particular circumstance. While we expect this information may be less efficient at addressing the nature of flaws and errors in observations over successive cycles, it may provide complementary information in addressing some of the larger

"configuration dependent" (e.g. the specific Kalman gain matrix applied) errors in EFSO employed forecasts. In addition to exploring the proposed advantage of "EFSO-components" based rejection we will investigate this potential complementary aspect of EFSO and "EFSO-components" in the next results subsection.

4.5 "EFSO-Components" Based PQC Experiments

4.5.1 "EFSO-Components" Based Observation Rejection Strategies

As discussed in chapter 3 we expect that application advantages of "EFSO-components" arise when more than one component term contributes substantively. For the Lorenz '96 OSSE testing apparatus this occurs when systematic errors are introduced. As such, we use a baseline experimental configuration for comparing EFSO based PQC and "EFSO-components" based PQC where systematic observation errors approximately an order of magnitude smaller than the standard deviation of random observation errors have been added ubiquitously, [$\mathcal{N}(\mu = 0.0125, \sigma = 0.1125)$], and model forcing error on the order of what was tested in Chen and Kalnay 2019, [$F = 8.00075$]. For this experimental observation configuration we test different "EFSO-components" based conditionals for rejecting observations within PQC. The intent in testing different "EFSO-components" conditionals is to add support in confirming or rejecting the hypothesized advantages of using "EFSO-components" as outlined and explained in section 4.4.

To assess the implications of "EFSO-components" conditionals as a basis for rejecting observations in PQC, scatter plots of 6-time step SOPA versus 6 time-step ROPA estimated impact calculations are provided in figure 4.4. ROPA and SOPA are

overwhelmingly the primary component contributors for this Lorenz '96 testing system, and have commonality in that they both represent component mechanisms associated with changes in forecast error anomalies. The EFSO and "EFSO-components" conditionals examined in figure 4.4 have corresponding symbol colors as follows:

- Observations where ROPA exceeds its 70th percentile, SOPA exceeds its 70th percentile, and EFSO does not exceed its 90th percentile are plotted in blue
- Observations where ROPA exceeds its 70th percentile, SOPA exceeds its 70th percentile, and EFSO exceeds its 90th percentile are plotted in yellow
- Observations where ROPA does not exceed its 70th percentile or SOPA does not exceed its 70th percentile, and EFSO exceeds its 90th percentile are plotted in red
- The complement of events indicated by blue, red and yellow symbols are plotted in cyan

We will hereafter refer to events where ROPA and SOPA exceed their respective thresholds as RS based rejection. RS based rejection is far more aggressive in rejecting quadrant-1 observations (i.e. where both ROPA and SOPA are positive) than EFSO based rejection, and it is overwhelmingly the case that ROPA is driving EFSO based rejection as would be expected given its contributions are approximately an order of magnitude larger than all other component terms. As outlined in section 4.4 our expectation is that the ability to identify and more aggressively reject quadrant-1 observations (i.e. observations with no substantive beneficial contributions) according to "EFSO-components" allows us to better address "original detrimental" growing errors while avoiding new error growth

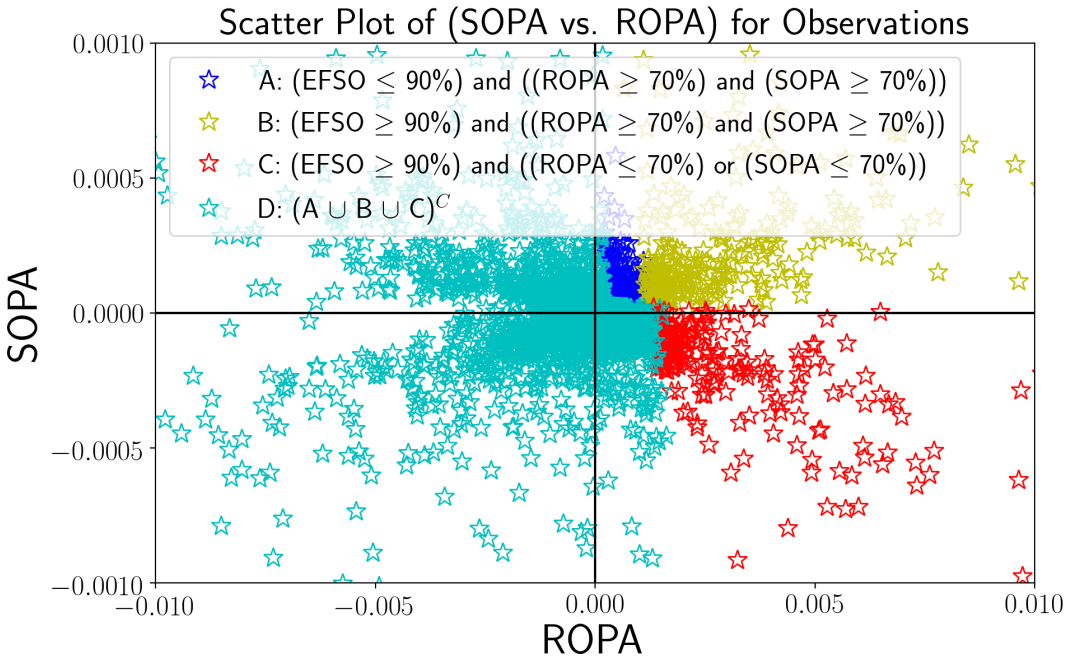

Figure 4.4: Scatter plot examining where observations satisfy "EFSO-components" and EFSO conditionals. Observations where ROPA exceeds its 70th percentile, SOPA exceeds its 70th percentile and EFSO exceeds its 90th percentile are indicated by blue symbols. Observations where ROPA exceeds its 70th percentile, SOPA exceeds its 70th percentile, and EFSO exceeds its 90th percentile are indicated by yellow symbols. Observations where ROPA does not exceed its 70th percentile or SOPA does not exceed its 70th percentile, and EFSO exceeds its 90th percentile are indicated by red symbols. The complement of events indicated by blue, yellow and red symbols are indicated with cyan symbols.

in the "original beneficial" direction. The observations rejected by EFSO at the 90th percentile, but not rejected according to RS for the particular thresholds relevant to figure 4.4 are overwhelmingly in quadrant-4 (i.e. where ROPA is detrimental and SOPA is beneficial). In considering conditionals where "EFSO-components" are applied, we would like to avoid inhibiting rejection of quadrant-4 outliers that correspond to some of the larger overall contributions to forecast error according to EFSO. We expect that the "simultaneous" aspect will be less of a consideration when rejecting these EFSO outliers, because rejection at these more stringent (i.e. higher) thresholds will tend to be limited to a smaller set of observations. As such, in an effort to explore the advantages of "EFSO-components" in PQC we apply conditionals that reject observations according to the union of EFSO exceeding a higher (e.g. 98%, 90% etc.) threshold and RS exceeding a lower (i.e. more aggressive) threshold for rejection. This approach allows us to take advantage of both the directional information from "EFSO-components" and the magnitude information that is better accounted for by using EFSO directly as discussed in section 4.4. We will hereafter refer to this **E**FSO **o**r (RS) conditional rejection approach as (EoRS). Rejection according to the union of events indicated by blue, yellow and red symbols in figure 4.4 is an instantiation of EoRS.

In running experiments to compare application of the EoRS conditional in PQC with pure-EFSO based PQC, the approach outlined in Chen and Kalnay 2019 is employed such that all rejection threshold percentiles used are based on 5000 cycles from a control experiment where no observations were rejected. This means that there will be variability in the number of observations rejected from cycle to cycle, consistent with what would generally be expected in practice. The rejection percentiles for EoRS, and

the corresponding percentage of observations rejected for 6 and 21 time-step EFSO lead times are listed in table 4.1. Note that "EFSO_C" is used in the table when referring to application of the overall EFSO calculation within the context of EoRS. To avoid ambiguity we will hereafter refer to PQC rejection based solely on the overall EFSO calculation as "pure-EFSO based" rejection. The corresponding "pure-EFSO" control experiment rejection percentiles are listed in table 4.2. For this study, we use the PQC_K and PQC_H methodologies (Hotta et al. 2017a,Chen and Kalnay 2019) for removing the rejected observation contributions from the analysis. The PQC_K methodology approximates removal of rejected observation analysis contributions by using the same exact Kalman gain and setting the innovations of the identified observations for rejection to zero; whereas the PQC_H approach removes rejected observation contributions from the columns of the observation operator $\mathbf{H}$ which involves an update to the Kalman gain matrix applied.

Table 4.1: A list of EoRS rejection percentile thresholds applied, and the corresponding percentage of observations rejected.

EoRS Rejection Percentile Thresholds Table				
EFSO_C	ROPA	SOPA	6-time step obs Rejected (%)	21-time step obs Rejected (%)
98	80	80	8.2	8.7
90	70	70	16.4	17.0
85	60	60	24.2	24.7
80	50	50	33.5	33.7
75	40	40	49.6	49.8
70	30	30	63.9	64.7
65	20	20	76.0	76.6

Cycling PQC is performed for each of the experimental EoRS rejection threshold rows in table 4.1 and for the corresponding "pure-EFSO" (i.e. control) rejection threshold

Table 4.2: A list of pure-EFSO rejection percentile thresholds applied.

