UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FORECAST SENSITIVITY TO OBSERVATIONS USING DATA DENIAL AND ENSEMBLE-BASED METHODS OVER THE DALLAS-FORT WORTH TESTBED

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

NICHOLAS ANTONIO GASPERONI
Norman, Oklahoma
2018

FORECAST SENSITIVITY TO OBSERVATIONS USING DATA DENIAL AND ENSEMBLE-BASED METHODS OVER THE DALLAS-FORT WORTH TESTBED

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Xuguang Wang, Chair

_____
Dr. Frederick Carr, Co-Chair

_____
Dr. Ming Xue

_____
Dr. Keith Brewster

_____
Dr. Deepak Devegowda

Acknowledgements

I am deeply grateful for the guidance of my main advisor, Dr. Xuguang Wang. She put me in an excellent position to succeed with my research goals despite the many challenges encountered. I would like to acknowledge all of my committee members – Drs. Frederick Carr, Keith Brewster, Ming Xue, and Deepak Devegowda – for their support. I am also very thankful for the input and support for the project provided by Drs. Fred Carr and Keith Brewster. The different perspectives allowed my research to grow in unique directions. This dissertation would not exist if not for the advice of my Master's thesis advisor, Dr. Ming Xue, who was supportive in pursuing for me the OU Alumni Fellowship and instrumental in convincing me to stay on to pursue this Ph.D. I would also like to thank several fellow past and present graduate students and colleagues for our research and general discussions including Lee Carlaw, Yongzuo Li, Sam Degelia, Aaron Johnson, Hristo Chipilski, and Matthew Morris.

Computer resources used for experiments in this dissertation were provided by the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma, the Texas Advanced Computing Center (TACC) at the University of Texas, and the National Science Foundation XSEDE program. I want to thank director Henry Neeman and the rest of the OSCER staff for being very responsive to any computing issues encountered. I would also like to acknowledge and thank the Oklahoma Mesonet and the Meteorological Assimilation Data Ingest System (MADIS) at NCEP for providing observation

datasets used in this work. I also want to thank Yoichiro Ota for providing me with his code to calculate EFSO within the GSI-EnKF system.

On a personal note, I absolutely must thank my wonderful, patient, caring wife Agatha-Beth, without whom I do not think I could have ever gotten through this most trying time of my life. You were there for me for all of my ups and downs, and I can unequivocally say I would not be where I am today without you. Special thanks to sisters Rachael and Vittoria, and close friends Amanda Kis, Wade DeJager, Virginia Silvis, and Mike Hunter for the great moments of fun we had among the long Ph.D. grind! Finally, I would like to think my parents, Deborah and Vittorio, for the love and support you provide me every day even from 1,100 miles away.

Table of Contents

List of Tables

List of Figures

Abstract

The 'Nationwide Network of Networks' (NNoN) concept was introduced by the National Research Council to address the growing need for a national mesoscale observing system. Part of this growing need is the continued advancement toward accurate high-resolution numerical weather prediction. The research testbed known as the Dallas – Fort Worth (DFW) Urban Demonstration Network was created to experiment with many kinds of mesoscale observations that could be used in a data assimilation system, in order to identify observational systems that are most impactful on high-resolution forecasts. Many observation systems have been implemented for the DFW testbed, including Earth Networks (ERNET) Weather Bug surface stations, Citizen Weather Observer Program (CWOP) amateur surface stations, Global Science and Technology (GST) mobile truck observations, CASA X-band radars, SODARs, and radiometers. These 'nonconventional' observations are combined with conventional operational data from METARs, mesonet, aircraft, rawinsondes, profilers, and operational radars to form the testbed network. A principal component of the NNoN effort is the quantification of observation impact from several different sources of information. This dissertation covers two main themes related to quantifying the impact that observations have on forecasts.

The first part is the quantification of impact using data denial experiments, or observational simulation experiments. The GSI-based EnKF data assimilation system was used together with the WRF-ARW model to examine impacts of observations assimilated for forecasting convection initiation (CI) in the 3 April 2014 hailstorm case.

Data denial experiments were conducted testing the impact of high-frequency (5-min) assimilation of nonconventional data on the timing and location of CI, as well as on the development of storms as they progress through the testbed domain. Results using ensemble probability of reflectivity and neighborhood ensemble probability of hail show nonconventional observations were necessary to capture local details in the dryline structure causing localized enhanced convergence and leading to CI. Diagnosis of denial-minus-control fields showed the cumulative influence each observing network had on the resulting CI forecast. It was found that most of this impact came from the assimilation of thermodynamic observations. Accurate metadata is found to be crucial to the application of nonconventional observations in high-resolution assimilation and forecasts systems.

The second part of this dissertation explored the application of the ensemble-based forecast sensitivity to observations (EFSO). First, tests using a global two-layer model were performed to identify areas of improvement in the localization methods needed to make EFSO estimates accurate. Due to the time-forecast component, localization of the EFSO metric is more complicated than during traditional assimilation because as forecast time increases the error correlation structures evolve with the flow. Experiments made use of the local ensemble transform Kalman filter (LETKF) with a simple two-layer primitive equation model and simulated observations. Application of an adaptive localization method – *regression confidence factors* (RCF) based on a Monte Carlo "group filter" technique – led to marked improvement especially for longer forecasts and at midlatitudes, when systematically verified against actual impact in RMSE and skill scores. Results showed that the shape,

location, time-dependency, and variable-dependency of RCF localization functions are consistent with underlying dynamical processes of the model. The impact estimates near the equator were not as effective due to large discrepancies between the RCF function and the localization used at assimilation time. These latter results indicated that there exists an inherent relationship between the localization applied during the assimilation time and the proper localization choice for observation impact estimates. Application of RCF for automatically tuned localization is introduced and tested for a single observation experiment.

Next, the EFSO method was applied to the high-resolution CI case from 3 April 2014 and evaluated for accuracy in terms of several verification metrics, including energy norms surface variables, and composite reflectivity. Static and advected localization were applied to EFSO and compared for accuracy to the actual forecast error reduction. The RCF method was also applied to the convective-scale EFSO estimation. Results showed that different verification metrics lead to different forecast length scales of useful estimates. The application of EFSO to reflectivity is hindered by the high nonlinearity of convection, though there were some qualitative insights in its use. The application of RCF localization, while found to reveal the underlying flow-dependence of the case study including the time-forecast component, did not improve upon the advected localization method. This is hypothesized to be due in part to insights gained from the two-layer model work, though other adaptive methods may yet yield better results. Nevertheless, the application of EFSO is appropriate for convective-scale systems on forecast time scales of 90 minutes or less.

Chapter 1: Introduction

Portions of this dissertation are direct excerpts of Gasperoni and Wang (2015; © American Meteorological Society). This includes parts of chapters 1, 2, and 4.

1.1 Background

The field of Numerical Weather Prediction (NWP) has enjoyed a steady increase in weather forecast skill of about one day per decade over the last 4-5 decades, representing a 'quiet revolution' of the accumulation of scientific knowledge over that timeframe (Bauer et al. 2015). According to Simmons and Hollingsworth (2002), this steady improvement can be broadly attributed to improvements in three areas: data assimilation (DA), modeling, and observing systems. Examples of DA improvements include the use of increasingly advanced methods such as four-dimensional variational analysis (4DVAR) and the use of ensembles within the DA methods to directly inform on daily forecast uncertainties. Ensemble methods include variations of the ensemble Kalman filter (EnKF) or hybrid methods that combine ensemble-based and variational-based methods. Improvements in modeling include the use of parameterization schemes to represent the physical processes at sub-grid scales, as well as the continued effort to increase the grid resolution to better resolve finer scale features. The clearest example of improvements to observing systems is the increasing use of global satellite radiance observations, which advanced the predictive skill of the southern hemisphere to match the skill in the northern hemisphere. It would be remiss to not mention a fourth contributor to the steady improvements -

advancements in computing power – without which the many advancements in the DA, modeling, and the use of large-volume observing systems would not be possible.

With the increasing computing power and sophistication of DA methods and NWP models, there is great potential to increase our understanding and forecasting accuracy of high-impact severe weather events. As NWP moves towards grid resolutions that explicitly resolve convection, there are unique challenges to be addressed in the pursuit of skillful high-resolution forecasts. For instance, in terms of DA, a significant unknown remains how to optimally design an ensemble such that smaller meso-beta-scale uncertainty is represented within the ensemble. In terms of modeling, increasing grid resolution reduces but does not eliminate the need for physical parameterization schemes. Further study is needed to refine schemes to work in a modeling environment that partially resolves some physical features but not others. For example, boundary layer rolls are partially resolved at 1 km horizontal resolution, but boundary layer parameterization is still necessary to capture the sub-km-scale processes.

Accurate high-resolution forecasts also require the availability of mesoscale observing networks. Dense observations are needed in both space and time to be able to capture the small scale and rapidly evolving features of severe events. Unfortunately, no such national mesoscale network currently exists. A report by the National Research Council (NRC; 2009) underscored the need for a national mesoscale network of observations and recommended the integration of existing and future mesoscale observations into a nationwide network of networks

(NNoN). Identified deficiencies in measurements include the height of the boundary layer, vertical profiles of humidity and temperature, and soil moisture and temperature profiles. The report further recommended the implementation of research testbeds in order to objectively evaluate the impacts of proposed observing systems for the NNoN concept. Such testbeds facilitate collaboration between researchers, forecasters, measurement specialists, and the private sector to accelerate the transition from research on observing systems to their operational implementation (Dabberdt et al. 2005; National Research Council 2009).

The Collaborative Adaptive Sensing of the Atmosphere (CASA; McLaughlin et al. 2009) Dallas-Fort Worth (DFW) Urban Demonstration Network (DFW Testbed, hereafter) was established as a testbed to carry out the vision of the NNoN concept (National Research Council 2012). It is an ideal location given the DFW area is home to a population exceeding 7 million, two major airports, other transportation centers and many large sports venues; additionally DFW experiences a wide range of hazardous weather from flash flooding to severe storms with large hail, damaging winds, and tornadoes throughout the year. Many in situ surface and remote sensing observation systems have been implemented for the DFW testbed. Surface observing networks include Earth Networks surface stations (ERNET; Earth Networks 2017), amateur stations from the Citizen Weather Observer Program (CWOP; Chadwick 2014), truck-mounted observations from the Mobile Platform Environmental Data network (MoPED; Heppner 2013), and weather stations operated by Understory Weather (2015). Remote sensing observations

include seven X-band radars operated by CASA, two SODARs operated by WeatherFlow (2017), and three microwave radiometers supplied by Radiometrics in partnership with Earth Networks (Bosse et al. 2012).

In an average sense, DA of observations in both time and space improves an NWP forecast. One important goal of the DFW testbed is to measure the impacts from each of the aforementioned observing systems in robust state-of-the-art DA and forecasting systems. In other words, we can investigate by instrument type, observation type, and location, which observations are the most impactful on a forecast. Additionally, we can avoid using observations that have negative impacts on a forecast. Evaluating the usefulness of observations is particularly important for operational NWP centers that operate under limited budgets and need to weigh the costs and benefits of adding more observations to an already large observational dataset.

There are a few basic approaches to quantifying the impact that assimilated observations have on a forecast. Traditionally, quantification of forecast impact from different observational systems is done through observation system experiments (OSEs) or observation system simulation experiments (OSSEs; Atlas 1997; Lord et al. 1997). OSSEs test the potential impacts of future observing systems by creating synthetic observations extracted from a model simulation, known as a "nature run", and treated as the "truth". Experiments with and without the synthetic observations are compared to this truth to quantify their potential benefits and limitations. On the other hand, OSEs test impacts from many different real observation systems already deployed, either in a data addition or data denial

sense. Depending on the OSE approach, the "control" experiment can refer to a baseline experiment before the observation networks to be tested are added one-by-one (data addition); or the control can refer to the "optimal" experiment where all available observations are assimilated and compared with subsequent experiments withholding subsets of observations (data denial).

There is an ever-expanding array of OSE-related research encompassing a wide variety of subtopics in NWP forecasting. The first OSEs were used to test the impact of mainly satellite and upper-air data on forecasts of synoptic systems with global models (e.g., Andersson et al. 1991; Zapotocny et al. 2002, 2007). With the increase in computing power and expanding use of observation networks, more recent OSEs tested data impacts using more sophisticated hybrid DA methods (e.g., Kutty and Wang 2015) and have included impacts of assimilating GPS-derived precipitable water, mesonet observations, and wind profiler data (e.g., Benjamin et al. 2010). Other recent OSEs have focused on specific regions and weather types (e.g., Singh et al. 2014; Coniglio et al. 2016; Zhang et al. 2016). With the increasing use of regional models at convection-allowing horizontal resolutions less than ~4 km, studies have also focused on the effects of assimilating Doppler radar radial winds and reflectivity on high-resolution forecasts of deep convection (e.g., Schenkman et al. 2011a, b; Snook et al. 2015).

While the OSE approach is a popular and straightforward method, it is computationally expensive because of the number of experiments required. A second method to quantifying observation impact is the adjoint-based approach, first explored by Langland and Baker (2004). Different from OSEs, this method can

provide observation impact estimates for all observations simultaneously, without the need for separate data denial or addition experiments. The adjoint method has been applied successfully as an important diagnostic tool (e.g., Cardinali 2009; Langland et al. 2009; Gelaro et al. 2010; Weissmann et al. 2012; Hamill et al. 2013). However, adjoints are generally difficult to create, and because of the tangent linear assumption their application is limited to shorter forecast lengths.

A third approach to evaluating observation impact on a forecast is the ensemble-based method. An approach analogous to the adjoint method of Langland and Baker (2004) was proposed by Liu and Kalnay (2008) with minor correction in Li et al. (2010). Kunii et al. (2012) successfully applied their method by evaluating the impact of real observations in a forecast of Typhoon Sinlaku, using the Weather Research and Forecasting (WRF) model together with the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007). Kalnay et al. (2012) derived a simpler formulation – the "Ensemble Forecast Sensitivity to Observations" or EFSO metric – that makes fewer approximations; it is more general and computationally efficient because it relies on readily-available EnKF products and can be used with any deterministic EnKF method. Ota et al. (2013) successfully applied the Kalnay et al. (2012) formulation to the National Centers for Environmental Prediction (NCEP) Global Forecasting System EnKF (GFS/EnKF; Whitaker et al. 2008), which is now part of the GFS hybrid data assimilation system (Wang et al. 2013).

The Kalnay et al. (2012) observation impact metric is appealing because ensemble perturbations take the place of the adjoint model in estimating

sensitivities. This is particularly beneficial with the expanding use and reliance on ensembles within the field of NWP. However, as with any ensemble method it suffers from sampling error, which occurs when the number of ensemble members is small compared to the degrees of freedom in a model and observing system – predominantly the case in ensemble NWP. Sampling error results in spurious correlations and can lead to filter divergence in deterministic EnKF assimilation, a condition where the ensemble spread becomes too small and disconnects from the true state. Houtekamer and Mitchell (1998) showed that the effects of sampling error can be suppressed by excluding distant observations from influencing the analysis at a given grid point. They experimented with filtering covariance estimates using a distance-dependent correlation function, referred to as *covariance localization* (Houtekamer and Mitchell 2001). Since then, much research has been done in developing localization methods to improve EnKF analyses using limited ensembles (e.g., Hamill et al. 2001; Houtekamer and Mitchell 2005; Anderson 2007; Kepert 2011; Anderson 2012; Anderson and Lei 2013; Holland and Wang 2013).

1.2 Motivation and dissertation overview

The initial success of applying the EFSO to synoptic-scale systems has motivated the question relevant to this dissertation – what are the pros and cons of the EFSO method, and to what degree can it be applied in a high resolution modeling framework? The DFW testbed gives a good testing framework for the application of such a method. Additionally, if successful, the EFSO metric provides

a means to evaluate relative impacts from differing observation platforms within the DFW testbed. The primary method of evaluation within the DFW testbed has been a data denial framework, using a selection of case studies to identify any challenges with the DA of various observations (e.g. Carlaw et al. 2015). However, month-long or seasonal studies are necessary to establish the statistical robustness of the differing observation system impacts within the testbed. The EFSO method is ideal to facilitate such a long-term impact study. However, it is unclear to what degree EFSO can be applied on the convective scale, and this dissertation explores it in detail.

To accomplish this goal, three main research topics are covered within this dissertation. The first is an OSE study within the DFW testbed. The OSE study explores observation impacts on the prediction of convection initiation (CI) within the domain. It also establishes a baseline for comparison for the later application of EFSO. The second topic is an exploration into the EFSO metric itself using a simplified two-layer model approach. This approach limits complications from model complexity so that the focus can be on the benefits and limitations of the method itself. In this way, areas of improvement can be identified and accounted for when transitioning to the application of EFSO to the DFW testbed system. The third area of research is the application of EFSO to convective scale, using the same case study previously used for OSE experiments.

This dissertation is organized as follows. In Chapter 2, ensemble based methods are introduced. This includes background on the EnKF as well as the EFSO method, with a discussion on localization methods to correct for sampling

error. Chapter 3 covers the OSE case study within the DFW testbed using the GSI-based EnKF system extended for use on the convective scale (Johnson et al. 2015). The case study, 3 April 2014, was a dryline CI case with numerous isolated storms that produced large hail and a few weak tornadoes.

In Chapter 4, the EFSO method is evaluated using a simple two-layer model approach to identify potential areas of improvement within the method (Gasperoni and Wang 2015). This is accomplished via application of an adaptive localization technique called *regression confidence factor* (RCFs), which reveals the underlying dynamics of the model using a Monte Carlo approach that utilizes groups of ensembles to evaluate sampling error.