Pure-EFSO Rejection Percentile Thresholds Table	
6-time step Pure-EFSO	21-time step pure-EFSO
91.8	91.3
83.6	83.0
75.8	75.3
66.5	66.3
50.4	50.2
36.1	35.3
24.0	23.4

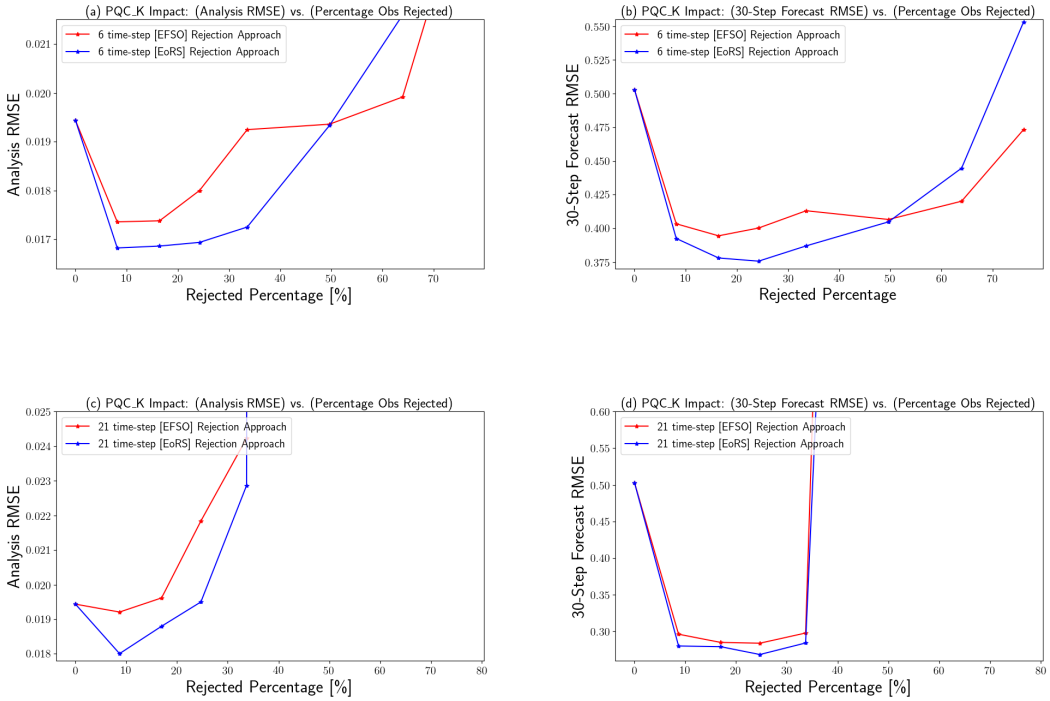

Figure 4.5: PQC_K baseline experiments where model forcing error is introduced in all cases [F=8.00075]. EoRS is plotted with blue symbols and lines, and "pure-EFSO" is plotted with red symbols and lines. As such, EoRS (i.e. application of "EFSO-components" in PQC) is more effective in reducing analysis and forecast RMSE than "pure-EFSO".: a) Analysis RMSE as a function of rejection percentage using 6-time step PQC_K. b) Same as (a), but forecast RMSE as a function of rejection percentage. c) Analysis RMSE as a function of rejection percentage using 21 time-step PQC_K, d) Same as (c), but forecast RMSE as a function of rejection percentage.

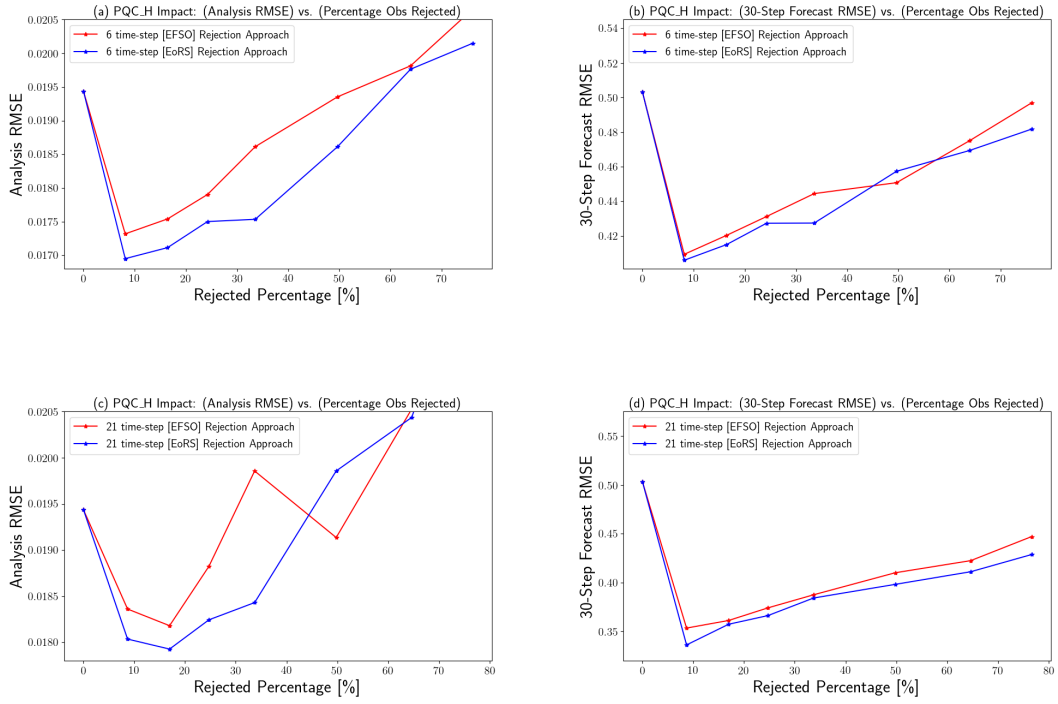

Figure 4.6: PQC_H baseline experiments where model forcing error is introduced in all cases [F=8.00075]. EoRS is plotted with blue symbols and lines, and "pure-EFSO" is plotted with red symbols and lines. As such, EoRS (i.e. application of "EFSO-components" in PQC) is more effective in reducing analysis and forecast RMSE than "pure-EFSO". : a) Analysis RMSE as a function of rejection percentage using 6-time step PQC_H. b) Same as (a), but forecast RMSE as a function of rejection percentage. c) Analysis RMSE as a function of rejection percentage using 21 time-step PQC_H, d) Same as (c), but forecast RMSE as a function of rejection percentage.

rows in table 4.2. The results when applying PQC_K are plotted in figure 4.5 and the results when applying PQC_H are plotted in figure 4.6. Note that the results for "pure-EFSO" based PQC, plotted in red, correspond to experiments where the rejection threshold percentile is consistent with the last two columns in table 4.1 (i.e. 100 - Percentage_Rejected), which is also reflected in the rejection thresholds listed in table 4.2. It is apparent that the EoRS conditional allows us to extract more "predictive" information in deciding which sets of observations should be rejected to reduce analysis and forecast

RMSE beyond what can be achieved using "pure-EFSO" based PQC. Furthermore, the analysis RMSE reduction grows throughout the forecast. According to figures 4.5 and 4.6 it is overwhelmingly the case that this result is irrespective of the percentage of observations rejected, the methodology for removing analysis contributions, and the forecast lead time employed in EFSO and "EFSO-components" calculations. We conclude that using the EoRS conditional as a basis for rejecting observations allows us to address both errors identified according to immediate considerations at a specific cycle time, via ROPA, and errors identified according to considerations that have accumulated over antecedent cycles via SOPA. As such, the advantage of the EoRS conditional is explained by its ability to better target observations, for rejection, according to a more comprehensive account of the nature and extent of flaws in observations, and its advantages relative to pure EFSO-based PQC are robust to the circumstances tested (e.g. the Kalman gain applied, forecast lead-time used for impact estimation, rejection threshold, impact metric etc.).

To further explore our conclusion that it is the combined information from ROPA and SOPA that provides advantages in rejecting observations, and to explore the potential complementary aspects of EFSO and RS, we consider results here from three additional sets of 6-time step PQC_K experiments that include application of EFSO component terms as a basis for rejecting observations: i) the intersection of **R**OPA and **S**OPA (RS) (i.e. R $\cap$ S), ii) the union of **E**FSO and **S**OPA (i.e. EFSO $\cup$ SOPA (EoS)), and iii) the union of **E**FSO and **R**OPA (i.e. EFSO $\cup$ ROPA (EoR)). The rejection thresholds for these experiments are provided in tables 4.3, 4.4, and 4.5 where again each row corresponds to an experiment. For comparison, "pure-EFSO" based PQC experiments are performed

Table 4.3: A list of RS rejection percentile thresholds applied, and the corresponding percentage of observations rejected.

RS Rejection Thresholds Table		
ROPA	SOPA	6-time step obs Rejected (%)
90	90	3.1
80	80	7.2
70	70	12.0
60	60	17.6
50	50	25.0
40	40	38.8
25	25	60.6

Table 4.4: A list of EoS rejection percentile thresholds applied, and the corresponding percentage of observations rejected.

EoS Rejection Thresholds Table		
EFSO	SOPA	6-time step obs Rejected (%)
98	80	10.9
90	70	24.7
85	60	36.8
80	50	49.0
75	40	61.1
70	30	72.9
65	20	84.7

according to rejection thresholds that are consistent with the last row of table 4.3 (i.e. 100 - Percentage_Rejected). Analysis and forecast RMSE changes according to RS, EoS, EoR and pure EFSO-based PQC experiments are plotted in figure 4.7. As expected, the EoS approach, plotted in cyan, which introduces rejection of observations according to SOPA (i.e. the secondary component mechanism in Lorenz '96) without consideration of the primary ROPA component mechanism leads to filter divergence. The EoR experiment results were more similar to "pure-EFSO" based PQC as might be expected given the prominence of ROPA contributions in the overall EFSO calculation, but the magnitude of

Table 4.5: A list of EoR rejection percentile thresholds applied, and the corresponding percentage of observations rejected.