Finally, in Chapter 5 the EFSO method is applied to the same 3 April 2014 DFW testbed case study presented in Chapter 3. Previous studies have focused on synoptic scale application verifying with energy norms. In this case, additional verification metrics more appropriate to convective-scale modeling, such as composite reflectivity and neighborhood probability, are explored for application. The accuracy of EFSO is tested against the impact found via data denial for various verifications. Additionally, the RCF method of Chapter 4 is tested on the convective scale to see if EFSO accuracy can be improved upon and extended for longer forecast periods

A summary and discussion of conclusions is presented in Chapter 6.

Chapter 2: Ensemble-based methods

2.1 Kalman filter

The DA technique used throughout this dissertation is based upon the Kalman filter (KF; Kalman 1960; Kalman and Bucy 1961). The goal of the KF, as with any DA method, is to optimally combine prior information about the atmospheric state (referred to as a background or first guess) with available observations. Often this background is taken from a previous model forecast. To achieve this, it is necessary to provide statistical information about the errors in both the background field and observations. In methods such as optimal interpolation (OI) and variational approaches (3DVAR and 4DVAR), the background error has to be specified for each analysis, referred to as background covariance matrix **B**. In the KF method, the background error covariance is advanced between each analysis step by some linear model. Thus the main difference between KF and other methods is estimation of analysis uncertainty or analysis covariances in addition to the optimal state[1].

Propagating the error within the KF crucially allows for each analysis to take into account the rapidly-evolving "errors of the day". Fully evolving the uncertainty would require numerically integrating the joint probability density function (pdf) of the state, which describes both the state (mean of the pdf) as well as its uncertainty (spread). However, numerically evolving this pdf is

[1] OI and 3DVAR methods assume static covariances, **B.** Within 4DVAR the background error is implicitly evolved by the model, however the estimate of optimal analysis error is unavailable. An initial estimate of background error covariances is thus required for each analysis time.

computationally prohibitive, thus the pdf is assumed to be Gaussian. This ensures that the final pdf, obtained by propagating the initial pdf by the linear model, will remain Gaussian with updated mean and covariance.

There are two steps within the KF update. The first step is the forecast or propagation step, evolving both the state and error covariance with a linear forward model $\mathbf{M}_{i-1}$ from time *i-1* to time *i*. The second step is the analysis or DA step, which for the KF is obtained through least squares minimization or minimization of the mean squared error of the analysis. These steps are shown in matrix form in the following equations.

Forecast step:

$$x_i^b = \mathbf{M}_{i-1} x_{i-1}^a \tag{2.1}$$

$$\mathbf{P}_i^b = \mathbf{M}_{i-1} \mathbf{P}_{i-1}^a \mathbf{M}_{i-1}^T + \mathbf{Q}_{i-1} \tag{2.2}$$

Analysis step:

$$x_i^a = x_i^b + \mathbf{K}_i (y_i^o - \mathbf{H} x_i^b) \tag{2.3}$$

$$\mathbf{K}_i = \mathbf{P}_i^b \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_i^b \mathbf{H}^T + \mathbf{R} \right)^{-1} \tag{2.4}$$

$$\mathbf{P}_i^a = \left(\mathbf{I} - \mathbf{K}_i \mathbf{H} \right) \mathbf{P}_i^b \tag{2.5}$$

Superscripts *a*,*b*, and *o* represent analysis, background, and observation, respectively, and subscript *i* refers to time index. Defining *n* and *p* as the number of model state variables and observations, respectively, each variable is described as follows with matrix sizes in parentheses. Variable *x* is the model state vector (*n* x 1); *y* is the vector of observations (p x 1); **P** is the error covariance matrix of the model state (*n* x *n*); **R** is the observation error covariance matrix (*p* x *p*); **Q** is the model error covariance matrix (*n* x *n*); **H** is the linear observation operator that

converts the state to observation locations (p x n); and **K** is the matrix of optimal weights for every observation-state pair known as the *Kalman gain*[2] (n x p). Equations (2.1)-(2.5) define a recursive algorithm because the background state and covariances for analysis at time i come from the model propagation of a previous analysis at time i-1. It should be noted that the least squares approach within the KF is the same as in OI, and that equations (2.1), (2.3), and (2.4) are equivalent to the equations used for OI, with equations (2.2) and (2.5) showing the addition of error covariance propagation and update.

Several assumptions were made to obtain the KF equations. The success of the KF method in finding the optimal solution depends on the degree to which these assumptions remain valid. The assumptions include: unbiased observations, an unbiased forecast model, uncorrelated observation and forecast errors, knowledge of the observation and model error covariances, and linearity of the model and observation operators. In atmospheric DA applications, many of these assumptions are invalid. The atmosphere is highly nonlinear, some observation types have highly nonlinear observation operators (e.g. reflectivity), and biases often exist within NWP models. Thus the KF is ill suited for NWP applications.

A variation of the KF known as the extended Kalman filter (EKF; Jazwinski 1970) was developed to incorporate nonlinearities of the forecast model and observation operator, though linearized model and observation operators are still used for the covariance propagation and update steps. Additionally, the linear observation operator is used in the Kalman gain calculation. While the EKF

[2] See Kalnay (2003) for a full derivation of the optimal weight matrix **K** obtained within a least-squares framework.

assumptions are better suited for NWP applications, typical model dimensions used in NWP make it computationally prohibitive and thus impractical to implement.

2.2 Ensemble Kalman filter (EnKF)

Both the KF and EKF are prohibitively costly for NWP modeling and other high-dimensional applications. The high cost is related to the propagation step of covariances, which would require on the order of *n* model propagations to fully resolve. A simplification to this step is a Monte Carlo approach utilizing *ensembles* of atmospheric states in the propagation step.

$$x_k^b = M\left(x_k^a(t_{i-1})\right) \qquad k = 1..K \tag{2.6}$$

Equation (2.6) replaces (2.2) in the KF formulation, where an ensemble of *K* model analysis states is propagated forward by the full nonlinear model *M()*. The forecast ensemble can then be used to estimate the forecast error covariances. Such an approach was first applied by Evensen (1994) for oceanographic applications and later applied to atmospheric applications by Houtekamer and Mitchell (1998).

There are several advantages to using the EnKF over the EKF. First, the ensemble forecast replaces the covariance propagation steps, eliminating the need for developing linearized model **M** and observation operator **H** or their corresponding adjoints. The cost of advancing the ensemble to estimate the error covariances is much less than fully evolving them, as for typical NWP applications the number of ensembles *K* is on the order of 10-100. This is a much smaller and more tractable number of model evolutions than if we were to propagate the full

model covariance, which would require n state evolutions (on the order of 10^6 or larger for typical NWP applications). Additionally, the use of ensembles in the DA automatically gives you an initialization ensemble analysis for probabilistic forecast purposes, rather than a final deterministic analysis. Finally, as with the EKF, the EnKF provides fully flow-dependent error covariances including cross-variable covariances, as opposed to OI or variational methods.

There are two types of EnKF algorithms: stochastic and deterministic. In stochastic algorithms, random perturbations consistent with observation error covariances **R** are added to the observations assimilated by each ensemble member(e.g., Houtekamer and Mitchell 1998; Hamill and Snyder 2002). In such a way, each member DA update maintains realistic independence as each ensemble member assimilates a different representation of the original set of observations. A stochastic EnKF has been operational at the Canadian Meteorological Centre (CMC) since 2005 (Houtekamer and Mitchell 2005). With deterministic algorithms, on the other hand, no perturbations are added to the observations. Instead, the observations are used to first update the ensemble mean, followed by an update of ensemble perturbations in a manner consistent with the analysis covariance update of the KF (Equation 2.5). The updated ensemble perturbations are then added to the ensemble mean to provide full updates for each ensemble member. Many flavors of deterministic EnKF algorithms exist, including the ensemble square root filter (EnSRF; Whitaker and Hamill 2002), the ensemble adjustment filter (EAKF; Anderson 2001), and the ensemble transform Kalman filter (ETKF;

Bishop et al. 2001). Differences in these approaches are summarized by Tippett et al. (2003).

Though the EnKF is computationally feasible within NWP applications, there are still computational concerns to take into account. For example, the ensemble-based approximation to the background covariances $\mathbf{P}^b$ can be calculated as follows:

$$\mathbf{P}^b = \frac{1}{K-1}\sum_{k=1}^{K}\left(x_k^b - \overline{x^b}\right)\left(x_k^b - \overline{x^b}\right)^T \tag{2.7}$$

However in practice, there is no need to calculate the full *n* x *n* covariance matrix in model space. Houtekamer and Mitchell (2001) showed that instead, ensemble-based approximations for $\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ can be directly evaluated for use in the Kalman gain computation (2.4):

$$\mathbf{P}^b\mathbf{H}^T = \frac{1}{K-1}\sum_{k=1}^{K}\left(x_k^b - \overline{x^b}\right)\left(H(x_k^b) - \overline{H(x^b)}\right)^T \tag{2.8}$$

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^T = \frac{1}{K-1}\sum_{k=1}^{K}\left(H(x_k^b) - \overline{H(x^b)}\right)\left(H(x_k^b) - \overline{H(x^b)}\right)^T \tag{2.9}$$

Equations (2.8) and (2.9) reduce the number of computations from (*n* x *n)* to (*n* x *p + p* x *p)*, which will be generally smaller as the number of observations *p* is usually much less than the number of state variables *n*. However, as the number of observations increases, the above may still be computationally expensive. Furthermore, the Kalman gain formulation in (2.4) requires a prohibitively costly matrix inversion.

Two different algorithm methods have been developed to avoid the numerical cost of large matrix inversion. The first is called sequential or serial

processing, where observations are assimilated either one at a time or in small batches (e.g., Anderson 2001) . If observations have independent errors, then assimilation sequentially should produce an identical result to assimilating all observations at once. This significantly reduces numerical cost because the matrix inversion becomes a trivial inversion of scalars (in the case of assimilating observations one at a time). The other approach is called the local ensemble transform Kalman filter (LETKF; Ott et al. 2004; Hunt et al. 2007). Within the LETKF, each grid point of the model domain is updated individually using all observations within a predefined radius. This allows for computational efficiency via parallelism, as domain decomposition can be readily implemented within the LETKF (each grid point can be updated independently).

The use of ensembles to estimate forecast covariances can cause problems within the EnKF that may lead to less accurate analyses or even filter divergence. Two of the major issues are sampling error and model error. Sampling error is caused by the use of a limited number of ensemble members relative to the degrees of freedom in the model, and it leads to spurious correlations within the covariance estimates. The remedy to treat sampling error is called covariance *localization*, where covariances are multiplied point-by-point with a correlation function. A review of localization methods is discussed in section 2.2.2.

Model error occurs because the ensemble cannot adequately resolve certain small-scale features or their interaction with larger scales. Additionally, the ensemble does not take into account errors related to physical parameterization schemes (e.g. scheme parameter uncertainty) or other approximations used in the

model (e.g. numerical methods, boundary conditions). When these sources of uncertainty are not accounted for, the final ensemble analysis spread is too small. In fact, the resulting ensembles may only explain roughly *one quarter* of the error variance in the ensemble mean (Houtekamer and Zhang 2016). Thus, covariance *inflation* is necessary to adjust for this underdispersion, by artificially inflating the spread after the EnKF analysis. There are many types of inflation methods that exist, such as additive inflation, multiplicative inflation, and relaxation to prior spread (RTPS; Whitaker and Hamill 2012). Often a combination of more than one type of inflation method is necessary to account for multiple sources of error within a cycled EnKF system.

2.2.1 Ensemble Square Root Filter

The forecast step of the EnSRF is the same as in stochastic EnKF where each ensemble member is propagated forward in time by the full nonlinear model, as in Equation (2.6). The forecast error covariances are then estimated using (2.8) and (2.9). The DA step can be written in two steps –

Mean update:

$$\overline{x^a} = \overline{x^b} + \mathbf{K}\left(y^o - \overline{H(x^b)}\right) \tag{2.10}$$

Perturbation update for each member *k*:

$$x_k'^a = x_k'^b - \tilde{\mathbf{K}} H(x_k^b)' \tag{2.11}$$

Here, $x_k'^a = x_k^a - \overline{x_k^a}$ and $x_k'^b = x_k^b - \overline{x_k^b}$ are ensemble perturbations for the analysis and background, respectively, and $H(x_k^b)' = H(x_k^b) - \overline{H(x^b)}$ are background ensemble

perturbations projected to observation space. Matrix $\tilde{\mathbf{K}}$ is called the *reduced Kalman gain*, which is used only during the perturbation update and is calculated in the following equation.

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T} + \mathbf{R}}}\right)^{-1} \mathbf{K} \tag{2.12}$$

Because the observations are processed serially, all terms in (2.12) are scalars. In the stochastic EnKF, it can be shown that as the ensemble size approaches infinity the ensemble based estimate of analysis covariances converges to the actual analysis covariance of the EKF, but *only if observations are perturbed* (Burgers et al. 1998). The correction in (2.12) is needed because using the full Kalman gain **K** in (2.11) would result in excess spread reduction inconsistent with the analysis covariance update of the EKF in Equation (2.5).

2.2.2 Review of localization methods

The Monte Carlo EnKF method first proposed by Evensen (1994) provided an easily implemented alternative to variational methods that, in theory, provided background flow-dependent error statistics for analysis updates more consistent with the underlying model. However in practice, the use of a limited set of ensembles to estimate errors within a nonlinear model with many more degrees of freedom led to spurious correlation that could lead to filter divergence. Fortunately, covariance localization can be used to mitigate the effects of sampling error by considering only spatially close locations, assuming that distant correlations go to zero. Houtekamer and Mitchell (1998) was the first to attempt

such a procedure, by implementing for each grid point analysis a cutoff radius beyond which observations were not used in the update. In a follow-up article by Houtekamer and Mitchell (2001), the localization method was modified to create more spatially smooth analysis increments. This avoids the introduction of noise and significant imbalance in the analysis from abrupt cutoffs, similar to the noise that data selection introduces within OI analyses (e.g., Cohn et al. 1998).

Houtekamer and Mitchell (2001) utilized the so-called Gaspari-Cohn (GC) localization function of Gaspari and Cohn (1999, Eqn. 4.10), a compactly supported (i.e. goes to 0 at a predefined distance) piecewise polynomial approximation to a Gaussian function. Hamill et al. (2001) showed that the GC function length scale can be optimally tuned. Additionally, the optimal scale increases with increasing ensemble size, allowing for more distant observations to update the analysis because of diminishing noise from sampling error. Their examination of the Eigen spectrum of background error covariance estimates showed that without localization the spectrum was too steep, but with localization it became flatter with sufficient projection onto the tails of the spectrum, effectively adding extra degrees of freedom back. This can also be seen in Figure 2.1. A limited 25-member ensemble shows spurious correlations far away from the observation location (Figure 2.1a), whereas a larger 200-member ensemble shows those locations should have near-0 values (Figure 2.1b). However, when the GC correlation function of Figure 2.1c is applied to the 25-member background covariances, the resulting field (Figure 2.1d) effectively resembles the 200-member correlations.

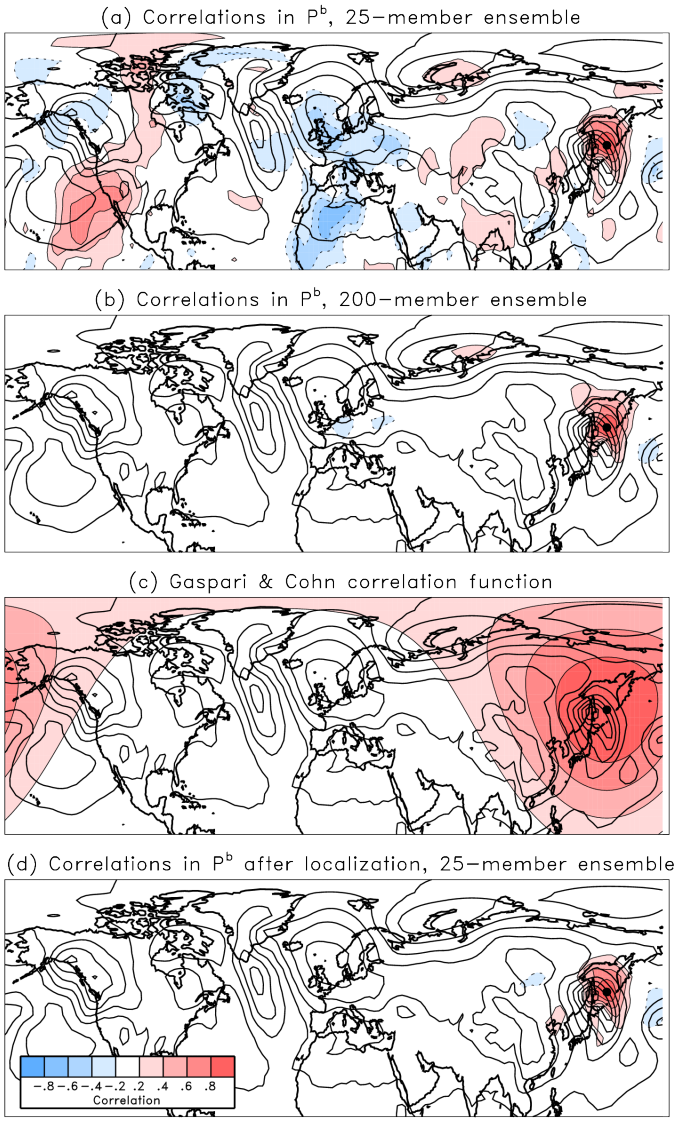

Figure 2.1. Depiction of the effect of covariance localization, adapted from Hamill (2006). Observation is located in East Asia and contours are sea-level pressure, with color contours representing correlations in (a,b,d) and the GC function in (c).

The GC function works well for horizontally removing spurious covariances, since the atmosphere is naturally inclined to diminishing correlation with increasing distance. However, localization in the vertical is also required and there is less theoretical basis for the proper choice of a function (e.g., Anderson 2012).

For instance, surface pressure is valid at the surface, but by the definition of η coordinates impacts all model variables (Houtekamer et al. 2005). Nevertheless, methods for vertical localization have been used which seemingly improve the analyses. Whitaker et al. (2004) created a vertical localization function in σ coordinates that was equal to 1.0 below σ = 0.2 and decreased linearly to 0.0 by σ = 0.05, above which no update was performed. Houtekamer and Mitchell (2005) implemented a vertical version of the GC function, with the natural logarithm of pressure as the vertical coordinate and a cutoff radius of two units in *ln p*.

Compounding the localization problem further are integrated observations that are not local in nature and cannot be represented by a unique spatial location, such as those from remote-sensing platforms (e.g. satellite and Doppler weather radar). Sobash and Stensrud (2013) investigated the impacts of differing horizontal and vertical GC localization scales on the assimilation of radar data for several modes of convection, using OSSEs. They found a large horizontal scale (12-18 km) and small vertical scale (3 km) could improve the radar analyses, which was different from the optimal localization used in Tong and Xue (2005). Several studies have examined the issue of how to apply vertical localization to different microwave-radiance channels of satellite observations, each of which are sensitive to different broad levels of the atmosphere but have significant overlapping regions (e.g., Houtekamer and Mitchell 2005; Houtekamer et al. 2005; Campbell et al. 2010).

There are several ways in which localization can be applied to limit spurious correlations in the ensemble covariances. The first is the so-called **B** localization of the following form in the Kalman gain matrix **K**,

$$\mathbf{K} = \left(\rho_{\mathbf{B}loc} \circ \mathbf{B}\right)\mathbf{H}^T\left[\mathbf{H}\left(\rho_{\mathbf{B}loc} \circ \mathbf{B}\right)\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{2.13}$$

where $\rho_{\mathbf{B}loc}$ is the localization function that operates on the **B** matrix. In practice, full **B** localization is rarely done because both instances of $\rho_{\mathbf{B}loc}$ require O(n^2) operations where n is the number of model state variables. Instead, it is done in observation space, modifying $\left(\rho_{\mathbf{B}loc} \circ \mathbf{B}\right)\mathbf{H}^T$ of (2.13) into $\rho_{\mathbf{BH}loc} \circ \left(\mathbf{BH}^T\right)$, which requires O($np$) operations. This greatly reduces number of operations for NWP models even for abundant observational platforms such as satellite (Campbell et al., 2010). There are other flavors of **B** localization which have been studied by Greybush et al. (2011), Janjić et al. (2011), and Holland and Wang (2013). For the EnSRF, localization can be applied in both the Kalman gain and reduced Kalman gain matrices (Holland and Wang 2013).

Another broad category of localization application is **R** localization, which takes the general form in the Kalman gain of

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \rho_{\mathbf{R}loc} \circ \mathbf{R}\right]^{-1}, \tag{2.14}$$

where now the localization function $\rho_{\mathbf{R}loc}$ operates on the observation error covariance matrix, requiring O(p^2) computations. Typically the LETKF adopts **R** localization, since assimilation is done grid point by grid point using patches of observations, though **B** localization is still possible as shown in Janjić et al. (2011)

and Holland and Wang (2013). The choice of how to apply localization is problem dependent, as there are tradeoffs to the various **B** and **R** methods.

Other types of localization have been proposed in the literature. Kepert (2009, 2011), based on dynamical balance considerations, proposed a 'balance-aware' localization in $\phi\psi\chi$ space rather than ϕuv-space, since streamfunction ψ, velocity potential χ, and geostrophic height ϕ covariances are typically isotropic in nature (u and v are horizontal and vertical components of wind). Such a transformation of variables has been used successfully in OI and variational methods as well. Buehner and Charron (2007) explored localizing in spectral space. Results showed that spectral localization affects correlations at all distances, in contrast to spatial localization, with an overall effect of smoothing correlations in grid point space. They found that the effects of combined spectral and spatial localization were complementary, but optimal combinations were case specific. Buehner (2012) further evaluated a scale-dependent spectral/spatial localization over a one-month assimilation experiment and found that it had similar forecast quality compared to experiments using spatial localization with double the ensemble size, and in some regions improved the forecasts. While there is significant computational cost of implementing spectral/spatial localization, it avoids the added costs of increasing the number of ensemble members.

2.2.2.1 Adaptive localization methods

The localization methods discussed thus far have required some form of global ad hoc tuning to find the optimal cutoff scale for use throughout

assimilation. Such methods could be referred to as static or non-adaptive localizations. However, the highly variable flow-dependent nature of evolving background error covariances means that the optimal cutoff may also vary with time and change shape, such as an elongation along of a cold front (Bishop and Hodyss 2007). There are also problems with vertical localization and localizing observations from nonlocal remote sensing platforms that have no obvious spatial localization. Further, defining the localization 'distance' between variables such as moisture and wind components is difficult, and different variables may require different localizations for the same observations (Anderson 2007). Often, observations may be temporally separated from the analysis time, inviting another complication to the problem (Anderson 2007). These considerations have motivated the development of *adaptive* or *dynamic* localization methods, which refers to methods for which localization functions adapt to the underlying dynamical flow of the model.

The first of these adaptive methods was the *group filter* (GF) method of Anderson (2007), which evaluates the sampling error in a given ensemble by utilizing groups of ensembles to compute an optimal weighting factor. The GF method was found to produce analyses at least as accurate as using GC localization with optimal width, with the benefit of automatic tuning. A drawback of the GF method is the need for g groups of K ensembles to produce a proper localization for the K-member ensemble, which is an added cost. However, Lei and Anderson (2014) have shown that just $g = 2$ groups may be sufficient to produce good results with the GF method. Zhang and Oliver (2010) presented a bootstrap

implementation of the GF method to estimate errors in ensemble covariances, where the use of bootstrapped samples avoids the cost of additional groups of ensembles.

Other adaptive methods have been developed to reduce sampling error using underlying model correlations. Bishop and Hodyss (2007) introduced an adaptive method referred to as Smoothed Ensemble Correlations Raised to a Power (SENCORP) localization. The ensemble correlation matrix of *smoothed* perturbations is raised to a power *m* element-by-element, with the intent of attenuating small spurious correlations while maintaining the real correlations. Then a matrix product smoother is applied to amplify nonzero correlations and the result is raised to another element-by-element power *r* to attenuate remaining spurious correlations. A modification to this localization developed by Bishop and Hodyss (2009a,b) is known as ECO-RAP localization, which contains a different smoothing procedure compared to SENCORP localization. Bishop and Hodyss (2009a) compared the new ECO-RAP method with SENCORP. While the ECO-RAP was not any better or worse than the SENCORP, it is more computationally efficient and contains one less tunable parameter. The drawback of both methods is the number of parameters that require tuning – 4 for SENCORP, 3 for ECO-RAP – and it is unclear how to proceed with optimally tuning other than laborious 'trial and error'.

For a petroleum reservoir application, Chen and Oliver (2010) used the prior ensemble covariance between an observation and a state variable to calculate localization. In another reservoir method, Emerick and Reynolds (2011)

used sensitivity matrices and principal correlation lengths of the model to compute different critical lengths for the GC localization of different observations. Anderson (2012) developed a related algorithm called *sampling error correction* (SEC) that computes localization as a function of ensemble size and pre-defined correlation between an observation and state variable. The SEC method works even with prior correlations as uninformative as having a uniform distribution between -1 and 1, though the algorithm can be made more efficient if the prior correlation distribution can be specified in a more informative way.

More recently, Anderson and Lei (2013) developed the *empirical localization function* (ELF), which computes localizations from the output of an Observing System Simulation Experiment (OSSE) in sets of pairs of observation and state variables binned by distance (e.g. all state variables between 100- and 200-km away from an observation). The ELF localization is computed by minimizing the RMS difference between the true values and posterior ensemble mean of the OSSE. Lei and Anderson (2014) compared the ELF to the GF method of Anderson (2007). In an ideal simulation where the true covariances are known, the ELF and GF show similar results, especially for larger ensembles. The ELF shows benefit over the GF in cases where there are biases in the spurious ensemble covariances. It also has the unique ability to detect underestimated ensemble spread and automatically inflate covariances. However, extension of the ELF to real atmospheric applications poses challenges. A large number of observation-state pairs are needed for accurate ELF localizations, especially for small localization values. Second, for real-data applications where there is no truth,

the method will likely require long assimilations for reliable localization estimations (Anderson and Lei 2013).

There are many localization methods in the literature, with very different applications. Until a method is proven to be superior, the choice of which adaptive method to use comes down to cost and case-by-case concerns. We have chosen to use a modified version of the group filter of Anderson (2007) for work in this dissertation evaluating an ensemble-based method for quantifying observation impact on forecasts. A discussion of this modification and application is given in chapter 4. In the next section, the ensemble-based impact technique is introduced.

2.3 Ensemble forecast sensitivity to observations (EFSO)

Following Kalnay et al. (2012), let $\overline{\mathbf{x}}_0^a$ represent the ensemble mean analysis and $\overline{\mathbf{x}}_{t|0}^f$ the deterministic forecast launched from the mean analysis (subscript "*t|0"* can be read as "valid at time *t*, initialized from analysis at time *0*"). A cost function, *J*, is defined in Langland and Baker (2004) to be the actual forecast error reduction – the difference in squared error between two adjacent forecasts,

$$J = \left(\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-n}^T \mathbf{e}_{t|-n}\right) = \left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}\right)^T \left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}\right) \tag{2.15}$$

where $\mathbf{e}_{t|0} = \overline{\mathbf{x}}_{t|0}^f - \mathbf{x}_t^{tr}$ and $\mathbf{e}_{t|-n} = \overline{\mathbf{x}}_{t|-n}^f - \mathbf{x}_t^{tr}$ are the forecast errors initialized from mean analyses at time $t = 0$ and $t = -n$, respectively. Here $\mathbf{x}_t^{tr}$ is the truth valid at time *t*; in the absence of the true state, a verifying analysis can be used. The actual forecast error reduction is shown schematically in Figure 2.2.

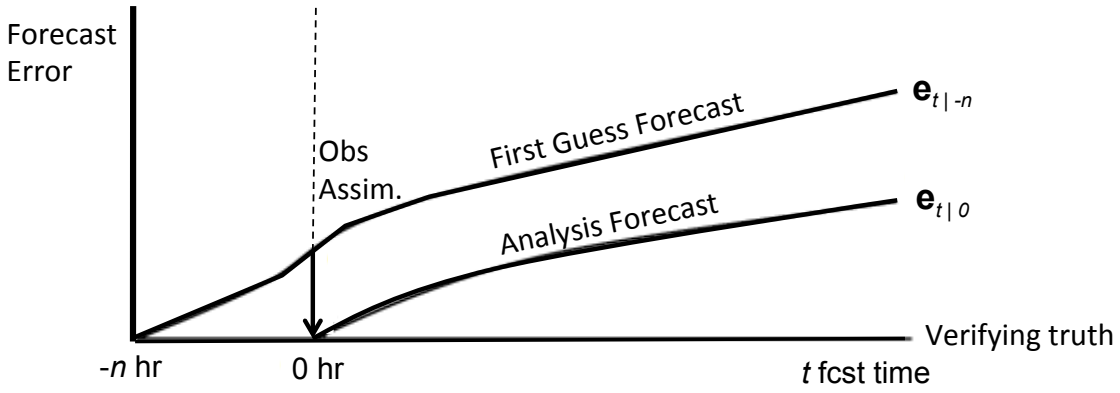

Figure 2.2. Schematic of actual forecast error reduction in a cycled DA system. Each line represents forecast error with time. Vertical line represents observations assimilated at time 0 using the first guess forecast as a background. Difference in error at time *t* is linked to assimilated observations at time 0.

The differences in forecast errors $\mathbf{e}_{t|0}$ and $\mathbf{e}_{t|-n}$ are due to the assimilation of observations at time *t = 0*, so (2.15) represents the impact of assimilating observations on a forecast. When *J* is negative (positive), the magnitude of error in $\mathbf{e}_{t|0}$ is less (greater) than the magnitude of error in $\mathbf{e}_{t|-n}$, which can be interpreted as *positive (negative) impact*.

It is shown in Kalnay et al. (2012) that (2.15) can be rewritten in ensemble form as

$$J \approx \frac{1}{K-1}\delta\mathbf{y}_0^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\mathbf{X}_{t|0}^{fT}\left(\mathbf{e}_{t|0}+\mathbf{e}_{t|-n}\right), \tag{2.16}$$

where $\delta\mathbf{y}_0 = \mathbf{y}_0 - H(\overline{\mathbf{x}}_0^b)$ is the observation innovation vector – the difference between the observations, $\mathbf{y}_0$, assimilated at time 0 and the mean background interpolated to observation space by forward operator *H( )*. **H** is the linearized forward observation operator, **R** is the observation error covariance matrix, *K* is

the ensemble size, and $\mathbf{X}_0^a$ and $\mathbf{X}_{t|0}^f$ are $n \times K$ analysis and forecast perturbation matrices, respectively *(n* = number of model state variables, *K* = number of ensemble members). Despite the use of the tangent linear model approximation to obtain (2.16) each column in $\mathbf{X}_{t|0}^f$ can be calculated using the full non-linear model *M( ),* such that the *i*th column is $M(\mathbf{x}_0^{a(i)}) - \overline{M(\mathbf{x}_0^a)}$. The expression in equation (2.16) is appealing as it can be applied using available assimilation products of any deterministic EnKF method. A full derivation of the EFSO metric can be found in Appendix A.

As with any method involving the use of ensemble to estimate covariances, covariance localization is needed to suppress the effects of sampling error from too small ensembles. The matrix product $\mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} = \mathbf{H}\mathbf{X}_0^a \mathbf{X}_{t|0}^{fT}$ is the ensemble estimate of model error covariance between the analysis in observation space and forecast valid at time *t*. Localization of (2.16) is applied to this $p \times n$ matrix (*p* = number of observations). Denoting localization matrix $\boldsymbol{\rho}_I$, the observation impact estimate modulated by the localization function becomes

$$J \approx \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} \left[\rho_I^T \circ \left(\mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} \right) \right] \left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n} \right). \tag{2.17}$$

The localization matrix $\boldsymbol{\rho}_I$ must be an $n \times p$ matrix, meaning that every grid point-observation pair can have a unique localization weight. Since localization in (2.17) is applied to $\mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT}$, in addition to spatial and cross-variable components, a level of complexity is added in the time-forecast component. In equation (2.17), there is no

requirement that ρ_I has to be the same localization as that used during EnKF assimilation, ρ_A.

The choice of ρ_I should attempt to take the time-forecast component into account, in addition to spatial and cross-variable components. To partially address the issue, Kalnay et al. (2012) proposed two methods of moving localization: (1) using a model-forecast nonlinear incremental evolution of the localization function, and (2) advecting the localization center using the climatological group velocity of dominant wavenumbers. Ota et al. (2013) applied a similar advected localization method, using the average forecast horizontal wind at each model vertical level. Both studies showed improvement relative to fixed GC localization; however, possible limitations exist with each method. The nonlinear evolution of the localization is computationally prohibitive for real NWP systems because a forecast is required for every observation. The advection methods are simpler to implement; however, they assume that the optimal localization is tied to the mean flow of the model and does not change in magnitude, size, or shape.

2.3.1 Studies applying the EFSO method

Kunii et al. (2012) successfully applied the original ensemble-based method by WRF model together with the LETKF. They found the ensemble method could capture the actual error reduction, though they noted a significant underestimation of the estimates. Generally all observation systems were found to have positive impacts, with upper level soundings providing the largest contribution. Ota et al. (2013) applied the EFSO to the NCEP GFS system for a one-

month period in early 2012, totaling 124 cases. The EFSO provided a useful online monitoring of impacts for different types, particularly to help identify observations that may not be optimally assimilated. They found that satellite radiance observations were the most important in reducing short-range forecast error, particularly in moisture. Other observations were found to be of high value as well, including aircraft, radiosonde, marine surface observations, and scatterometer winds.

Another application of the Ota et al. (2013) study was in the identification of detrimental observations that led to 24-hour forecast failures. Such a 'proactive quality control' (PQC) concept was further explored and developed by Hotta et al. (2017b). In PQC, the EFSO is used to identify detrimental observations on a 6-hour forecast. The analysis is then performed again withholding the detrimental observations. In a semi-operational context, Hotta et al. (2017b) found that applying such a procedure significantly reduced the number of forecast 'dropouts' – abrupt drops in 24-hour forecast skill. In 18 of 20 dropout cases, withholding detrimental observations identified by EFSO improved the 24-hour forecast error, with over 30% improvement in seven of the cases. This improvement was found to persist beyond 5 days in forecast. They were careful to note, however, that a negative impact does not necessarily indicate the observation is bad. The EFSO estimates impact from the observed-minus-background innovations associated with each observation. The source of error in a detrimental observation could thus be related to the observation, background, or errors related to the DA system (e.g. misspecification of observation error). The most obvious application of the PQC

method is for reanalysis purposes. The 6-h forecast requirement to identify detrimental observations makes it challenging to apply in an operational context, though the authors do discuss the potential for such an application and how some of the challenges could be addressed. A modification of the EFSO system is the Ensemble Forecast Sensitivity to Observation Error Covariance, or ESFR, developed by Hotta et al. (2017a) to help facilitate objective and systematic tuning of observation error covariances **R**.