EoR Rejection Thresholds Table		
EFSO	ROPA	6-time step obs Rejected (%)
98	80	10.0
90	70	20.1
85	60	30.1
80	50	40.1
75	40	50.2
70	30	60.2
65	20	75.1

RMSE reduction was less than that realized when applying "pure-EFSO" based PQC. As such, the experimental results indicate that application of ROPA and SOPA components individually is not helpful in providing improvements beyond those from "pure-EFSO" based PQC, further supporting the explanation that it is the combined directionality from ROPA and SOPA that enables advantages in making rejection decisions using EoRS. The substantial differences between the EoR and EoS experiment results can be explained by the prominence of the ROPA component in the Lorenz '96 system. While the RS conditional was successful in improving analysis RMSE relative to "pure-EFSO" based PQC, unlike the EoRS conditional it did not provide advantages in reducing forecast RMSE. This suggests that for the Lorenz '96 system the relative component magnitude information from the overall EFSO calculation provides significant complementary information in addressing errors that tend to be configuration dependent (e.g. the specific Kalman gain used, the errors in other observations assimilated, the particular growing forecast errors at any given cycle time (etc.)).

For the remainder of this chapter the EoRS conditional is employed to further

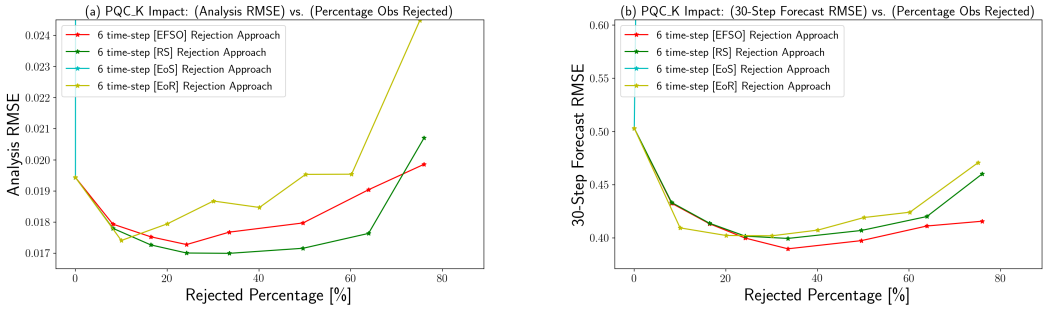

Figure 4.7: PQC_K baseline experiments for "pure-EFSO" (red), EoS (cyan), EoR (yellow) and RS (green) rejection approaches. Model forcing error [F=8.00075] was introduced in all cases. The purpose of this plot and the associated experiments is to demonstrate that the explained advantages of the EoRS rejection approach according to figures 4.5 and 4.6 are twofold: i) EoRS utilizes the combined directionality offered by "EFSO-components" as discussed in section 4.4, ii) EoRS employs complementary aspects of "EFSO-components" and EFSO such that the advantages grow throughout the forecasts. a) Analysis RMSE as a function of rejection percentage using 6-time step PQC. b) Same as (a), but forecast RMSE as a function of rejection percentage.

explore the potential advantages of applying "EFSO-components" within the context of PQC. We will explore these advantages for a variety of configurations, and run tests designed to more directly confirm that "EFSO-components" does allow for better targeting of observations according to their errors and flaws.

4.5.2 EoRS PQC Experiments for Varying Observation Flaws

With the intent of comparing EoRS and "pure-EFSO" based PQC for a more comprehensive set of configurations, we run 6-time step PQC_K and PQC_H experiments based on rejection according to the second row of thresholds in tables 4.1 and 4.2 for 25 combinations of random and systematic observation flaws. The second row of thresholds in tables 4.1 and 4.2 are roughly representative of where "pure-EFSO" and EoRS based PQC tend to provide the most significant improvements in reducing both analysis and

forecast RMSE.

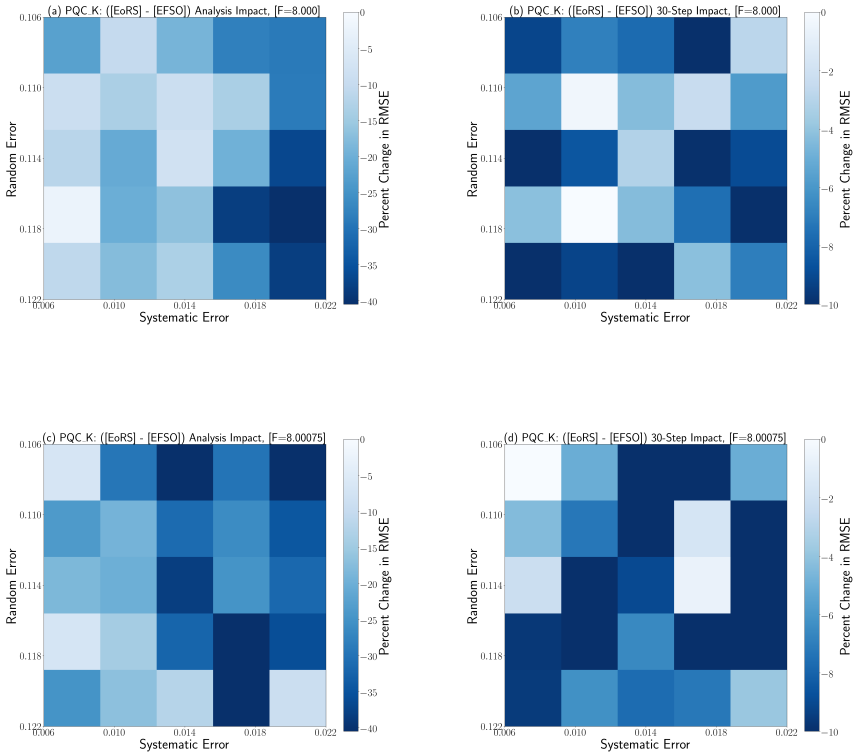

Figure 4.8: 2D percent change impact plots representing differences in RMSE reduction between 6 time-step EoRS PQC_K and 6 time-step "pure-EFSO" based PQC_K for 25 combinations of systematic and random errors. It is overwhelmingly the case that the "EFSO-components" information utilized in EoRS provides advantages in reducing analysis RMSE and 30-step forecast RMSE as indicated by the shades of blue in the vast majority of rectangles. a) Analysis RMSE where no model forcing error is introduced. b) Same as (a), but 30-step forecast RMSE performance metric. c) Analysis RMSE where model forcing error is introduced (i.e. F=8.00075 as opposed to F=8.000). d) Same as (c), but 30-step forecast RMSE performance metric.

The percentage change in RMSE impacts from EoRS compared to "pure-EFSO" based PQC are represented by the 2D plots in figures 4.8 and 4.9 for analysis and 30-step forecast metrics, where each rectangle in the plots corresponds to an experiment comparing EoRS and "pure-EFSO" based PQC rejection decisions. It is overwhelmingly the case that the "EFSO-components" information utilized in EoRS provides advantages

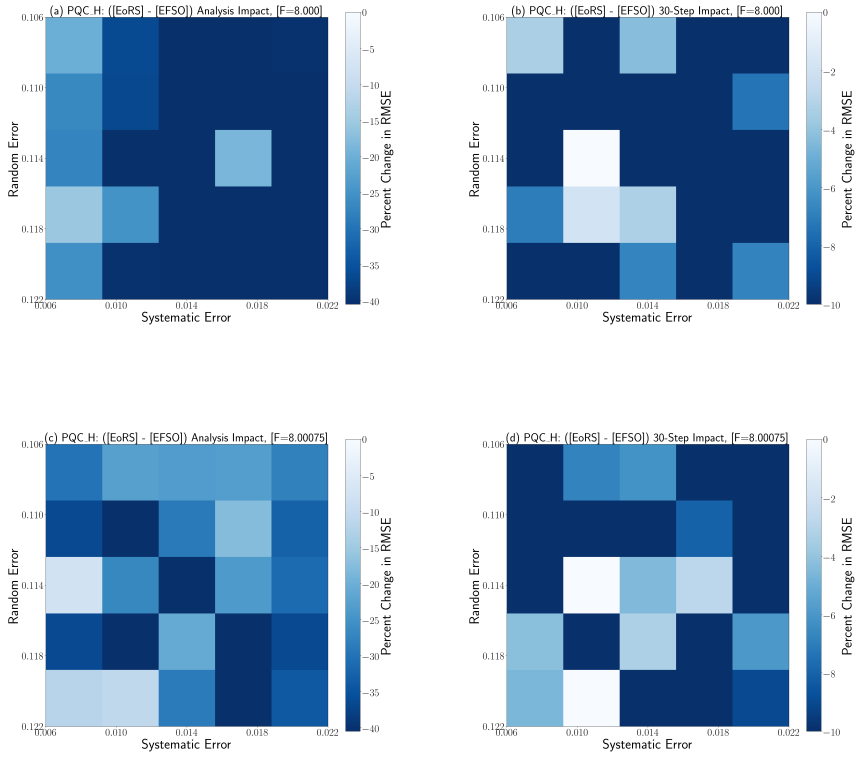

Figure 4.9: 2D percent change impact plots representing differences in RMSE reduction between 6 time-step EoRS PQC_H and 6 time-step "pure-EFSO" based PQC_H for 25 combinations of systematic and random errors. It is overwhelmingly the case that the "EFSO-components" information utilized in EoRS provides advantages in reducing analysis RMSE and 30-step forecast RMSE as indicated by the shades of blue in the vast majority of rectangles. a) Analysis RMSE where no model forcing error is introduced. b) Same as (a), but 30-step forecast RMSE performance metric. c) Analysis RMSE where model forcing error is introduced (i.e. F=8.00075 as opposed to F=8.000). d) Same as (c), but 30-step forecast RMSE performance metric.

in reducing analysis RMSE that are maintained and grow throughout the forecasts as indicated by the shades of blue in the vast majority of rectangles. Note that the absolute errors in these Lorenz '96 system configurations are approximately 30 times larger for the 30 time-step forecasts compared to the analysis time, and as such, the colorbar scales for 30-step forecast RMSE reduction correspond to RMSE reduction that is nearly an order of magnitude larger than that realized at the analysis time. These advantages are robust

to both the PQC_K and PQC_H methodologies and to the introduction of forcing error in the Lorenz '96 model. The analysis RMSE reduction advantages of EoRS relative to "pure-EFSO" based PQC tend to be larger where the ratio of systematic observation error to random observation error is also larger. This result primarily reflects that the ROPA component contributions in the overall EFSO calculation are effectively hiding systematic observation flaws, examples of this were explored in chapter 3 of this dissertation. For the 30 time-step RMSE metric results, the extent of the advantages in using EoRS to better reject observations according to the nature of their errors and flaws are generally not related to the ratio of systematic observation error to random observation error. As such, for the 30 time-step results the improved rejection of EoRS is expressed according to where specifically the growing forecast errors occur in any particular circumstance. This demonstrates the potential application advantage of "EFSO-components" in the context where the intent of applying tangent-linear approximations at shorter forecast lead-times in the EFSO calculation is ultimately to obtain information that improves forecasts at longer lead-times where the tangent-linear approximation may be inadequate. Furthermore, the magnitude of the PQC_H advantages associated with using EoRS are relatively larger than what is realized for the PQC_K experiments, indicating that the advantages of better rejecting observations according to the extent of their flaws and errors is invariant to the Kalman gain used, as would be expected.

4.5.3 PQC Observation Targeting Comparison Experiments

To more directly confirm that ROPA and SOPA information in the EoRS conditional enables improved rejection of observations according to the extent of errors and flaws beyond what can be achieved using "pure-EFSO" based PQC, staggered observation flaws experiments are again employed. For all 6 time-step PQC_K experiments here, the observation flaws are only introduced at odd grid-points in the Lorenz '96 system, and the prescribed standard deviation of observation error is again 0.1 everywhere. The random and systematic observation errors introduced for the relevant staggered experiments are provided in table 4.6.

Table 4.6: The observation flaws placed at odd grid-points in the Lorenz '96 system for six PQC_K observation rejection targeting experiments.

Observation Flaws at Odd Grid-Points		
	Systematic Observation Errors	Std. of Random Observation Errors
Exp-1	-0.022	0.115
Exp-2	-0.022	0.122
Exp-3	-0.015	0.122
Exp-4	-0.010	0.122
Exp-5	-0.005	0.122
Exp-6	0.000	0.122

Since in this case the location of the actual observation flaws at the odd grid-points is known, the extent to which the EoRS or "pure-EFSO" based approaches preferentially reject observations at these flawed observation locations is a direct measure of how well the approaches reject observations according to their flaws. For enumerated experiments 1 through 6 listed in table 4.6, the ratio of systematic observation error to random observation error decreases monotonically. The vertical grid lines are included to help

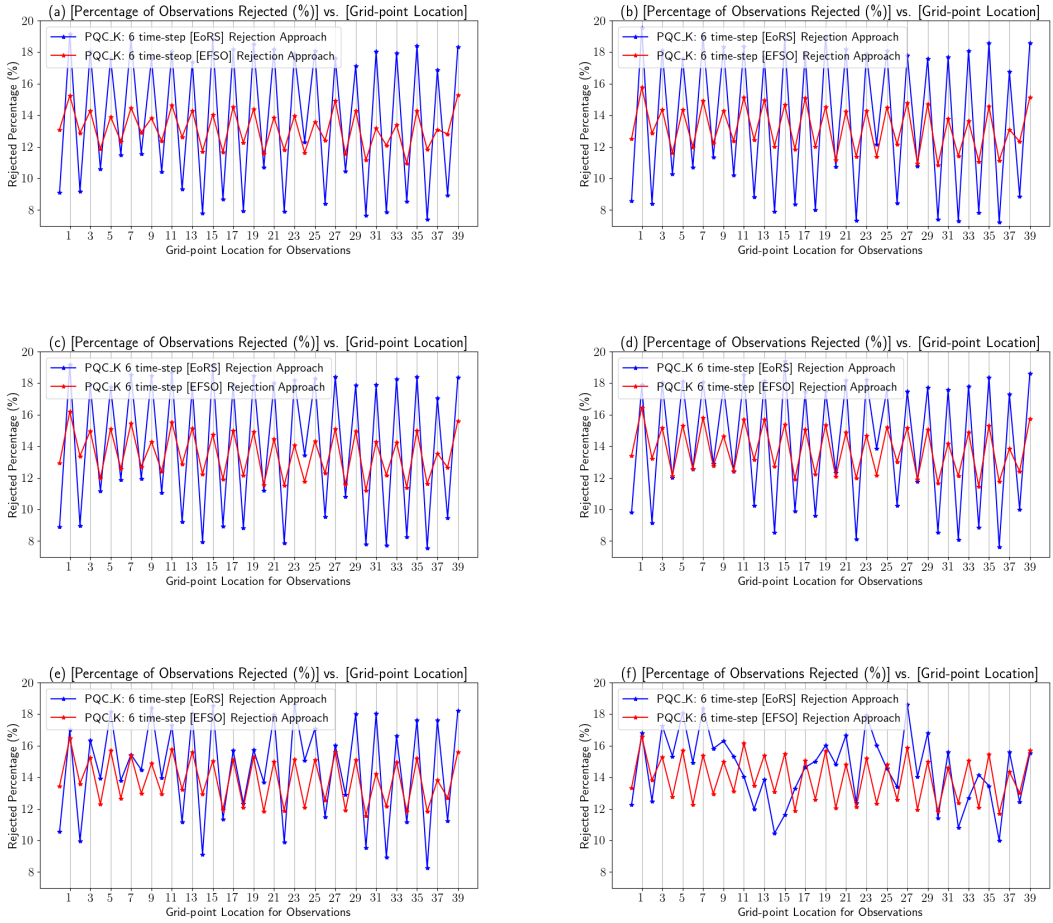

Figure 4.10: Experiments to assess the extent to which EoRS based rejection preferentially targets observations at odd grid-points in the Lorenz '96 system, where observation flaws were introduced, compared to "pure-EFSO" based rejection. Plots a through f correspond to experimental observation flaws configurations 1 through 6 (i.e. exp-1 through exp-6) as listed in table 4.6.

identify the percentage of observations rejected at the odd grid-points in figures 4.10, a through f. For experiments 1 through 5 plotted in figures 4.10, a through e, it is overwhelmingly the case that the EoRS approach preferentially rejects observations at the odd grid-points more so than the "pure-EFSO" based rejection. This comparative result is irrespective of whether the systematic observation flaws are large or small relative to the random error flaws. These results in part relate to the hiding of systematic

observation flaws in the overall EFSO calculation as was discussed in chapter 3 section 3.3.3 of this dissertation. For the Lorenz '96 system the systematic observation errors are effectively transferred (or dumped) into forecast error anomaly contributions via the SOPA mechanism. Over successive cycles this serves to increase ROPA contributions all things being equal, and as such, the systematic observation flaws are effectively hidden when access to "EFSO-components" is unavailable. For the experiments represented in figures 4.10d and 4.10e where the systematic observation errors and SOPA component are relatively small, the differences in application of the directional information from the combined use of ROPA and SOPA, as compared to what is primarily relative component magnitude information from the overall EFSO calculation, explains the improved EoRS based rejection. For the special case experiment (i.e. figure 4.10f) where there are only random observation error flaws, and as such, only ROPA contributes substantively in the Lorenz '96 system, the extent to which "pure-EFSO" based rejection targets observations at the odd grid-points is similar to that for EoRS based rejection. We do not anticipate the special case where only one component contributes substantively to forecast error changes to be a realistic circumstance in an operational setting, as it would require an absence of systematic observation errors among other considerations.

4.5.4 Testing Application of "EFSO-components" for non-Gaussian Observation Error Distributions

Operational NWP systems, effectively or otherwise, assimilate observation types or observing systems that have associated non-Gaussian error distributions (e.g. Kazumori

2014). In general, from a PQC application standpoint we expect that the efficacy of approaches in identifying and preferentially rejecting observations with unusually large errors (i.e. the outliers) will have added significance where non-Gaussian associated "heavy-tails" or "long-tails" are prominent in observation error probability distributions. As such, we compare the efficacy of EoRS based PQC and "pure-EFSO" based PQC_K in reducing analysis and forecast RMSE where non-Gaussian observation error distributions have been introduced. The parametric non-Gaussian distributions we draw from and the associated statistical properties of the introduced observation errors for the EoRS and "pure-EFSO" based PQC_K experiments are provided in table 4.7. The Gumbel and logistic distributions characterized have been scaled and shifted so that the first and second moments are similar to those from the baseline experiment in section 4.4.1.