The first study to apply the EFSO on a convective-scale forecast system was Sommer and Weissmann (2014). They applied the EFSO to the German-developed Kilometre-scale Ensemble Data Assimilation (KENDA) system together with the Consortium for Small- scale Modeling (COSMO) model, run at 2.8-km horizontal grid resolution. Using 0, 3, and 6-hour forecast valid times they found good agreement between the data denial impact and the EFSO estimates for different observation types. Additionally, the EFSO method was able to properly identify cases where an observation type is assimilated with suboptimal observation error. They also found little sensitivity in varying the localization length scale, though they did not take into account the time-forecast component. Later, Sommer and Weissmann (2016) modified the original EFSO method to verify against observations directly. This eliminates the need for a verification analysis and thus avoids the potential problem of the forecast correlating with the verification field.

Chapter 3: Assessing Impacts of the High-Frequency Assimilation of Surface Observations for the Forecast of Convection Initiation on 3 April 2014 within the Dallas-Fort Worth Testbed

3.1 Introduction

The 2009 NRC report noted that the highest priority of observational need is measurements of the planetary boundary layer (PBL), the layer directly influenced by contact with the earth's surface (National Research Council 2009). Many mesoscale features are prominently featured in the PBL including horizontal convective rolls, surface boundaries such as drylines and fronts, and other gradients and diurnal variations influenced by topography. While radar data provide crucial measurements of ongoing precipitating systems, the current network of S-band radars overshoots a large part of the PBL. Additionally radar data do not directly measure temperature or moisture, so important features that may trigger or maintain convection under favorable conditions are missed, such as areas of enhanced moisture convergence or location and strength of cold pools. For these reasons, recent high-resolution DA studies have placed increasing emphasis on the assimilation of mesonet surface observations in addition to radar data (e.g., Carlaw et al. 2015; Johnson et al. 2015; Snook et al. 2015; Chen et al. 2016). Despite positive results, surface observations remain an operationally underutilized dataset, partly due to mismatches between coarse model terrain and actual observation heights, particularly in areas of high varying terrain (Pu et al. 2013), and other concerns about siting and instrument quality.

One fundamental limitation of radar data is the inability to measure pre-convective environments and the different features that lead to convection

initiation (CI). Given a favorable large-scale environment, variations on the order of 1 g kg^{-1} of specific humidity have a large influence on the existence and location of storm initiation, owing to small-scale features in the PBL such as dryline bulges and convective rolls (Weckwerth 2000; Weckwerth and Parsons 2006). The prediction of CI can be potentially improved with the utilization of mesonet surface observations. Xue and Martin (2006a) and Liu and Xue (2008) explored the prediction of CI of two separate cases as part of the International H2O Project (IHOP; Weckwerth et al. 2004). They performed hourly assimilation of mesonet observations in addition to special IHOP upper-air and surface observations. Xue and Martin (2006a) found moderate improvements in the CI forecast; however, Liu and Xue (2008) found mixed results, showing that reducing the assimilation to 3-hourly actually improved the forecast in some ways. They suggested that too frequent surface DA may weaken the surface forcing responsible for CI, leading to a worse forecast.

Sobash and Stensrud (2015; hereafter SS15) was the first study to examine sub-hourly assimilation of mesonet observations for the prediction of CI. They assimilated mesonet observations every 5 minutes for 1, 2, and 3 hours prior to CI using an EnKF system. Results showed that high-frequency assimilation of mesonet observations led to improvements in timing and placement of CI as compared to hourly assimilation experiments, with 3-hours of 5-min. cycling performing best. These improvements persisted throughout their 3-hour free forecast period owing to better representation of the surface moisture field and dryline strength. SS15 concluded that mesonet observations provided important

information on the diurnal progression of the dryline mixing eastward, including important small-scale variability such as dryline surges, that helped to provide more accurate initial conditions (IC) and a more accurate short-term forecast. Furthermore, frequent assimilation of mesonet observations helped to constrain the positive moisture bias of the PBL scheme used in their study. They note, however, that frequent mesonet DA may not be as impactful in other cases where model biases are minimal or in cases where CI is driven by more large-scale forcing rather than surface features.

Although SS15 was an important foundation in showing the benefit of frequent DA of surface mesonet observations, that work was focused on the use of mesonet data as a whole, with the predominant impacts coming over areas well-covered by Oklahoma mesonet (McPherson et al. 2007) and West Texas mesonet (Schroeder et al. 2005) observations. Unfortunately, many areas such as the DFW metroplex do not have a federal or state sanctioned and maintained mesonet network available; additionally, although the number of mesonet networks is increasing in different states and regions across the US, budget constraints make a national implementation of such a network unfeasible. In order to have the NNoN vision succeed, then for cases where there are no mesonet networks we must leverage the use of nonconventional surface observations from other sources, such as the CWOP and ERNET data within the DFW testbed. In this work, we explore the use of nonconventional surface observing networks within the DFW area for the prediction of a CI event. In contrast to SS15, we will examine impacts from

different nonconventional networks individually to assess benefits and limitations from using the different systems.

The first OSE study using nonconventional data within the DFW testbed was done by Carlaw et al. (2015). They tested the utility of nonconventional surface data, specifically from CWOP and ERNET, on the forecast of a tornadic supercell that caused EF3 damage near Cleburne, TX on 15 May 2013. Results showed that these nonconventional surface observations helped improve the surface analysis of low-level thermodynamic fields in an otherwise data sparse region. These observations caused enhanced instability ahead of the storm and led to a stronger updraft and vertical vorticity within the storm, which better matched observations. Though not the first case study using these nonconventional networks, this study is different from Carlaw et al. (2015) in that we are using these data to predict storm initiation rather than as a supplement to radar data for an ongoing storm.

In section 3.2, we summarize the severe weather event used for this study, 3 April 2014, and further detail the different observation sources available within the DFW testbed for this case. Section 3.3 describes the experiment setup, including model and DA configurations, as well as descriptions of the data denial experiments. In section 3.4, the results of the data denial experiments are discussed, with a diagnosis of impacts tied to observations that led to forecast differences in ensemble CI performance. In section 3.5, the case study is summarized with implications for potential of utilizing these nonconventional observations in future DA systems.

3.2 3 April 2014 case overview

3.2.1 Summary, synoptic setup, and storm evolution

According to the National Center for Environmental Information (NCEI; 2014), severe storms in the Dallas area on April 3, 2014 caused an estimated $500 million damage to 35,000 vehicles and 22,000 homes, mainly in Denton County, Texas. Reports at the Storm Prediction Center (SPC) included all types of severe weather across N. Texas on that day, including numerous hail reports with the largest up to softball size (108 mm or 4.25 in. diameter), a high wind report of 41 m s^{-1} (82 kt), and three confirmed tornadoes NE of the DFW area causing EF0 and EF1 damage (Figure 3.1). Outside of Texas there were numerous severe reports, altogether totaling 288 filtered reports across the United States – including 16 tornado reports.

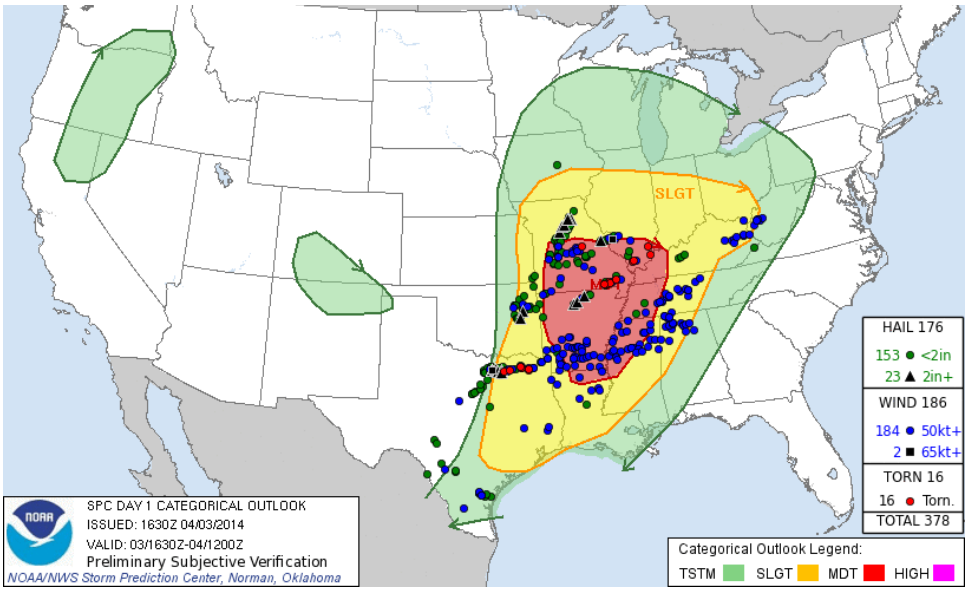

Figure 3.1. SPC preliminary storm reports valid from 1200Z on 4/3/17 to 1200Z on 4/4/17, overlaid on the day 1 categorical outlook issued at 1630 UTC on 04/03/17. Markers indicate tornado reports (red dots), significant hail reports greater than 2 in. diameter (black triangles), other severe hail reports (green dots), significant wind reports above 65 kts (black squares), and other severe wind reports (blue dots).

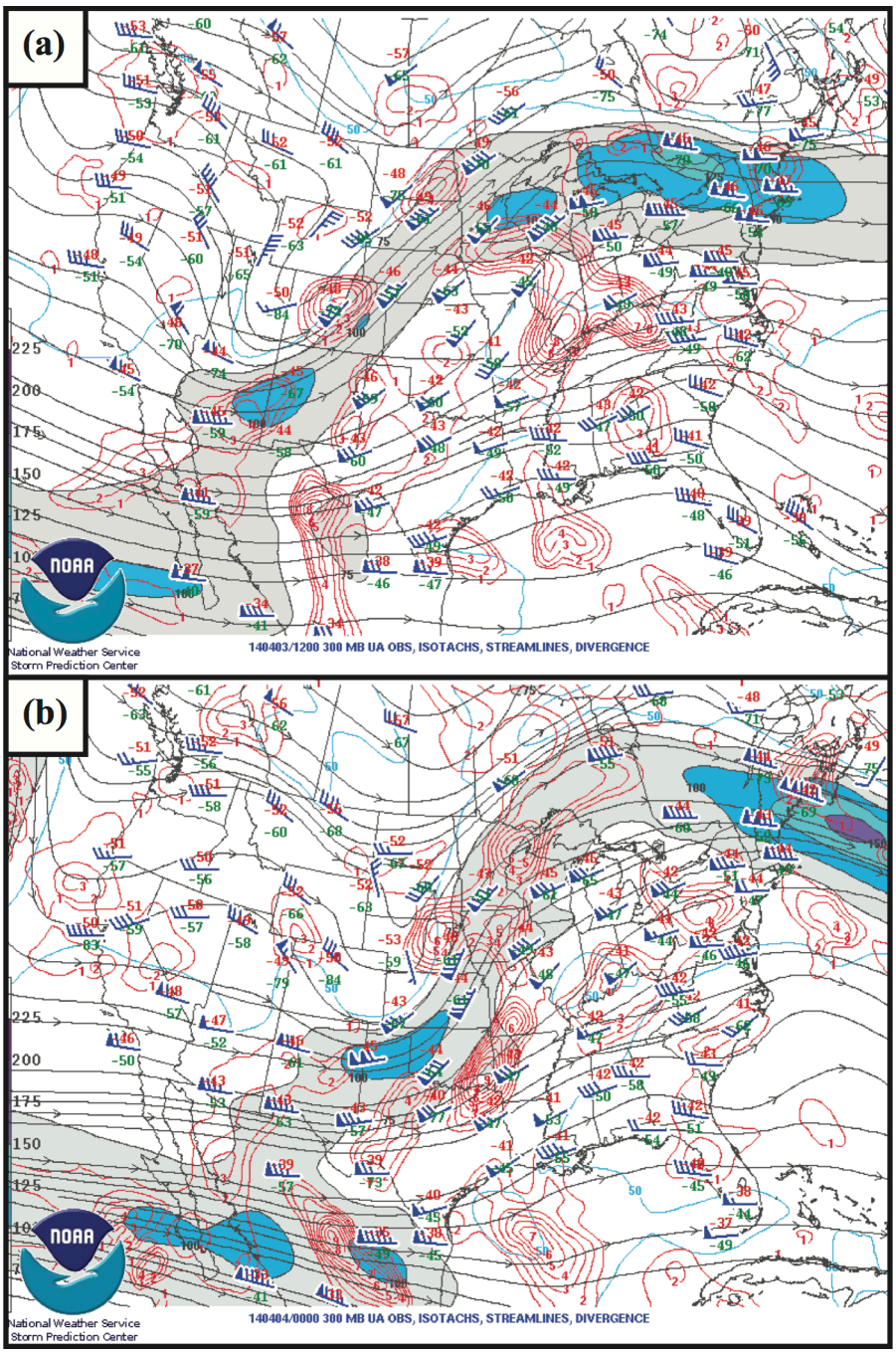

Figure 3.2. Upper-air 300-hPa observations, isotachs (blue and color-filled contours), streamlines (black contours), and divergence (red contours) valid at (a) 1200 UTC 3 April 2014 and (b) 0000 UTC 4 April 2014.

On the synoptic scale, a high-amplitude, positively tilted short-wave trough was located over the central and southern Rockies region at 12Z on 3 April 2014 (Figure 3.2a). As the trough progressed eastward it became more neutral to slightly negatively tilted, with a region of upper-level divergence extending from NE Texas through Oklahoma, Arkansas, and Missouri by 00 UTC 4 April 2014 (Figure 3.2b). In response the associated surface low, located in southern Kansas at 12 UTC on April 3, deepened and moved towards NE Missouri by 00 UTC on April 4. A cold front extended to the SW across NE and central Oklahoma through the Texas panhandle, with a dryline intersecting it near central OK and extending SW through Texas, west of the DFW area (Figure 3.3).

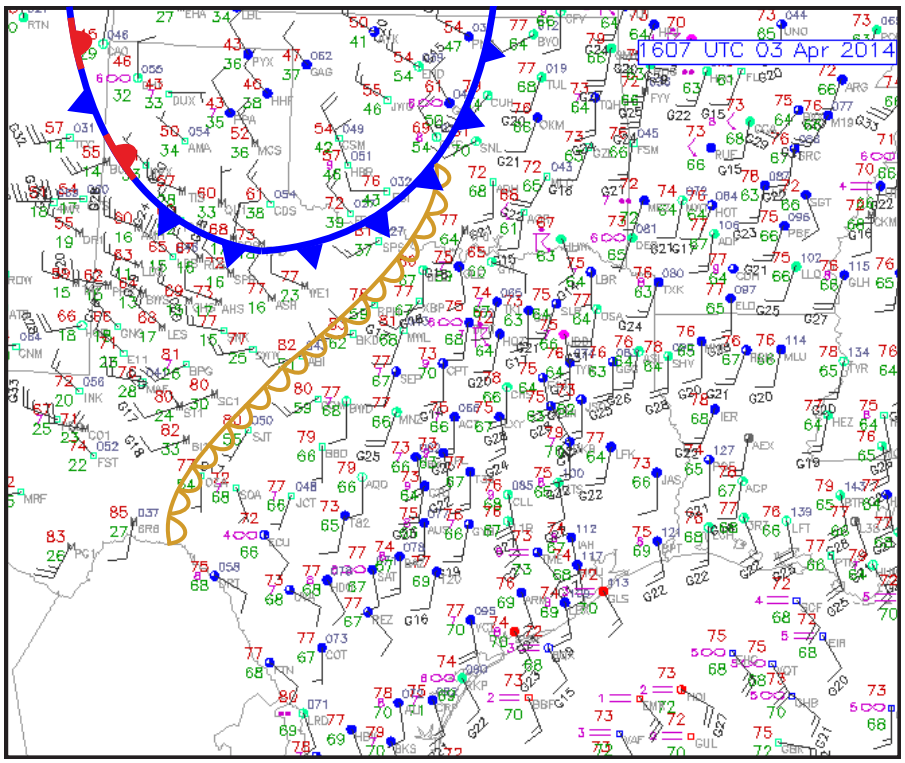

Figure 3.3. Surface map of observations with approximate locations of cold front (blue line with triangles), stationary front (mixture of blue-triangles and red-scallops), and dryline (brown scalloped line), valid 1607 UTC 3 April 2014.

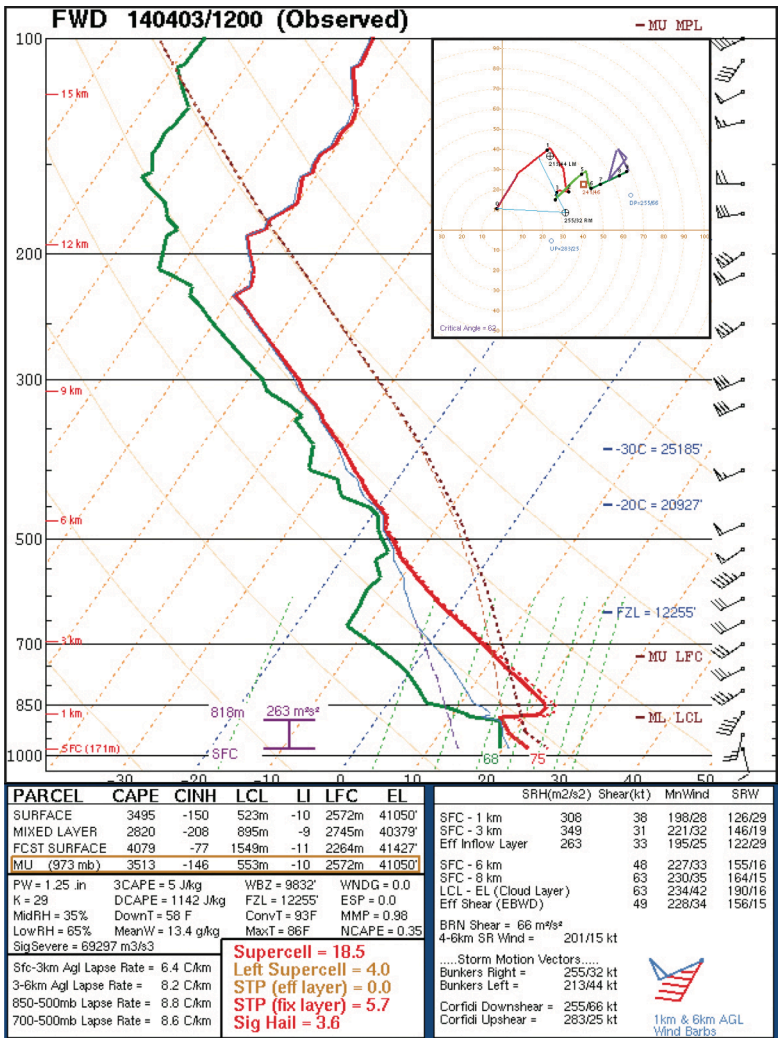

Figure 3.4. Fort Worth, TX (FWD) sounding and hodograph (upper right corner) valid 1200 Z 3 April 2014, taken from the SPC Experimental Sounding Analysis System which uses sounding analysis program NSHARP (Thompson and Hart 2017).