Table 4.7: The statistical properties of non-Gaussian observation errors introduced

Statistical Properties of non-Gaussian Distributions Tested				
	Mean	Standard Deviation	Skew	Kurtosis
Gumbel	0.0123	0.1121	1.130	2.313
Skewed	0.0355	0.0934	0.0275	0.0359
Logistic-1	0.0121	0.1123	-0.0018	1.1590
Logistic-2	0.0203	0.1123	-0.0018	1.1590

The rejection percentile thresholds applied for EoRS again follow table 4.1, and for the purpose of performing meaningful comparisons, the "pure-EFSO" based rejection percentile thresholds are again set to be consistent with the percentage of observations rejected by EoRS (i.e. analagous to table 4.2). Changes in analysis RMSE and 30-step forecast RMSE according to EoRS and "pure-EFSO" based observation rejection are plotted in figures 4.11 and 4.12 for the Gumbel, Skewed, Logistic-1 and Logistic-2

distributions characterized in table 4.7. Overall the EoRS rejection approach is more effective in reducing forecast and analysis RMSE than pure EFSO-based rejection, indicating that the advantages of using "EFSO-components" extend to circumstances where non-Gaussian observation errors are present.

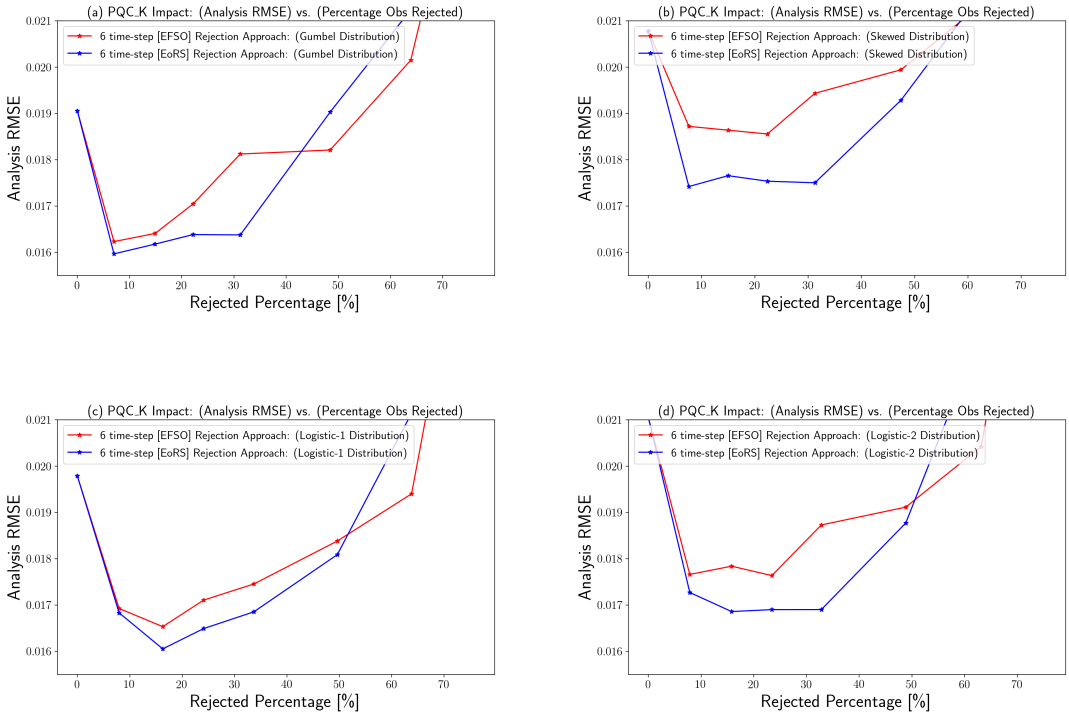

Figure 4.11: Analysis RMSE versus rejection percentage for 6-time step PQC_K experiments where non-Gaussian observation error distributions characterized in 4.7 have been introduced. Model forcing error [F=8.00075] is introduced in all cases. EoRS is plotted with blue symbols and lines, and "pure-EFSO" is plotted with red symbols and lines. As such, EoRS (i.e. application of "EFSO-components" in PQC) is more effective in reducing analysis RMSE. a) Results for the "Gumbel" distribution, b) same as (a), but for the "Skewed" distribution, c) same as (a), but for the "Logistic-1" distribution, d) same as (a), but for the "Logistic-2" distribution.

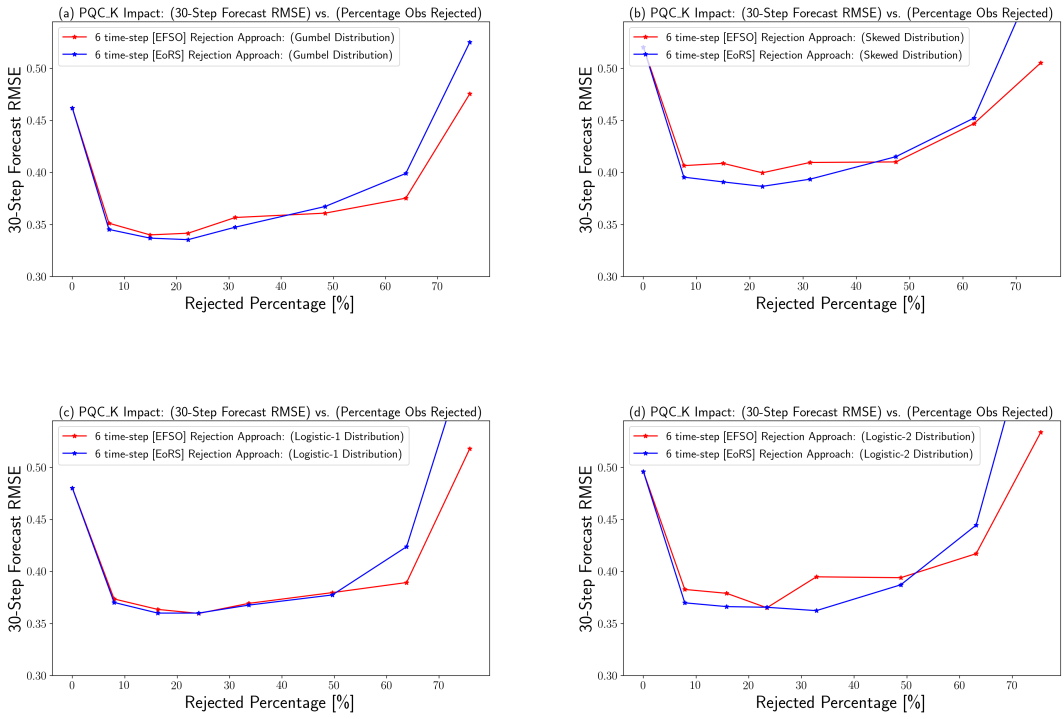

Figure 4.12: 30-step forecast RMSE versus rejection percentage for 6-time step PQC_K experiments where non-Gaussian observation error distributions characterized in 4.7 have been introduced. EoRS is plotted with blue symbols and lines, and "pure-EFSO" is plotted with red symbols and lines. As such, EoRS (i.e. application of "EFSO-components" in PQC) is generally more effective in reducing 30-stef forecast RMSE. a) Results for the "Gumbel" distribution, b) same as (a), but for the "Skewed" distribution, c) same as (a), but for the "Logistic-1" distribution, d) same as (a), but for the "Logistic-2" distribution.

Chapter 5: PEFSO: Testing EFSO Emulation for Unassimilated Observations and Potential Observing System Design Applications

5.1 Introduction

In this chapter the Lorenz '96 testing apparatus is used to assess the potential application of PEFSO. As was discussed in chapter 2 of this dissertation, PEFSO is a straightforward extension of the EFSO formulation, and is intended to predict EFSO impact estimation for unassimilated observations. That is, the intent in using PEFSO is that it can provide useful impact estimation for unassimilated observations had they been hypothetically assimilated. Such a capability would make it possible to quantitatively compare the efficiency of multiple observing systems in reducing forecast error without having to perform and run additional analyses and forecasts. This represents a possible extension to the general "non-Cycled" EFSO approach introduced in Lien et al. 2018. Furthermore, the intent is that PEFSO could help in achieving representative EFSO summary statistics for operational NWP systems where data selection procedures are applied in consideration of computational costs. As such, in this chapter we assess experiments performed to test the accuracy of PEFSO in emulating EFSO, and the extent to which PEFSO might provide comparative information as to the efficiency with which

hypothetical observing systems reduce forecast error without having to perform and run additional analyses and forecasts. Additionally, sample comparisons between PEFSO and EFSO using EnSRF products from a low resolution experimental version of the 2016 GFS are provided and assessed.

5.2 PEFSO Testing Methodology Overview

Lorenz '96 OSSEs are performed where a "control" set of observations are assimilated at all grid-points. For the "control" sets of observations, flaws are introduced at only the odd grid-points. Using background forecasts from these control experiments, analyses are then performed where observations hypothetically located at odd grid-points are deleted from the corresponding columns of the observation operator (i.e. staggered deletion). PEFSO calculations are then performed for the deleted "control" set of observations and "experimental" sets of unassimilated observations located at the odd grid-points. The potential applicability of PEFSO is then assessed according to the extent to which PEFSO is able to emulate EFSO calculations for the "control" observations and the extent to which PEFSO provides information about expected increases or decreases in forecast error for the unassimilated "experimental" observation sets located at odd grid-points.

5.3 Random Flaws PEFSO Experiments

A "control" Lorenz '96 OSSE was performed in which the prescribed observation error standard deviation is again 0.1 in all cases, but the standard deviation of observation

errors introduced at the odd grid-points is 0.135. A representative scatter plot of PEFSO versus EFSO for the "control" set of random observation flaws located at odd grid-points is provided in figure 5.1. For this random flaws experiment, it is apparent that PEFSO

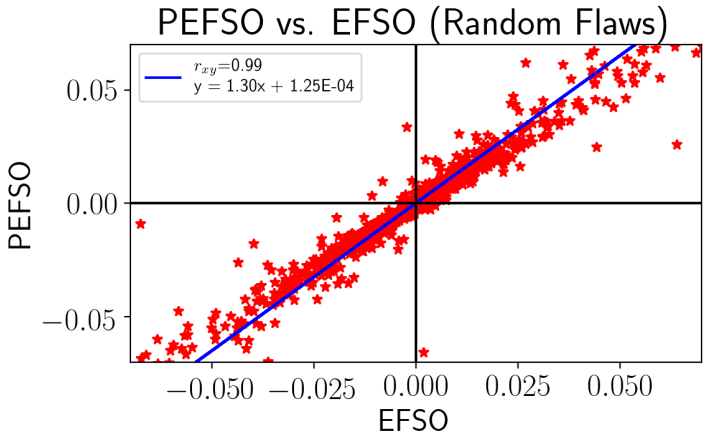

Figure 5.1: Representative scatter plot of PEFSO versus EFSO for a "control" set of observations located at odd grid-points in the Lorenz '96 system. The "control" set of observations at the odd grid points included random observation errors exceeding what was prescribed. The linear regression of PEFSO to EFSO is indicted by the blue line. The Pearson correlation coefficient and linear regression coefficients are indicated in the legend.

is largely able to emulate the directionality of EFSO. However, the PEFSO estimates are inflated relative to the EFSO quantities, reflecting additional contraction of the ensemble when all observations are assimilated relative to what is realized when only half of the Lorenz '96 sector space is observed. The offset in this case is generally 2 orders of magnitude smaller than the impact estimate quantities.

To assess the potential utility of PEFSO in providing comparative information for future (or hypothetical) observing systems without having to perform and run additional analyses and forecasts, PEFSO calculations are performed for 4 experimental observation sets located at odd grid-points. In all cases the background forecasts used are from

OSSEs where the full "control" set of observations were assimilated. The errors in the experimental observation sets have been scaled relative to the control, and the standard deviation of observation error for these scaled experimental observation sets is provided in table 5.1. The offsets when performing linear regression of PEFSO to the "control" EFSO

Table 5.1: The standard deviation of errors for "experimental" observation sets and the offset when performing linear regression between PEFSO calculations for the "experimental" observation sets and EFSO calculations for the "control" set of observations located at odd grid-points.

Random Flaws Experiments		
Experiment Name	Observation Error Standard Deviation	Linear Fit Offset
Exp-1	0.100	-7.67E-04
Exp-2	0.120	-2.58E-04
Control	0.135	1.25E-04
Exp-3	0.152	5.6E-04
Exp-4	0.170	1.03E-03

calculations are also provided in table 5.1. The implied cross-experiment directionality changes according to these offsets indicate that PEFSO was successful in providing useful information as to which observation sets were more or less flawed. That is, the PEFSO offset increases monotonically with the extent of flaws evaluated.

5.4 Systematic Flaws PEFSO Experiments

A Lorenz '96 OSSE was performed for a "control" set of assimilated observations where systematic observation errors of 0.035 are located at odd grid-points. A representative scatter plot of PEFSO versus EFSO for this "control" set of observations located at odd grid-points is provided in figure 5.2. For this "control" systematic flaws experiment, it is apparent that PEFSO is largely able to emulate the directionality of

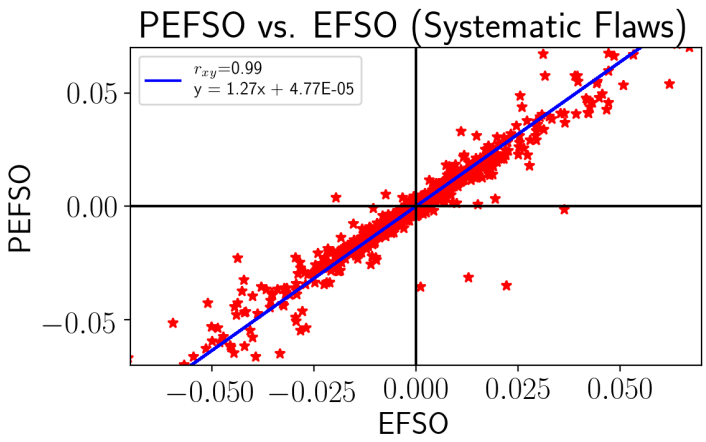

Figure 5.2: Representative scatter plot of PEFSO versus EFSO for a "control" set of observations located at odd grid-points in the Lorenz '96 system. The "control" set of observations at the odd grid points includes systematic observation errors. The linear regression of PEFSO on EFSO is indicted by the blue line. The Pearson correlation coefficient and linear fit parameters are indicated in the legend.

EFSO. However, the PEFSO estimates are inflated relative to the EFSO quantities, again reflecting the additional contraction of the ensemble when all observations are assimilated relative to what is realized when only half of the Lorenz '96 sector space is observed. The offset in this case is generally 2 orders of magnitude smaller than the impact estimates.

To assess the potential of PEFSO in providing comparative information for future (or hypothetical) observing systems without having to perform and run additional analyses and forecasts, PEFSO calculations were performed for 4 experimental observation sets located at odd grid-points in the Lorenz '96 system. In all cases the background forecasts used are from OSSEs where the full "control" set of observations were assimilated. The experimental observation sets have associated systematic errors as indicated in table 5.2. The offsets when performing linear regression of PEFSO to the "control" EFSO calculations are also provided in table 5.2. The implied directionality

Table 5.2: The systematic errors associated with experimental observation sets and the offset when performing a linear fit between PEFSO for the "experimental" observations and EFSO for a "control" set of observations located at odd grid-points.

Systematic Flaws Experiments		
Experiment Name	Systematic Observation Error	Linear Fit Offset
Exp-1	-0.07	-4.1E-03
Exp-2	0.07	1.34E-03
Control	0.035	4.77E-05
Exp-3	-0.004	-1.45E-03
Exp-4	0.004	-1.14E-03

change according to the reported offset for "experiment-1" in 5.2, relative to the PEFSO "control" offset, is not consistent with what would be expected from a realized impact standpoint. The magnitude of systematic errors in the "experiment-1" observation sets are larger than the magnitude of systematic errors in the "control" observation sets at odd grid-points, but the PEFSO calculations suggest that the experiment-1 observation sets are more efficient in reducing forecast errors. To understand this result, recall from chapter 3 that SOPA directionality reflects tugging on cross-covariances associated with assimilation of systematic observation errors over successive antecedent cycles. This tugging ultimately establishes directions in which cross-covariances tend to project onto forecast error anomalies. For the PEFSO calculations here, this tugging is from the "control" observation set which provides the background forecasts in all cases. As such, PEFSO is providing information about the direction of the "experimental-1" systematic errors w.r.t these established directions. However, from an application standpoint the intent is to obtain comparative information regarding the distance of "experimental" or "control" observation sets from the "truth".

5.5 Lorenz '96 PEFSO Experiments Summary

The experimental results in this chapter suggest that PEFSO will largely emulate directionality of EFSO when the error characteristics of the unassimilated observations are consistent with the error characteristics of assimilated observations. Furthermore, since the PEFSO approach to comparing multiple observing systems would involve using a common set of background forecasts (i.e. the "corrective" information is constant across experiments) it is effective at comparing the extent of random errors in unassimilated (or hypothetical) observing systems. However, when systematic observation errors are present in the "control" cycling that provides the common background forecasts, PEFSO information regarding systematic errors in "experimental" datasets is relative to established directions in applied cross-covariances associated with tugging from the "control" set of observations over successive antecedent cycles. From an application standpoint, this information regarding systematic observation errors is inconsistent with what is intended.

We conclude that PEFSO, "non-cycled" EFSO (Lien et al. 2018), "cycled" EFSO and "EFSO-components" provide complementary utility. That is, controlling for background forecasts provides utility in identifying random flaws in observations, but has the potential to obscure information regarding systematic observation errors that pertain to established directions over antecedent cycles. From a standpoint of determining where to apply "non-cycled" EFSO (Lien et al. 2018), PEFSO or "cycled" EFSO to compare the effectiveness of future observing systems, availability of "EFSO-components" may be helpful in identifying where established directions in cycling are prominent. As such,

"EFSO-components" could be used to determine the extent to which "cycling" EFSO, "non-cycled" EFSO (Lien et al. 2018) or PEFSO should be applied in any particular circumstance.