As the afternoon progressed the dryline propagated eastward towards the western edge of the DFW domain. The Fort Worth, TX radiosonde launched at 12 UTC on April 3rd was representative of the potential instability available ahead of the dryline near Dallas, with convective available potential energy (CAPE) ranging from 2000 to 4000 J kg^{-1} (Figure 3.4). The high values of CAPE are attributable to steep mid level lapse rates exceeding 8 ºC km^{-1} with a very moist boundary layer

including surface dewpoints 18-21°C. A substantial capping inversion existed, inhibiting convection from developing before 1800 UTC. As the afternoon progressed, the cloud layer just east of the dryline near Dallas cleared and moved east (Figure 3.3), allowing for the solar insolation necessary to erode the capping inversion and result in conditions favorable for CI. The 0-6km deep layer shear is supportive of supercells with sufficient low-level shear to allow for tornadoes; however, the veer-back-veer wind profile in the hodograph was not conducive for strong, long-track tornadoes (Figure 3.4).

The first CI event in north Texas appears at 1820 UTC in SW Wise County, TX (Figure 3.5a). As this storm grows and moves NE, a second storm initiates at approximately 1900 UTC in NW Parker County, TX, approximately 25 km to the SW of the first CI event (Figure 3.5b). Both storms develop into strong hail-producing supercells, though they remain nontornadic as they travel through the north side of the DFW testbed (Figure 3.5c). The second storm tracks through Denton, TX around 2030 UTC, producing very large hail of up to 100 mm around 2045 UTC (Figure 3.5d). Other areas of CI occur outside of the DFW area, including a cell along the Northern border of Comanche County, TX around 2000 UTC (Figure 3.5c) and additional locations S and SW of there between 2100 and 2200 UTC, though these storms are generally not as well-organized as the DFW supercells (Figure 3.5e). Additional storm development occurs after 2200 UTC and progresses eastward, eventually producing the EF0 and EF1 tornadoes NE of the DFW area.

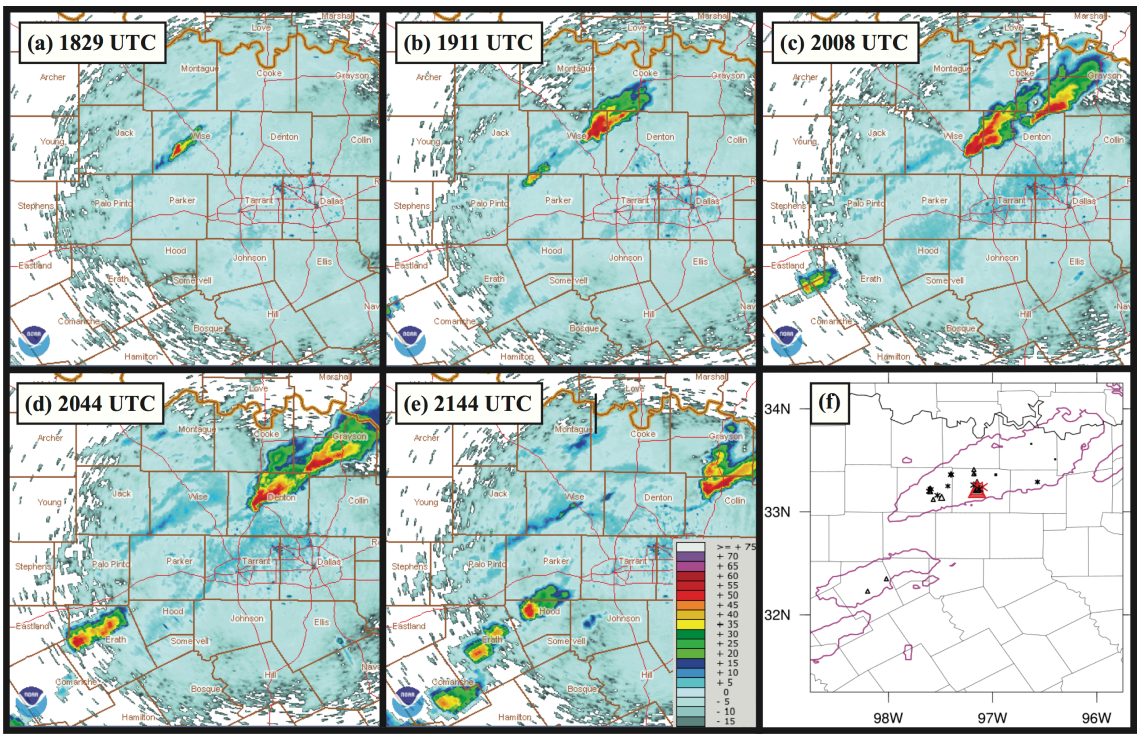

Figure 3.5. (a-e) Observed digital hybrid reflectivity from terminal Doppler radar TDAL, valid at 1829, 1911, 2008, 2044, and 2144 UTC, respectively. (f) SPC severe hail reports (triangles) and MPING hail reports (asterisks) between 1800 and 2200 UTC. Red markers indicate significant hail reports greater than 2 in, and magenta lines indicate the 30-dBZ contour of maximum observed composite reflectivity over the 1800-2200 UTC time period.

Hail report locations and relative sizes are shown for the 1800 – 2200 UTC period on 3 April 2014 in (Figure 3.5f). These reports were taken from the SPC as well as the Meteorological Phenomena Identification Near the Ground (mPING; Elmore et al. 2014) project, run by the National Severe Storms Laboratory (NSSL); mPING observations use crowd sourcing to provide observations of precipitation type across the U.S. The largest hail occurs right over the city of Denton, TX, and hail is reported from E and NE Wise County through Denton County and NE to Collin and Grayson Counties.

3.2.2 Description of observations used

The DFW Testbed is a medley of many observational platforms, summarized in Table 3.1. Observing systems are separated into two categories: conventional and nonconventional. Here, conventional data refers to observations that are already typically used in operational data assimilation systems. They include upper air soundings launched twice daily, wind profilers, aircraft data from the Aircraft Communications Addressing and Reporting System (ACARS; Moninger et al. 2003), surface data from Automated Surface Observing System (ASOS) and Automated Weather Observing System (AWOS), and integrated water vapor data from Global Positioning Satellite Precipitable Water Vapor (GPS PWV) observations. In this study we also consider the state-sanctioned Oklahoma and West Texas mesonets to be conventional datasets whose quality is similar to that of ASOS and AWOS. Additionally, radar data from the Next Generation Radar (NEXRAD) network of WSR-88D S-band radars is used as a conventional data source. Nonconventional data are all the other sources that are not yet fully implemented in many systems. For radar, these include the low-cost CASA X-band radars deployed in the DFW Testbed as well as C-band Terminal Doppler Weather Radars (TDWRs) available at major airports. Because this study focuses on forecasts of CI, the CASA and TDWRs are not used during assimilation. Additionally, the radiometers available for this case were not assimilated because here we are focusing on the impacts of surface observing networks.

Table 3.1. All observing platforms available within the DFW testbed as of 2017. Used column represents whether that observation platform was used in this study (Y) or not (N); n/a means "not available for this study". The last column specifies whether the platform is considered a "conventional" or "nonconventional" dataset

Observation	Type	Provider	Used ?	Conv. (C) or Nonconv. (NC)?
NEXRAD	S-band radar	NWS	Y	C
TDWR	C-band radar	FAA	N	NC
CASA	X-band radar	CASA	N	NC
MDCRS/ACARS	Aircraft	Airlines via MDCRS/ACARS	Y	C
Radiosonde	Weather Balloon	NWS	Y	C
NOAA Profiler Network	Wind Profiler	NOAA	Y	C
GPS PWV	Precipitable Water Vapor	Suominet	Y	C
ASOS/AWOS	Surface	NWS & FAA	Y	C
Oklahoma Mesonet	Surface	Oklahoma Climatological Survey	Y	C
West Texas Mesonet	Surface	Texas Tech University	Y	C
CWOP	Surface	NWS	Y	NC
ERNET	Surface	Earth Networks	Y	NC
Understory	Surface & Hail	Understory	N (n/a)	NC
Other Mesonet	Surface	Misc. Federal/State Agencies	Y	NC
MoPED	Surface	GST	Y	NC
SODAR	Wind Profile	WeatherFlow	N (n/a)	NC
Radiometer	Thermodynamic Profile	Radiometrics via Earth Networks	N	NC

Nonconventional surface observations include CWOP, ERNET, and MoPED. CWOP data come from a volunteer group of over 10,000 amateur radio operators who transmit weather data using the Automatic Position Reporting System as a Weather Network (APRSWXNET; Chadwick 2014). These observations are found

near volunteers' homes, backyards, farms, and businesses. They are ingested into the Meteorological Assimilation Data Ingest System (MADIS) and subject to the same quality-control checks as other datasets within MADIS (Miller et al. 2005). ERNET is a network of over 8000 surface observations across the country, initially installed on or near public schools and buildings for real-time weather data to be used during television broadcasts. An additional benefit is the ability for the schools with ERNET stations to use them within their science and technology curriculum. These data are also ingested via MADIS and have been made available for experiments with the DFW testbed. Both CWOP and ERNET were installed without enforcing siting standards that ASOS and AWOS employ. As such, they may exhibit additional representativeness errors and biases, such as the low wind speed bias noted in Carlaw et al. (2015). MoPED data are weather sensors mounted on trucking fleets involved in interstate transportation. Global Science and Technology (GST) started the program in 2009 in partnership with NOAA and the National Mesonet Program (Dahlia 2013). These sensors provide measurements of pressure, temperature, humidity, and precipitation; however, wind measurements cannot be obtained due to contamination from vehicle motion at high speeds. There are a few other nonconventional mesonet-type observations within the MADIS dataset that are separate from any previously mentioned. These observations are still used during the assimilation procedure described in the next section; however, they are not a primary focus of this study, so they have been separated into a miscellaneous nonconventional category.

3.3 Experiment setup

3.3.1 Model configurations and initial ensemble

The model chosen for this study is Weather Research and Forecast (WRF) model, specifically version 3.7 of the Advanced Research WRF (ARW) core (Skamarock and Klemp 2008). The ensemble used here was initialized from a combination of ensemble members from other systems, shown in Table 3.2, chosen such that each ensemble member had its own unique lateral boundary conditions (LBCs) from their respective model perturbations. The ensemble initial condition (IC) was provided by interpolating each ensemble member at 03 UTC on April 3, 2014 to a 12-km horizontal grid with 250 × 180 grid points, centered over N Oklahoma (Figure 3.6). This intermediate outer grid was chosen to facilitate smooth transition from the coarse ICs to a convection allowing model (CAM) grid resolution. The inner grid, shown in Figure 3.6, has 2.4-km horizontal grid spacing with 351 × 351 grid points, centered over the DFW domain. This grid is initialized at 15 UTC on April 3, 2014 using two-way nesting within the WRF. Prior to any DA on the inner grid, an hour of model integration was performed to allow the model to spinup the small-scale processes that can be resolved on the 2.4-km grid (Figure 3.7). Further details on the assimilation procedure can be found in the next two sections.

Table 3.2. WRF model settings and physics parameterization schemes

WRF setting	Option Chosen
Number of Ensembles	43
Initial Conditions	SREF (21 members); GEFS (21 members); NAM (1 member) at 03 UTC 3 April 2014
Horizontal grid (outer domain)	250 × 180, Δx = 12 km
Horizontal grid (inner domain)	351 × 351, Δx = 2.4 km
Vertical grid	50 levels, p_{top} = 50 hPa
Cumulus scheme	Outer: mixed – Kain-Fristch (Kain 2004) Betts-Miller-Janjic (Janjic 1994) Grell-Freitas Ensemble (Grell and Freitas 2014) Grell 3D Ensemble (Grell and De´ve´nyi 2002) Inner: none
PBL scheme	Mellor–Yamada Nakanishi Niino (MYNN) Level 2.5 (Nakanishi and Niino 2009)
Microphysics scheme	Thompson (Thompson et al. 2008)
LW radiation scheme	Rapid Radiative Transfer Model for global climate models (RRTMG; Iacono et al. 2008)
SW radiation scheme	New Goddard (Chou and Suarez 1999)
Land surface scheme	Noah (Tewari et al. 2004)

Table 3.2 summarizes the different model settings for the WRF, including all parameterization schemes used for the outer and inner grids. Cumulus parameterization is only done on the outer grid, with four schemes evenly mixed throughout the ensemble, similar to that done in Johnson and Wang (2017). The MYNN PBL scheme was chosen in part based on results from Coniglio et al. (2013), who found that the MYNN scheme is almost unbiased in PBL depth, moisture, and potential temperature when verified against sounding observations. The MYNN

scheme was unique among local schemes, as others tended to produce too shallow and too moist PBLs. The Thompson microphysics scheme in combination with the MYNN PBL scheme produced forecasts with storm mode most closely matching reality for the supercells in the DFW domain in the case study here.

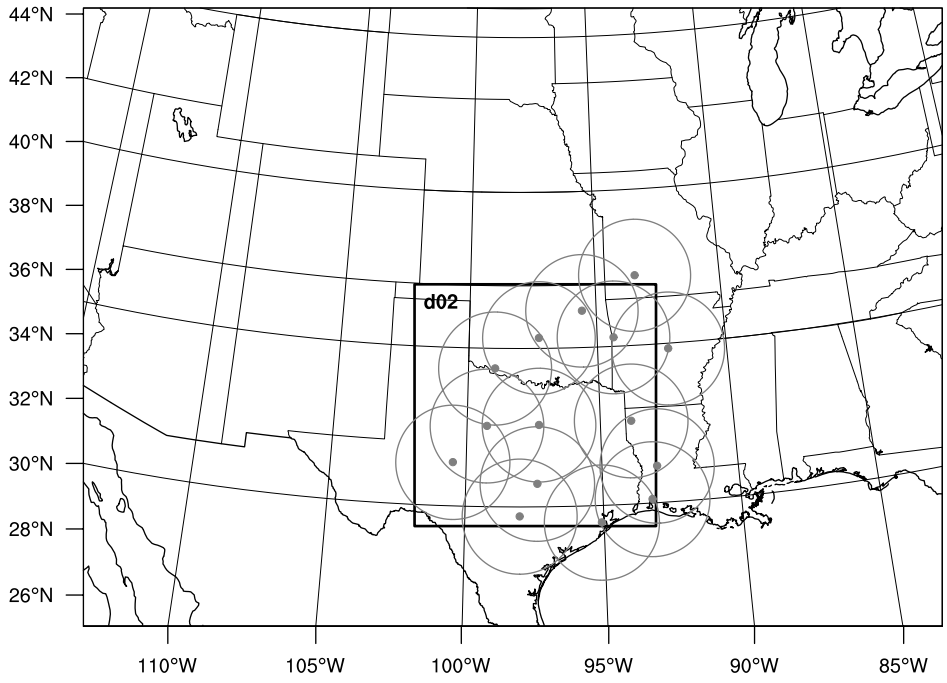

Figure 3.6. WRF model two-way nested grid setup. Outer domain has 12-km horizontal resolution, and inner grid has 2.4-km horizontal resolution (5:1 ratio). Dots indicate locations of WSR-88D radars used during inner cycle DA, with circles representing 200-km ranges for each radar.

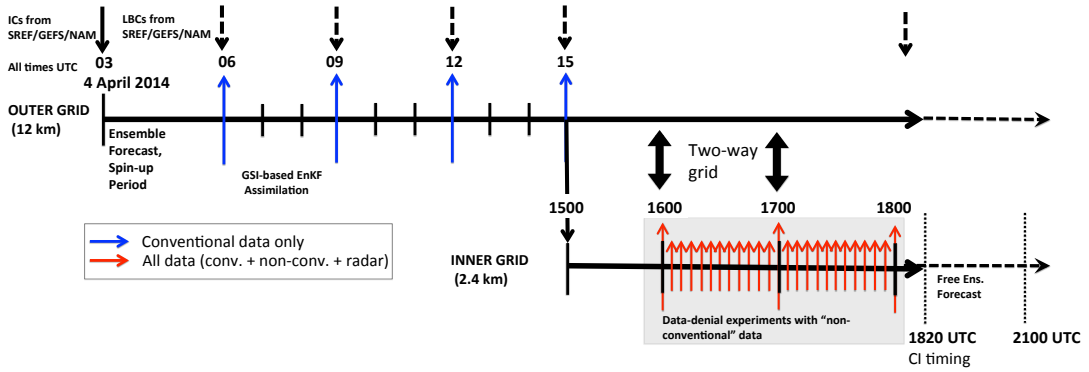

Figure 3.7. Cycled DA experiment diagram. On the outer grid, after 3-hour spinup four 3-hourly DA cycles of conventional observations were performed, inner grid initialized at 1500 UTC. After 1-hour spinup, 5-min. DA cycling of all observations was performed for 2 hours (1600 – 1800 UTC). Free ensemble forecast initialized at 1800 UTC and run for 3 hours.