5.6 Comparisons of PEFSO and EFSO for Real Observations

In consideration of computational efficiency, data selection procedures had been applied within EnSRF (Whitaker and Hamill 2002) data assimilation supporting the 2016 4DEnVar GFS (Kleist and Ide 2015). Specifically, observations were discarded in the applied EnSRF where the estimated ratio of analysis error covariance to background error covariance exceeded 0.98. This data selection can significantly reduce the sets of assimilated observations relative to what is considered (Todling and El Akkraoui 2018). In this context it may be of interest to assess whether PEFSO can help in achieving representative EFSO summary statistics for operational NWP systems where such data selection procedures are applied. In figure 5.3 sample scatter plots of PEFSO versus EFSO are provided for observations where the estimated ratio of analysis error covariance to background error covariance exceeded 0.98 according to EnSRF (Whitaker and Hamill 2002) applied data selection procedures. Consistent with the Lorenz '96 results, where the error characteristics of assimilated and unassimilated observations are similar, PEFSO is largely able to account for directionality in EFSO (Guo-Yuan Lien and Wen-Hsin Teng reported similar results for the Central Weather Bureau (CWB) system: Email Correspondence). Furthermore, the magnitudes of the impact estimates are inflated in a manner similar to that for the Lorenz '96 experiments. However, the

extent of this PEFSO inflation, relative to EFSO, does seem to depend on the type of observing system. For example, the inflation for aircraft observations is larger than that for advanced microwave sounding unit-A (AMSUA) and infrared atmospheric sounding interferometer (IASI) satellite-based radiances. A possible explanation is that ensemble contraction along aircraft flight tracks is more pronounced relative to what is generally co-located with the more homogeneous distribution of satellite radiance observations.

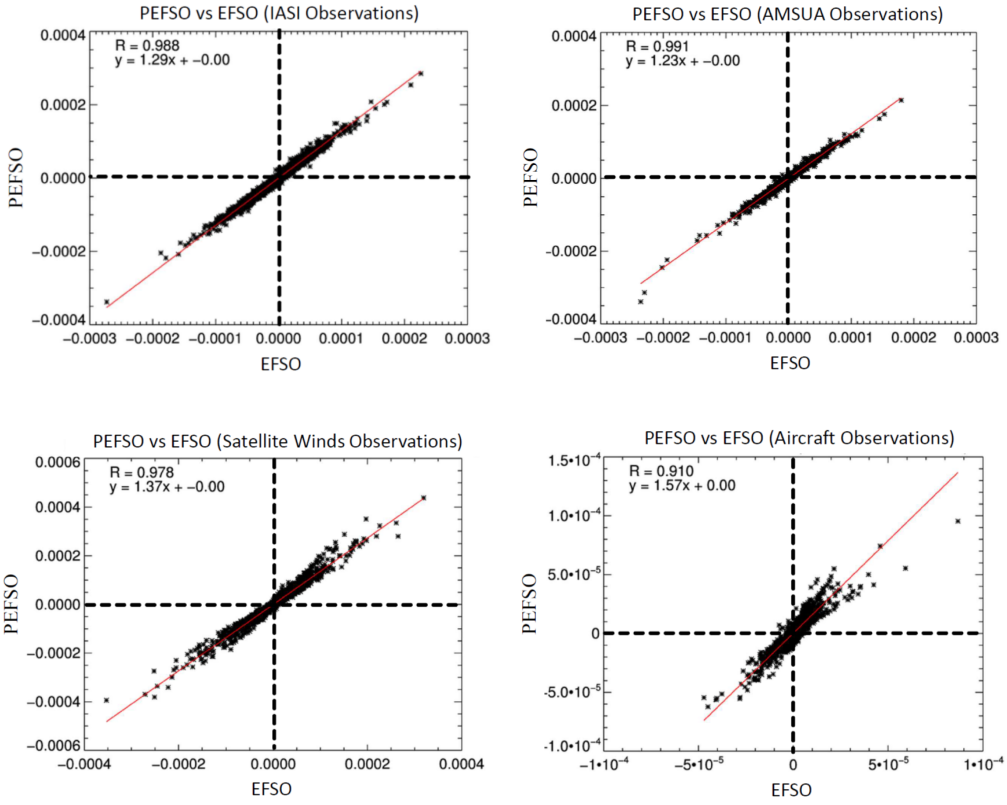

Figure 5.3: PEFSO versus EFSO for an experimental low resolution version of the 2016 GFS. The PEFSO calculations correspond to observations where the estimated ratio of analysis error covariance to background error covariance exceeded 0.98 according to EnSRF applied data selection procedures. The Pearson correlation and coefficients corresponding to linear regression of PEFSO on EFSO is indicated in the legend.

Chapter 6: Summary and Future Application Directed Development

6.1 Summary

A new quantitative framework intended to further extend application of EFSO information in NWP was introduced. Proof of concept Lorenz '96 EFSO OSSEs demonstrate that this quantitative framework, we call "EFSO-components", provides advantages in detection of observation flaws beyond what can be achieved when access to only the overall EFSO calculation is available, and provides rejection decision advantages for cycling PQC relative to what can be achieved using "pure-EFSO" based rejection. The PQC advantages enabled by "EFSO-components" are robust to:

- the lead-time used in EFSO and "EFSO-components" impact estimation
- the extent to which systematic or random observation flaws are present
- the introduction of model forcing error
- the performance metric applied (e.g. analysis and 30-step forecast RMSE reduction)
- the PQC algorithm and associated Kalman gain used to remove analysis contributions from rejected observations

- the extent to which assimilated observation distributions include kurtosis and skew

The aforementioned "EFSO-components" cycling PQC advantages are in part explained by the distribution of "corrective" innovation contributions across components. This distribution of "corrective" contributions provides direct and indirect advantages for cycling PQC observation rejection decisions in mitigating the extent to which "cumulative" contextual considerations hide observation errors and flaws. Furthermore, the combined directionality information from "EFSO-components" enables observation rejection decisions in PQC that account for both immediate considerations at a specific cycle time, via ROPA, and directions that are established over successive antecedent cycles relating to assimilation of systematic observation errors, via SOPA. This more comprehensive approach to making observation rejection decisions provides advantages in addressing potential application limitations associated with the "simultaneous" aspect of EFSO as discussed in chapter 4 of this dissertation. Experiments were performed to demonstrate that "EFSO-components" based PQC preferentially rejects observations according to the extent of their errors and flaws relative to "pure-EFSO" based PQC. As such, we conclude that "EFSO-components" provides fundamental advantages in better extracting the "predictive" information discussed in chapter 1 of this dissertation.

6.2 Future Application Directed Development

We propose testing the fundamental application advantages of "EFSO-components", in better extracting "predictive" information, to higher dimensional operational NWP systems. In general, we expect that advantages of "EFSO-components"

will be most significant where all four component mechansisms identified contribute substantively. This expectation is based primarily on the fundamental combined directionality advantages enabled by "EFSO-components" and the presence of cross-component interactions that effectively serve to hide observation flaws in the overall EFSO calculation.

We expect that a complication with implementing "EFSO-components" in an operational system is the space and time dependent nature of model bias. As such, we propose application of the "decaying average" approach that was originally introduced in (Cui et al. 2012) as a post-processing algorithm for bias correction of National centers for environmental prediction (NCEP) and meteorological service of canada (MSC) global ensemble forecasts. This approach could be applied to obtain space and time dependent systematic contributions to innovation and ensemble mean forecast errors applied in EFSO calculations. Furthermore, it is expected that best rejection strategies using "EFSO-components" in an operational system will depend on which components provide substantive impact in any given circumstance, and as such, it is expected this would require experimental testing and tuning of rejection thresholds by component.

It is expected that concepts developed here can be successfully extended to forecast sensitivity to observation error covariance applications (Hotta et al. 2017b,Daescu and Todling 2010). That is, we propose that an "EFSR-components" capability could be introduced to better extract "predictive" forecast-sensitivity based observation error covariance information. In particular, we expect availability of combined directionality advantages, as demonstrated in this study, could be used to improve observation error covariance tuning decisions.

Another advantage of "EFSO-components" is that it enables identification of specific mechanisms that project onto forecast error. This more specific information could be used in NWP applications to improve the effectiveness of observations assimilated in reducing forecast errors. More specifically, we expect that the SOPA component could be used as a basis for improving observation bias correction in NWP. In association with a JCSDA newsletter and related conference presentations, Groff 2017 demonstrated that EFSO impacts for CrIS satellite radiances depend significantly on the result of the bias correction process. Access to the SOPA component may provide guidance regarding such dependencies and opportunities for improving observation bias correction in NWP.

Bibliography

Alpert J. C., D. L. Carlis, B. A. Ballish, and V. K. Kumar. Using Pseudo RAOB Observations to Study GFS Skill Score Dropouts. In *23rd Conference onWeather Analysis and Forecasting/19th Conference on Numerical Weather Pre-diction, Omaha, NE, American Meteorological Society, Extended Abstract.*, 2009.

E. Anderson and H. Järvinen. Variational quality control. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 1 1999. ISSN 00359009. doi: 10.1002/qj.49712555416.

M. Atkins. Quality control, Selection and Processing of Observations in Meteorological Office's Operational Forecast System. In *Workshop on Use and Quality Control of Meteorological Observations*, pages 255–290, Shinfield Park, Reading, 1984. ECMWF.

P. Bauer. Observing System Experiments (OSE) to Estimate the Impact of Observations in NWP. In *ECMWF Workshop on Diagnostics of data assimilation system performance*, Shinfield Park, Reading, 6 2009. ECMWF.

C. H. Bishop, B. J. Etherton, and S. J. Majumdar. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, 129 (3), 3 2001. ISSN 0027-0644. doi: 10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2.

V. Bjerknes. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (The problem of weather prediction, considered from the viewpoints of mechanics and physics) (translated and edited by VOLKEN E. and S. BR ONNIMANN. – Meteorol. Z. 18 (2009), 663–667). . 1904.