3.3.2 Data assimilation settings

The DA system chosen for this study is the Gridpoint Statistical Interpolation analysis system (GSI) based EnKF extended to directly assimilate radar observations (Johnson et al. 2015). The GSI-based EnKF uses the ensemble square root filter (EnSRF) introduced by Whitaker and Hamill (2002). On the outer domain, four 3-hourly cycles were performed at 6, 9, 12, and 15 UTC using all available conventional observations except radar to better simulate the mesoscale environment prior to assimilating small-scale features at CAM resolution (Figure 3.7). For these analyses, first guess at appropriate time (FGAT) was used with output model first-guess fields every 30-min until the end of the DA cycling window (1.5 h after analysis time). On the inner grid, 5-min cycling is used to capture rapidly-evolving storm-scale structures within the DA system. In addition to conventional observations, nonconventional surface observations and NEXRAD radar radial velocity and reflectivity are also assimilated on the inner grid.

The purpose of the radar DA is to capture existing convection within the 1600-1800 UTC time period, as well as to eliminate any spurious convection that develops within the model. To accomplish this, both precipitation and clear-air reflectivity observations are assimilated from the 15 radars shown in Figure 3.6, including assimilating missing observations as clear-air. Past studies have shown assimilating clear-air reflectivity to be effective in suppressing spurious cells (e.g., Tong and Xue 2005; Aksoy et al. 2009). Here, we consider the precipitation threshold to be 15 dBZ; that is, all observed reflectivity ≥ 15 dBZ were assimilated as precipitation observations. Any missing observation or observation less than 5 dBZ was set to 0 dBZ, the clear air value. Additionally, any first guess reflectivity value less than 0 dBZ was reset to 0 dBZ, to ensure clear-air reflectivity could only have the effect of precipitation suppression. Note that there is a gap between 5 and 15 dBZ where the reflectivity observations were ignored (not assimilated) to prevent overlap in assimilation of precipitation and clear-air observations near the edges of precipitation.

One challenge for ensemble assimilation methods is how to treat error sources within the system. The first error source is sampling error owing to use of a relatively small ensemble compared to the large number of degrees of freedom within the model, causing spurious covariances. Fortunately, covariance localization can be used to ameliorate sampling error (Houtekamer and Mitchell 1998, 2001). This study employs the commonly-used Gaspari and Cohn (1999) function, which approximates a Gaussian function but implements an explicit cutoff radius. For the outer grid, a cutoff radius of 700-km was used horizontally,

the same as in Johnson et al. (2015). On the inner grid, horizontal localization is varied dependent on observation type and network density, summarized in the last column of Table 3.3. All conventional observations save radar use a localization scale of 200-km on the inner grid. For radar, the 20-km cutoff is the same as used in Johnson et al. (2015), and similar to the 18-km radius that worked best in Sobash and Stensrud (2013). For surface mesonet observations two different scales are used. For Oklahoma and West Texas mesonets we chose 80 km to allow some overlapping of localization functions based on their average station separation of approximately 35 km. Nonconventional surface observations such as CWOP and ERNET have much denser networks, particularly near major cities, thus for each of these categories a 40-km localization radius was chosen. In SS15, a 60-km localization radius was used for all mesonet observations, so the values of 40-km and 80-km are similar but chosen to more accurately reflect the different surface mesonet networks rather than all of them at once. Vertical localization for all observations was chosen to be 0.55 in natural log pressure coordinates.

Another error source that must be accounted for is intrinsic model error that is not represented within the ensemble. Covariance inflation methods can be used to correct for model error. Two methods of covariance inflation are used in this work, following Johnson et al. (2015). The first is height-dependent multiplicative inflation, to account for model errors that cannot be represented within the ensemble (Whitaker and Hamill 2012). The amount of multiplicative inflation at the surface is set to 15% every 3 h on the outer domain, and 0.3% for each 5-min cycle on the inner domain. The multiplicative inflation smoothly tapers

to about 9% at 200 mb and 3% at the 50-mb model top to avoid excess spread, as in Johnson et al. (2015). The second inflation method is called relaxation to prior spread (RTPS; Whitaker and Hamill 2012). RTPS inflates posterior ensemble spread to some fraction, α, of the prior ensemble spread, accounting for excessive spread reduction in regions where there are many observations assimilated. Here we chose a value of 0.95 for α, meaning the posterior ensemble spread is inflated to 95% of the prior ensemble spread. RTPS is applied for both outer and inner domain DA.

Table 3.3. Initial observation error table values used for surface and radar datasets. Last column indicates localization cutoff scale used for each data source.

Data Source	P_s (hPa)	T (ºC)	RH (%)	uv (m s^{-1})	Z (dBZ)	V_R (m s^{-1})	Loc. (km)
NEXRAD	--	--	--	--	5.0	2.0	20
ASOS/ AWOS	0.5426	0.8	2.705	1.5	--	--	200
OK/WTX Mesonet	0.75	1.0	3.5	1.5	--	--	80
ERNET	1.125	1.5	5.0	2.0	--	--	40
CWOP	1.5	2.0	7.0	2.5	--	--	40
MoPED	1.125	1.5	5.0	--	--	--	40
Misc. Mesonet	1.125	1.5	5.0	2.0	--	--	40

3.3.3 Observation processing and quality control

All conventional observations minus radar were obtained from the NAM Data Assimilation System (NDAS) in prepbufr format. These observations were put through various automated quality control (QC) checks within the prepbufr processing (Keyser 2017). Nonconventional surface observations were obtained through MADIS and subjected to the same QC checks within the MADIS system, which uses several levels of checks including temporal and spatial consistency. As in Johnson et al. (2015), radar observations were processed using the Warning Decision Support System–Integrated Information (WDSSII; Lakshmanan et al. 2007) software. The GSI system itself performs additional QC gross error checks based on the ratio of the observation innovation to the observation error for a given observation.

An important consideration for this work is how to specify observation error values. Carlaw et al. (2015) showed that misspecification of error, in their case specifying CWOP data as accurate as ASOS data, led to degradation in verification scores. Values for initial observation error are shown in Table 3.3. They were based on similar values used in Carlaw et al. (2015) but further tuned for the GSI system. The basic idea, though, is that we consider ASOS to have the least amount of error, conventional mesonets to be nearly as accurate, and nonconventional data to be less accurate, with CWOP data being the least trusted due to the amateur nature of the observations. These values are initially smaller than what was used in Carlaw et al. (2015) because the GSI automatically adjusts (inflates) observation errors based on mismatches between observation pressure

and the first guess pressure, including assessing a higher penalty if the observation is found below the model terrain height. Since nonconventional data have enhanced siting concerns, the degree of observation error inflation is generally higher than in any of the conventional datasets. Other conventional observations for both inner and outer domains used the default observation error table values found within the GSI software.

3.3.4 Data denial experiment setup

Data denial experiments were done within the 5-min DA cycling from 1600-1800 UTC on the inner grid (Figure 3.7). The locations of all observations assimilated within this period are shown in Figure 3.8. Surface data are abundant within the DFW metroplex, but near the dryline the observations are relatively sparse, though there is some coverage of nonconventional observations (Figure 3.8b). The temporal interval of surface observations varies by dataset; ASOS and AWOS were available typically every 20 min, Oklahoma and West Texas mesonets every 5 min, and ERNET every 5 min with a few exceptions. CWOP data were much more intermittent; while most locations had observations available every 5 minutes, some stations were available every 10-15 minutes, with a few remote locations reporting even less frequently. Additionally, the miscellaneous mesonet data have a wide variety of reporting intervals, from 15 minutes to as infrequent as once per hour.

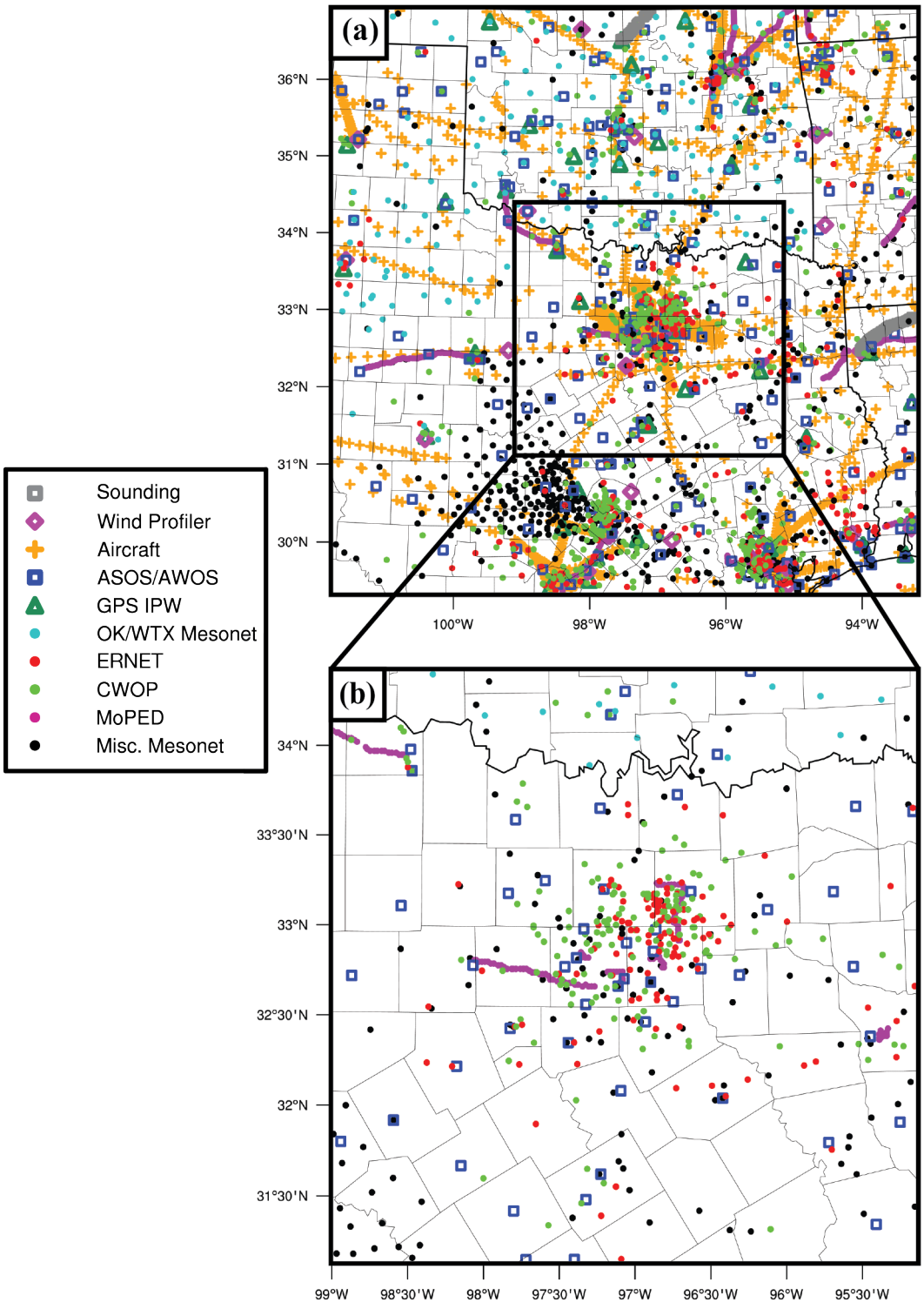

Figure 3.8. (a) Locations of all observations assimilated on inner domain between 1600 and 1800 UTC. (b) All surface observations available from 1600-1800 UTC, zoomed into the DFW region.

All data denial experiments are listed in Table 3.4. The control experiment, CNTL, uses all available data. Experiment NOSFC is a baseline experiment to show the effects of not having any surface data DA on the inner domain for the CI forecast, while NONEWSFC denies just the nonconventional surface observations in DFW. The other denial experiments will be compared with CNTL to evaluate the relative impacts of each dataset. The experiment denying ASOS data, denyASOS, is presented as another point of comparison with the other nonconventional denials. Finally, given past results of Carlaw et al. (2015) that showed the majority of positive benefit coming from thermodynamic variables, an experiment denying just wind information from CWOP and ERNET (denyCWwnd) is also considered to separate out the relative impacts from thermodynamic and wind observations.

Table 3.4. Summary of surface data denial experiments.

Experiment Name	ASOS/ AWOS	OK/WTX Mesonet	ERNET	CWOP	MoPED	Misc. Mesonet
CNTL	Yes	Yes	Yes	Yes	Yes	Yes
NOSFC	No	No	No	No	No	No
NONEWSFC	Yes	Yes	No	No	No	No
denyASOS	No	Yes	Yes	Yes	Yes	Yes
denyERNET	Yes	Yes	No	Yes	Yes	Yes
denyCWOP	Yes	Yes	Yes	No	Yes	Yes
denyCW	Yes	Yes	No	No	Yes	Yes
denyMISC	Yes	Yes	Yes	Yes	Yes	No
denyCWwnd	Yes	Yes	Yes (thermo)	Yes (thermo)	Yes	Yes

3.4 Results

There are a few methods of evaluating the experiment forecast output for CI. First, we use a subjective evaluation of ensemble probability of composite reflectivity (CREF) exceeding 35 dBZ for various time intervals. This threshold is common for identification of active ongoing convection (e.g., Roberts and Rutledge 2003). The focus for evaluating ensemble probability is on the first hour of the free ensemble forecast, chosen since the actual CI occurs around or just before 1820 UTC. Given that this is an extreme hail event for the DFW testbed, another metric to examine is the maximum hail diameter produced within the model throughout the 3-h free forecast. Given the even larger uncertainty involved with hail prediction, we employ a neighborhood ensemble probability (NEP) method, where for each ensemble member the fraction of points exceeding the threshold value within a given neighborhood radius is calculated at all model grid points, then averaged across all ensembles (Schwartz et al. 2010). In hail two thresholds are considered – 10 mm and 25 mm. The former is used as the lower limit of hail, slightly larger than pea size; the latter is approximately the threshold for severe hail used by the US National Weather Service. The NEP output will be compared with SPC and mPING report locations gathered between 1800 and 2100 UTC. The experiments are then further analyzed by tying the differences in ensemble probabilities to differences seen in surface fields, as well as comparison of root mean square differences (RMSD) and bias statistics among experiments.

3.4.1 Evaluation of CNTL experiment

The overall view of the severe event on the inner domain is shown in Figure 3.9,where observations at 1930 and 2030 UTC are compared to forecast probabilities of 35-dBZ composite reflectivity. Aside from the two supercells in the DFW region, another area of CI is observed in NE Oklahoma along the advancing cold front between 1800 and 1900 UTC (Figure 3.9a,b). This line of storms stretches SW into central Oklahoma. Another area of CI occurs in SW Texas between 1900 and 1930 UTC. Throughout the 3-h free ensemble forecast, CNTL captures the Oklahoma line of storms with very high confidence, as probabilities are shown exceeding 90% for a large portion of the line and as high as 100% for an early part of the line (Figure 3.9c). There is a bit less confidence in the southwestern extent of the line, but still over half the ensemble members indicate convection extending that far south. The storms of interest near Dallas are also captured very well in this 3-h view, with probabilities exceeding 90%. Convection also initiates to the SW in the ensemble and appears overaggressive compared to reality. There are additional points of observed CI in this area after 2100 UTC (not shown), so we consider this area forecast to be early CI rather than completely spurious.

Within the DFW region, the ensemble shows CI activity very quickly after the IC time of 1800 UTC, as shown in the top row of Figure 3.10. Already 10-20 minutes into the forecast, probabilities of CREF ≥ 35 dBZ as high as 90% are seen in SW Wise County, centered less than 10-km from the actual CI location (Figure 3.10a). Probabilities remain above 70% by 1850 UTC with a general collocation of

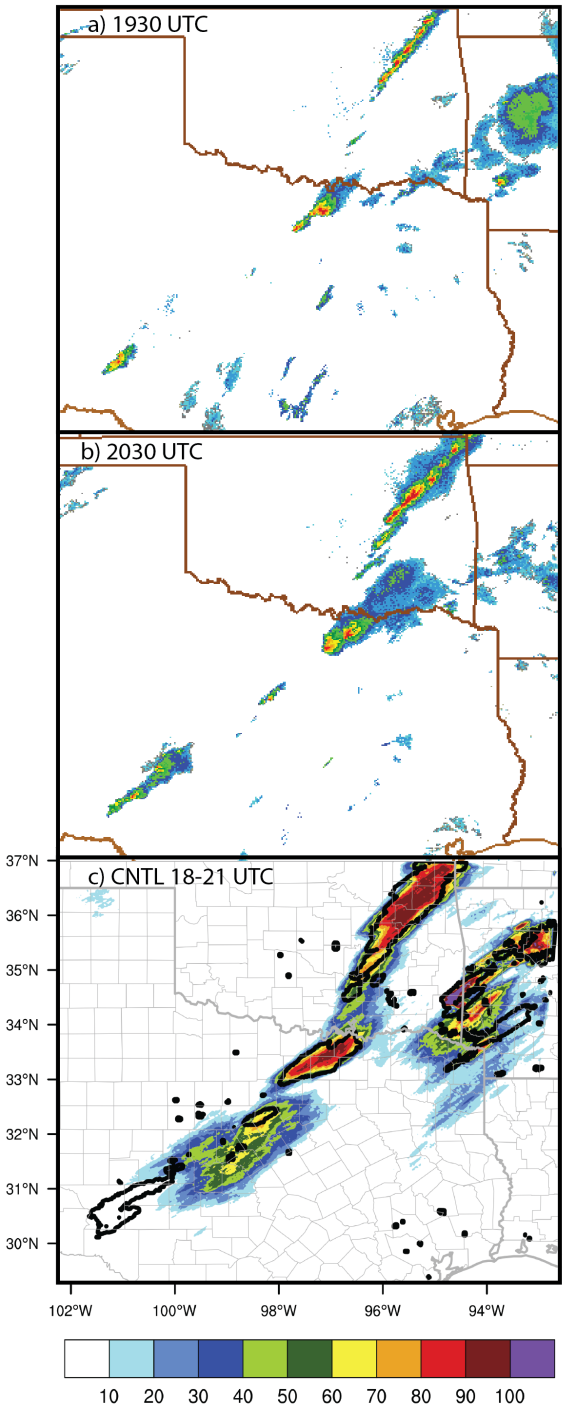

Figure 3.9. (a-b) Observed composite reflectivity mosaic valid 1930 and 2030 UTC on 3 April 2014, respectively. (c) Ensemble probability of maximum 3-h (18-21 UTC) composite reflectivity ≥ 35 dBZ (%), with observed 35-dBZ maximum 3-h composite reflectivity (black contours).

the actual storm extent (Figure 3.10c). There are, however, spurious CI locations nearby, with a couple areas to the NE showing probabilities greater than 50% in the ensemble. By 1850 UTC the swath of probabilities exceeding 30% extends from SW Wise County up to the Oklahoma-Texas border (Figure 3.10c).

The output of maximum hail size NEP is shown in Figure 3.11. The ensemble has high confidence in hail exceeding 10 mm and covers nearly all hail reports with above 50% probability (Figure 3.11a). There is a slight northern bias to the hail swath, however, as probabilities above 70% extend as far north as Oklahoma.

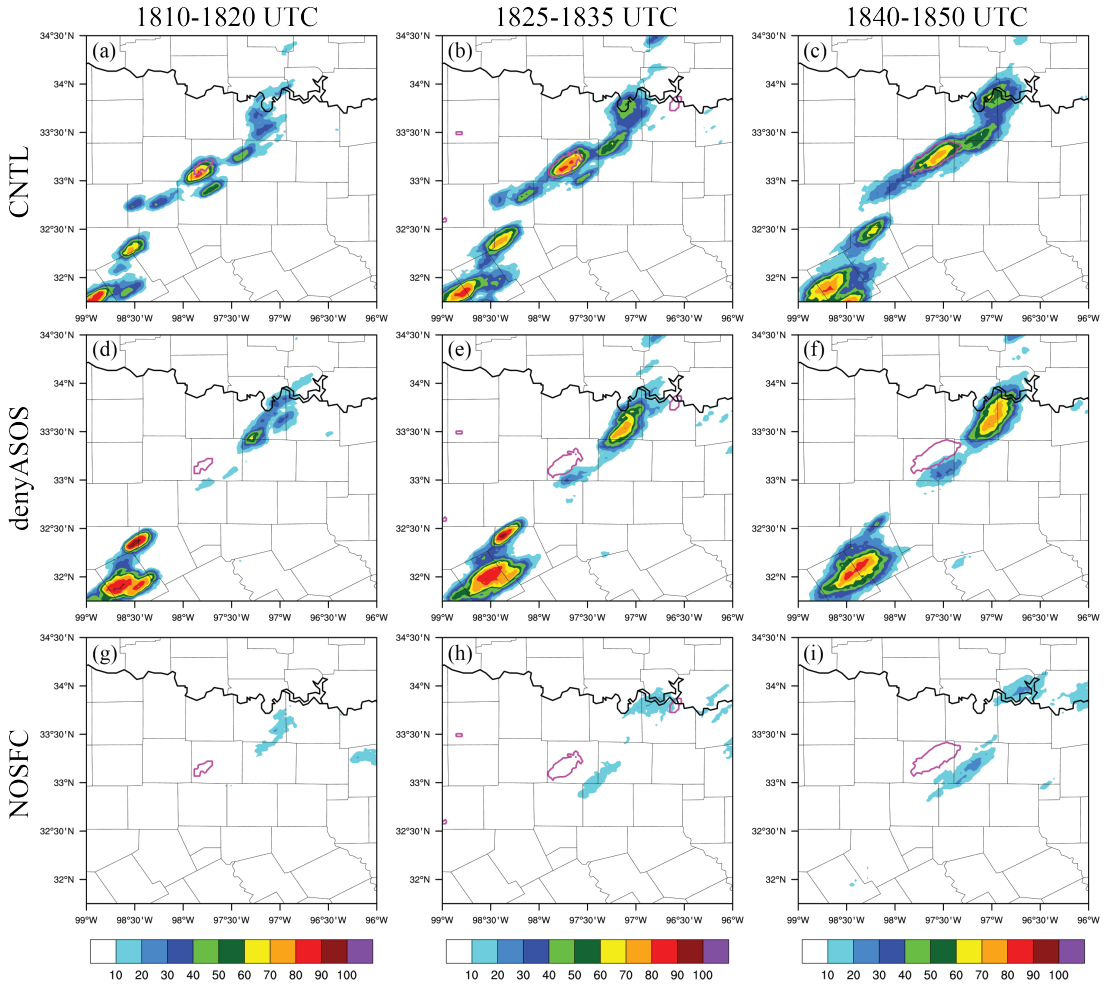

Figure 3.10. Ensemble probability of 15-min maximum composite reflectivity greater than 35 dBZ for experiments CNTL (a-c), denyASOS (d-f), and NOSFC (g-i). Magenta contour indicates observed maximum 15-min composite reflectivity for each time period indicated by columns.

With the higher severe threshold of 25 mm, probabilities are significantly reduced. There are two distinct centers of probability exceeding 50%. The first is in SW Wise County, just to the SW of the hail reports – an indication that the model grew the storms too quickly and aggressively compared to reality. The second peak is stronger and covers a larger area, with the center approximately 20-30 km NE of the maximum hail size reports located in Denton, TX. Despite the small location

errors, the CNTL ensemble was able to predict with high confidence severe hail in the north DFW region.

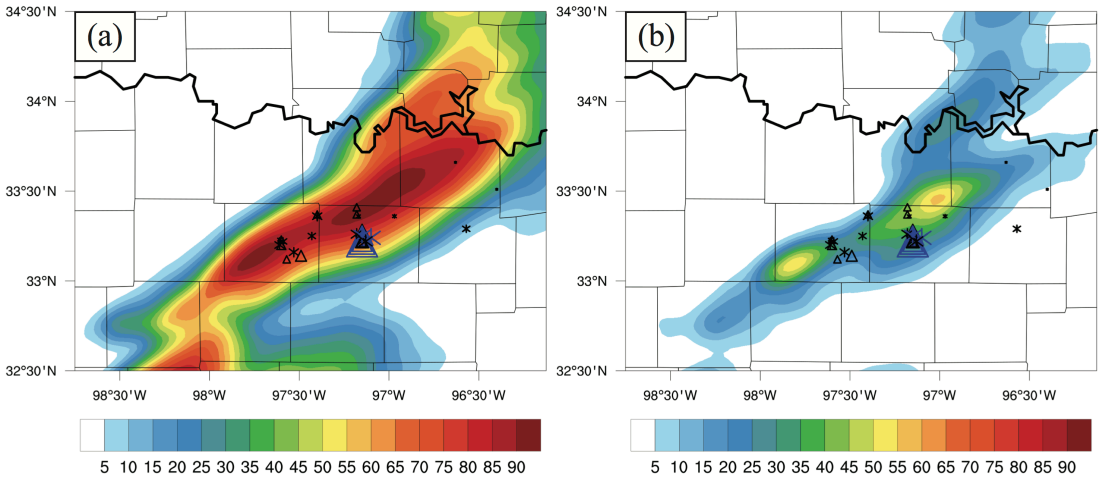

Figure 3.11. Neighborhood ensemble probability of maximum hail size in the entire column exceeding 10 mm (a) and 25 mm (b) for the entire 3-h forecast period (18-21 UTC) plotted for the CNTL experiment, using a neighborhood radius of 9.6 km. Triangles indicate SPC hail reports, and asterisks indicate mPING hail reports, sized according to size of the hail reported.

3.4.2 Evaluation of denial experiments

In addition to CNTL, experiments denyASOS and NOSFC are shown in Figure 3.10. These experiments were conducted as different baselines to compare to the nonconventional denials. One might anticipate a marked negative impact in denying ASOS and AWOS, since it is considered the most reliable surface dataset assimilated with the smallest observational errors. Additionally information is spread out at a larger scale than other surface data sets, due to the larger 200-km localization. As seen in Figure 3.10d-f, there is a substantial impact in the CI forecast, with the main area of CI shifted about 65 km to the NE, close to the Oklahoma border. Additionally this CI was delayed by about 15 minutes compared

to the CNTL. An even larger negative impact is seen when assimilating no surface observations (Figure 3.10g-i), where minimal CI occurs in the ensemble within the first hour. This extends further into the free forecast, as between 1900 and 2000 UTC less than one third of ensemble members show convection within the DFW domain (not shown). This result of experiment NOSFC is an important baseline to show the impact that *any* surface observations have on the CI forecast in this case – that CI is highly sensitive to surface observations. Comparing NOSFC to denyASOS, we can infer that nonconventional observations have added some value in capturing CI, even though there are larger timing and location errors than CNTL.

Ensemble probabilities of CREF ≥ 35 dBZ for nonconventional denial experiments are shown in Figure 3.12. The differences are generally smaller than the differences shown in Figure 3.10; however, there are still some notable effects. When denying all nonconventional data (NONEWSFC), the CI in SW Wise County is reduced in probability to below 40%, though two additional CI points just S and SW show probabilities in the 40-50% range. Additionally, a spurious point of CI occurs by the Oklahoma border, with probabilities over 80% (Figure 3.12a). This storm dissipates rapidly within the ensemble, as probabilities quickly diminish. For experiment denyERNET, there are many more spurious CI locations within the ensemble; additionally, the Wise County CI shows probabilities no higher than 60%, which is a small reduction compared to CNTL (Figure 3.12d-f). On the other hand, denyCWOP shows a marked reduction in the number of spurious CI points within the DFW region, though the CI in Wise County is reduced 10% more in probability as compared to denyERNET. Taken together, there are mixed signals in

positive and negative impacts – we can infer that CWOP data are important in establishing the CI in Wise County; however, CWOP also contribute to the amount of spurious CI in the ensemble. While ERNET data have a bit less impact on the CI itself, they do help counteract the spurious activity.

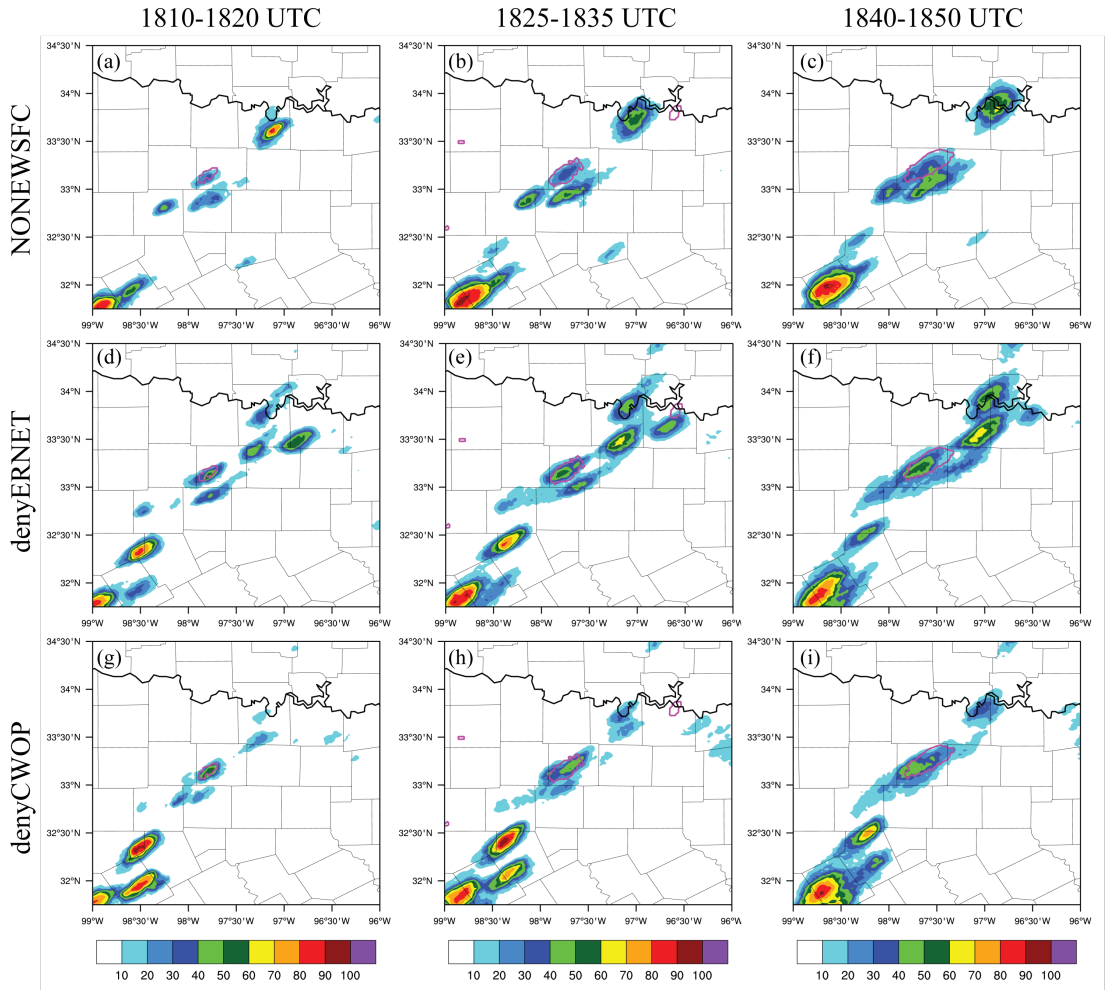

Figure 3.12. As in Figure 3.10, but for experiments NONEWSFC (a-c), denyERNET (d-f), and denyCWOP (g-i).

When looking at NEP of maximum hail ≥ 10 mm (Figure 3.13) denyASOS is again displaced to the NE, close to the Oklahoma border. Denying ERNET has a minor effect on the 10 mm hail forecast, but denying CWOP data constrains the hail swath to more closely match the report locations – eliminating much of the northerly extent of hail (Figure 3.13c). Experiment NONEWSFC shows reduced probabilities – less than 70% - and smaller hail swaths, though the location still matches the hail reports. A similar pattern can be seen with the NEP of 25 mm hail (Figure 3.14). There is a small NE displacement of the max hail probability in denyERNET, and probabilities are reduced slightly (around 10%) compared to CNTL (Figure 3.14b). There is less displacement in denyCWOP, but the probabilities are again reduced by around 10% compared to CNTL (Figure 3.14c). In NONEWSFC, less than a quarter of ensemble members show severe hail, which is as much as 40 percentage points lower than the CNTL (Figure 3.14d); less than 15% of ensemble members have severe hail in NOSFC (Figure 3.14f).