A. Blum. *The Weather Machine*. HarperLuxe, 9 2018. doi: 10.1080/00431672.2020.1736477.

M. Buehner, P. Du, and J. Bédard. A New Approach for Estimating the Observation Impact in Ensemble–Variational Data Assimilation. *Monthly Weather Review*, 146(2), 2 2018. ISSN 0027-0644. doi: 10.1175/MWR-D-17-0252.1.

T.-C. Chen. *Applications of Ensemble Forecast Sensitivity to Observations For Improving Numerical Weather Prediction*. PhD thesis, University of Maryland, 2018. URL https://drum.lib.umd.edu/handle/1903/21165?show=full.

T. C. Chen and E. Kalnay. Proactive quality control: Observing system simulation experiments with the Lorenz '96 model. *Monthly Weather Review*, 147(1), 2019. ISSN 15200493. doi: 10.1175/MWR-D-18-0138.1.

T.-C. Chen and E. Kalnay. Proactive Quality Control: Observing System Experiments Using the NCEP Global Forecast System. *Monthly Weather Review*, 148(9), 2020. ISSN 0027-0644. doi: 10.1175/mwr-d-20-0001.1.

W. G. Collins and L. S. Gandin. Comprehensive Hydrostatic Quality Control at the National Meteorological Center. *Monthly Weather Review*, 118(12), 12 1990. ISSN 0027-0644. doi: 10.1175/1520-0493(1990)118⟨2752:CHQCAT⟩2.0.CO;2.

B. Cui, Z. Toth, Y. Zhu, and D. Hou. Bias Correction for Global Ensemble Forecast. *Weather and Forecasting*, 27(2), 4 2012. ISSN 0882-8156. doi: 10.1175/WAF-D-11-00011.1.

D. N. Daescu and R. Todling. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quarterly Journal of the Royal Meteorological Society*, 136(653), 10 2010. ISSN 00359009. doi: 10.1002/qj.693.

D. P. Dee, L. Rukhovets, R. Todling, A. M. da Silva, and J. W. Larson. An Adaptive Buddy Check for Observational Quality Control. *Quarterly Journal of the Royal Meteorological Society*, 127(577), 10 2001. ISSN 00359009. doi: 10.1002/qj.49712757714.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 10 2005. ISSN 00359009. doi: 10.1256/qj.05.108.

G. DiMego, P. A. Phoebus, and J. E. McDonnel. Data processing and quality control for optimum interpolation analysis at the National Meteorological Center. *NMC Office Note, 306, 38 pp*, 1985.

G. Dyson. *Turing's Cathedral*. Vintage, 1st edition, 12 2012.

M. Ehrendorfer, R. M. Errico, and K. D. Raeder. Singular-Vector Perturbation Growth in a Primitive Equation Model with Moist Physics. *Journal of the Atmospheric Sciences*, 56(11), 6 1999. ISSN 0022-4928. doi: 10.1175/1520-0469(1999)056⟨1627:SVPGIA⟩2.0.CO;2.

L. S. Gandin. Complex Quality Control of Meteorological Observations. *Monthly Weather Review*, 116(5), 5 1988. ISSN 0027-0644. doi: 10.1175/1520-0493(1988)116⟨1137:CQCOMO⟩2.0.CO;2.

R. Gelaro and Y. Zhu. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 61(2), 2009. ISSN 02806495. doi: 10.1111/j.1600-0870.2008.00388.x.

J. Gleick. *Chaos: Making a New Science*. Penguin Books, 1988.

D. Groff. Assessment of Ensemble Forecast Sensitivity to Observation (EFSO) Quantities for Satellite Radiances Assimilated in the 4DEnVar GFS. *JCSDA Newsletter*, 54:1–5, 2017. URL `https://static1.squarespace.com/static/5bad1a12c2ff616821035c9f/t/5c2fa1c58a922de016a4429d/1546625484217/2017_01JCSDAQuarterly.pdf`.

R. N. Hoffman and R. Atlas. Future Observing System Simulation Experiments. *Bulletin of the American Meteorological Society*, 97(9), 9 2016. ISSN 0003-0007. doi: 10.1175/BAMS-D-15-00200.1.

A. Hollingsworth and P. LÖnnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2), 1 1986. ISSN 1600-0870. doi: 10.3402/tellusa.v38i2.11707.

D. Hotta. *Proactive Quality Control based on Ensemble Forecast Sensitivity to Observations*. PhD thesis, University of Maryland, College Park, MD, 2014.

D. Hotta, T. C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi. Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Monthly Weather Review*, 145(8), 2017a. ISSN 15200493. doi: 10.1175/MWR-D-16-0290.1.

D. Hotta, E. Kalnay, Y. Ota, and T. Miyoshi. EFSR: Ensemble forecast sensitivity to observation error covariance. *Monthly Weather Review*, 145(12), 2017b. ISSN 15200493. doi: 10.1175/MWR-D-17-0122.1.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 1st Edition, 12 2002.

E. Kalnay, Y. Ota, T. Miyoshi, and J. Liu. A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1), 2012. ISSN 0280-6495. doi: 10.3402/tellusa.v64i0.18462.

M. Kazumori. Satellite Radiance Assimilation in the JMA Operational Mesoscale 4DVAR System. *Monthly Weather Review*, 142(3), 3 2014. ISSN 0027-0644. doi: 10.1175/MWR-D-13-00135.1.

D. T. Kleist and K. Ide. 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), 2 2015. ISSN 0027-0644. doi: 10.1175/MWR-D-13-00350.1.

R. H. Langland and N. L. Baker. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, 56(3), 5 2004. ISSN 0280-6495. doi: 10.1111/j.1600-0870.2004.00056.x.

E. Larson. *Isaac's Storm*. Random House, 1999.

D. Laskin. *Children's Blizzard*. New York: Harper Collins, 2004.

G. Y. Lien, D. Hotta, E. Kalnay, T. Miyoshi, and T. C. Chen. Accelerating assimilation development for new observing systems using EFSO. *Nonlinear Processes in Geophysics*, 25(1), 2018. ISSN 16077946. doi: 10.5194/npg-25-129-2018.

J. Liu and E. Kalnay. Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 134(634), 7 2008. ISSN 00359009. doi: 10.1002/qj.280.

Lönnberg P and Shaw DB. Data selection and quality control in the ECMWF analysis system. In *Workshop on Use and Quality Control of Meteorological Observations, 6-9 November 1984*, Shinfield Park, Reading, 1984.

S. Lord, G. Gayno, and F. Yang. Analysis of an Observing System Experiment for the Joint Polar Satellite System. *Bulletin of the American Meteorological Society*, 97(8), 8 2016. ISSN 0003-0007. doi: 10.1175/BAMS-D-14-00207.1.

A. C. Lorenc and O. Hammon. Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *Quarterly Journal of the Royal Meteorological Society*, 114(480), 1 1988. ISSN 00359009. doi: 10.1002/qj.49711448012.

E. N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20 (2), 3 1963. ISSN 0022-4928. doi: 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2.

E. N. Lorenz. Predictability – a problem partly solved. In T. Palmer and R. Hagedorn, editors, *Predictability of Weather and Climate*. Cambridge University Press, Cambridge, 1996. doi: 10.1017/CBO9780511617652.004.

E. N. Lorenz and K. A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55(3), 1998. ISSN 00224928. doi: 10.1175/1520-0469(1998)055⟨0399:OSFSWO⟩2.0.CO;2.

K. Onogi. A Data Quality Control Method Using Forecasted Horizontal Gradient and Tendency in a NWP System. *Journal of the Meteorological Society of Japan. Ser. II*, 76(4), 1998. ISSN 0026-1165. doi: 10.2151/jmsj1965.76.4{_}497.

Y. Ota, J. C. Derber, E. Kalnay, and T. Miyoshi. Ensemble-based observation impact estimates using the NCEP GFS. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 2013. ISSN 1600-0870. doi: 10.3402/tellusa.v65i0.20038.

R. J. Purser. A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. In *AMS 10th Conference on Weather Forecasting and Analysis*, pages 102–105, Clearwater Beach, Florida, 1984.

E. Satterfield, D. Hodyss, D. D. Kuhl, and C. H. Bishop. Investigating the Use of Ensemble Variance to Predict Observation Error of Representation. *Monthly Weather Review*, 145(2), 2 2017. ISSN 0027-0644. doi: 10.1175/MWR-D-16-0299.1.

R. Todling and A. El Akkraoui. The GMAO Hybrid Ensemble-Variational Atmospheric Data Assimilation System: Version 2.0. *Technical Report Series on Global Modeling and Data Assimilation*, 50, 2018.

J. S. Whitaker and T. M. Hamill. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130(7), 7 2002. ISSN 0027-0644. doi: 10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2.

J. Woollen. New NMC operational 0 1 quality contro. In *American Meteorological Society, 9th conference on numerical weather prediction*, pages 24–27, Denver, CO, 1991.

X. Zeng, R. Atlas, R. J. Birk, F. H. Carr, M. J. Carrier, L. Cucurull, W. H. Hooke, E. Kalnay, R. Murtugudde, D. J. Posselt, J. L. Russell, D. P. Tyndall, R. A. Weller, and F. Zhang. Use of Observing System Simulation Experiments in the United States. *Bulletin of the American Meteorological Society*, 101(8), 8 2020. ISSN 0003-0007. doi: 10.1175/BAMS-D-19-0155.1.