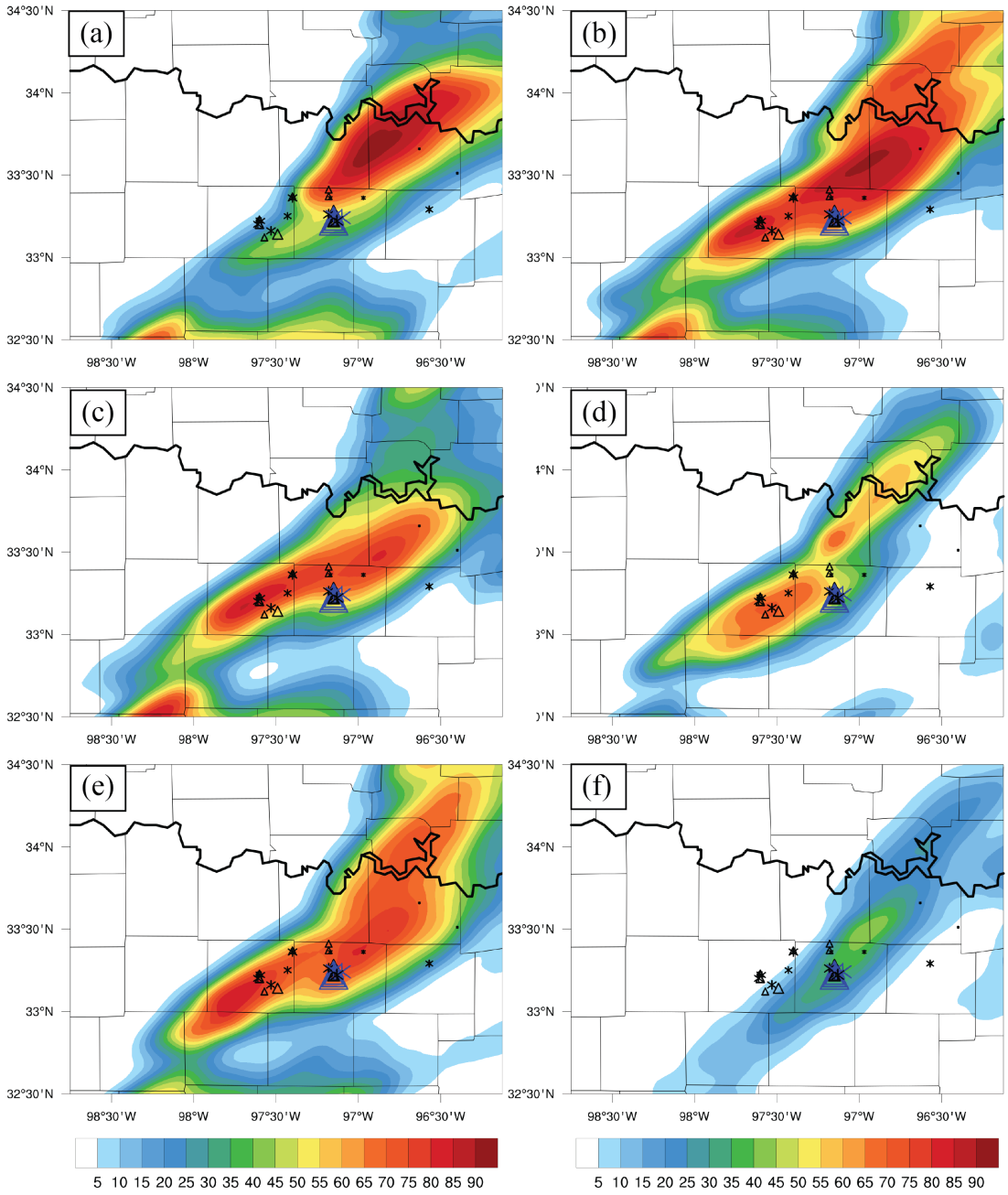

Figure 3.13. As in Figure 3.11a, but for denial experiments (a) denyASOS, (b) denyERNET, (c) denyCWOP, (d) NONEWSFC, (e) denyMISC, and (f) NOSFC

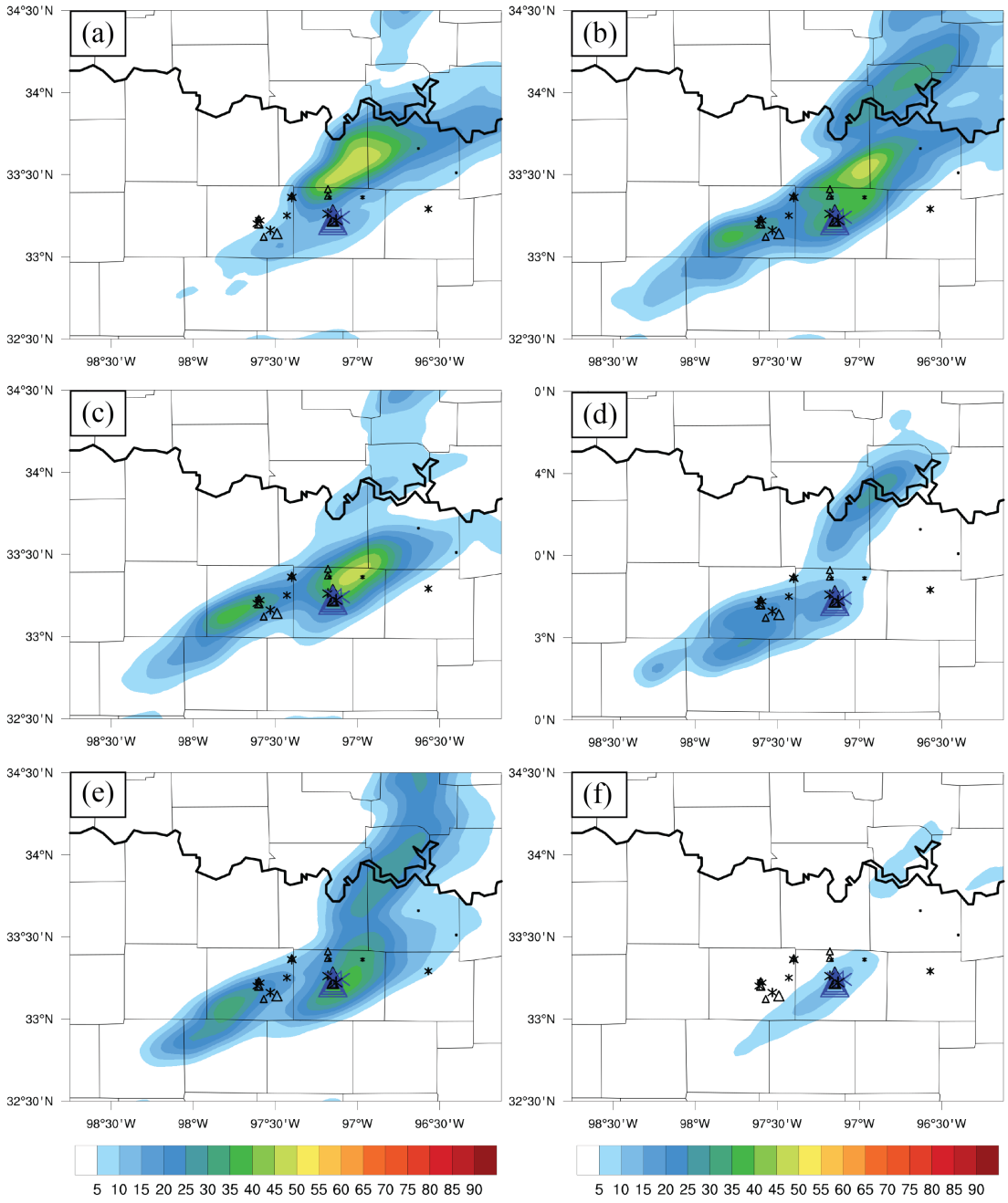

Figure 3.14. As in Figure 3.11b, but for denial experiments (a) denyASOS, (b) denyERNET, (c) denyCWOP, (d) NONEWSFC, (e) denyMISC, and (f) NOSFC

3.4.3 Evaluation of surface fields

Since CI is especially sensitive to boundary layer moisture for dryline cases, it is important to look at the surface moisture field to gain insight into some of the

differences shown in the ensemble probability fields. Dewpoint temperature for the final mean analysis is shown for CNTL and NOSFC in Figure 3.15a and b, respectively. The actual dryline placement error is not very large in NOSFC compared to CNTL, perhaps one county further east at most. However, there are several other notable differences in the dryline that led to CI in the proper location and time. First, the dryline in CNTL has stronger gradients and shows higher values of moisture in general – the dryline in NOSFC is over-mixed and has coarser gradients, as a result it is drier than it should be at locations too far east. Additionally, there are important kinks in the dryline in CNTL that are not as apparent in NOSFC. Several studies have linked CI to dryline bulges and kinks such as that seen in CNTL (e.g., Hane et al. 1997; Hane et al. 2002). Near the CI location, the local dryline is approximately east-west in orientation in CNTL, whereas in NOSFC no such small-scale variation exists. Due to the east-west local variation, an area of enhanced convergence exists owing to the largely southerly flow in the moist regime being locally perpendicular to the dryline. Additionally, there appears to be a small pocket of enhanced moisture in that location.

Of course, it can be challenging to look at a moisture plot and simply point to an exact location of CI. Not all dryline kinks result in convection. One avenue to address this is to look at plots of mass convergence or moisture flux convergence (MFC) to identify favorable areas of CI (Banacos and Schultz 2005).

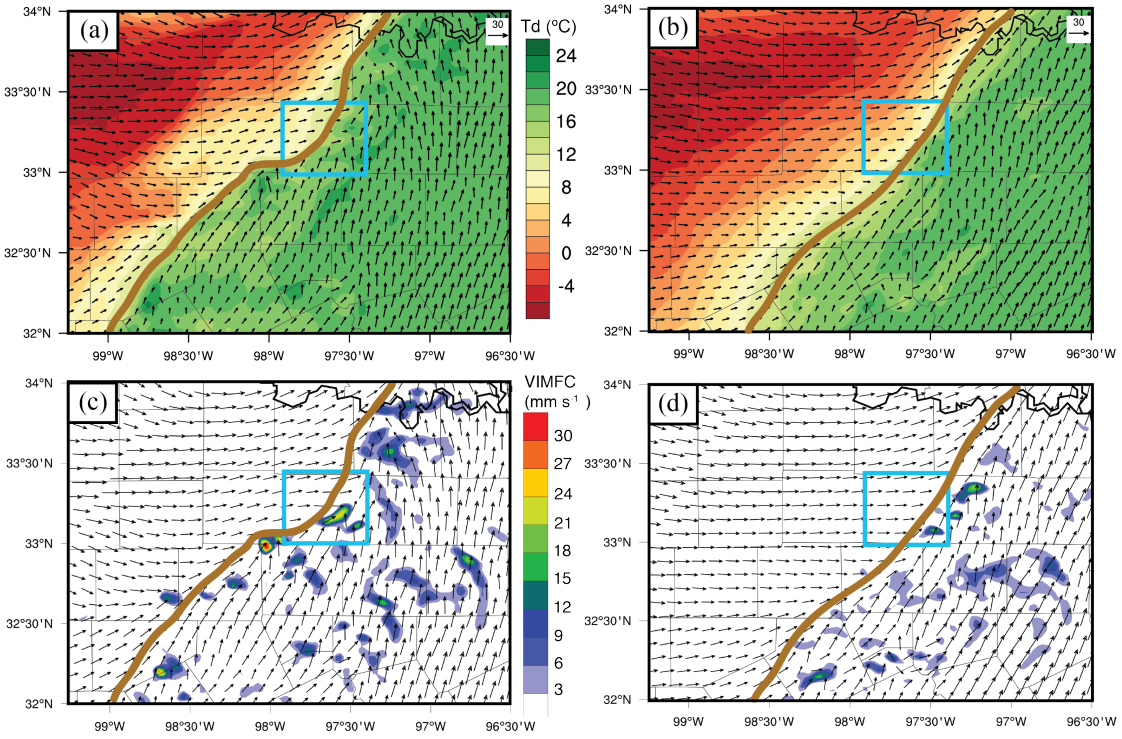

Figure 3.15. (a-b) Final analysis mean 10-m dewpoint temperature (color fill) and wind (vectors), valid 18 UTC 3 April 2014, for experiments CNTL and NOSFC, respectively. (c-d) Vertically-integrated moisture flux convergence (mm(H_2O) s^{-1}) computed for the lowest 2-km above ground level, with wind (vectors) at 2 km above ground, valid 18 UTC 3 April 2014 for experiments CNTL and NOSFC, respectively. Thick brown line indicates approximate dryline location, and blue box indicates location of Wise County, TX.

While the utility of a snapshot of MFC is questionable in a forecasting sense, it can be used as an important visual tool for CI case studies. One such example is Xue and Martin (2006a,b), who with the aid of MFC plots developed a conceptual model for dryline CI that posits locations of enhanced convergence occur at intersections between boundary layer convective rolls and the primary dryline convergence band. It is possible the east-west kink in Figure 3.15a is a reflection of such a process, though the horizontal resolution of this forecast is too coarse to fully resolve convective rolls in the PBL.

A drawback of MFC is that it only reflects a single level, while triggering CI is a 3D process that requires convergence at more than one level such that parcels can be lifted to the level of free convection. Additionally, plots of MFC along the dryline often show many areas of enhanced convergence – it may show favorable areas of CI, but does not discriminate very well between CI from none. For this reason, we have instead plotted vertically integrated moisture flux convergence (VIMFC) over the lowest 2km of the atmosphere, to reflect the amount of forced lifting in the PBL. VIMFC was explored as a discriminator on synoptic scales by van Zomeren and van Delden (2007); however, to the best of our knowledge has not been used for convective scales. VIMFC is analogous to precipitable water in that it reflects the depth of liquid water within a column – in the case of VIMFC, it reflects the liquid equivalent depth of water entering the column per second. As seen in Figure 3.15c, CNTL has a strong center of VIMFC associated with the east-west local kink near Wise County. With a strong center of VIMFC clearly CI is imminent, and this helps explain why timing and location errors were very small in CNTL for this cell. On the other hand, VIMFC shows much more limited favorable CI areas in NOSFC (Figure 3.15d), which helps to show why convective activity was low in the experiment.

Difference fields of dewpoint are shown in Figure 3.16 to further focus on factors that lead to the enhanced VIMFC shown in the CNTL experiment. Overlaid on these difference plots are the denied observations minus the analysis, which can be thought of as the observation innovation that the denied observations would have had if they were assimilated in the denial experiment. The goal of this

is to clearly tie the differences between the CNTL and denial experiments – which include nonlinear effects of model integration – to the actual observations had they been assimilated. As such, we can see many important impacts from each observational dataset. From Figure 3.16a, there are a few ERNET observations upstream of the CI location in relatively data-sparse locations that help feed moisture to the CI location. The CWOP data – which had a larger impact on the CI probability – have fortuitous data coverage near the CI location to help capture that east-west kink in the dryline, as locally there are dry and moist observations directly north and south of each other right next to that CI location (Figure 3.16b). ASOS data, which had the largest impact of the surface datasets, can be seen to have more broad-scale impacts in Figure 3.16c, which helps to constrain the gradient and location of the dryline from the over-mixing tendency found within NOSFC.

Finally, though not discussed thus far, the miscellaneous mesonet data had a surprising impact on the moisture field as well, as observations far away SW of the CI location were found to feed directly to that location in a river of moisture after 2 hours of model integration. Denying this data did have a slightly larger negative impact in terms of severe hail (Figure 3.14e) than either denyCWOP or denyERNET. These observations to the SW were from a hydrometeorological observation network, located often next to small lakes and streams. As such it is possible they have an additional representativeness error that was unaccounted for, though it is also worth noting the observations in Figure 3.16d also lie right

along the dryline gradient so there is an added sensitivity to these observations because of proximity to this feature.

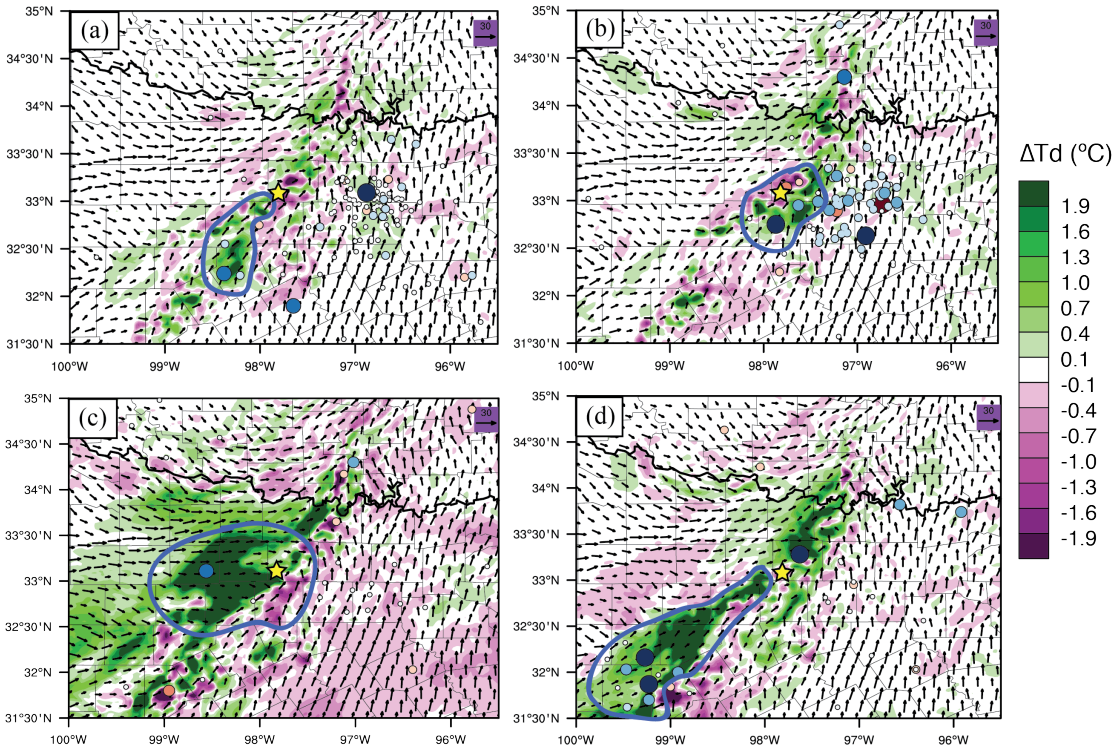

Figure 3.16. Difference fields for ensemble mean 2-m dewpoint temperature (color fill) and 10-m wind (vectors) for final analysis time 18 UTC 3 April 2014: (a) denyERNET minus CNTL, (b) denyCWOP minus CNTL, (c) denyASOS minus CNTL, (d) denyMISC minus CNLT. Color-fill dots indicate respective observations from each denial experiment for each plot, with colors and sizes indicating the O-A values for each denial dataset (i.e. what the observation innovations would have been if denied observations were assimilated). Yellow star indicates approximate CI location in SW Wise County, and blue outlines highlight relevant observations influencing the CI forecast.

In summary, Figure 3.16 indicates that there is a cumulative effect of all these different observations. Each network is adding moisture in a different way. This helps explain why the location and timing of CI was not greatly affected by the nonconventional denials, but the ensemble probabilities of CI and hail were reduced.

The RMSD and bias statistics for all experiments are shown in Figure 3.17, computed against all surface observations available. The negative time period RMSD and biases were calculated using first guess and analysis ensemble means during DA cycling creating a "sawtooth" appearance, while the positive period uses ensemble means from the 3-h free forecast period (18-21 UTC). Both temperature and wind speed show little difference in RMSD among all experiments during the free forecast, aside from a slight degradation in RMSD of wind for denyASOS. Most of the notable differences occur in specific humidity. Experiment NOSFC is generally 0.5 g kg^{-1} higher than CNTL. Interestingly, CNTL is somewhere in the middle of the experiments in terms of RMSD, while denyCWOP and denyCWwnd are generally lowest. In terms of bias, there tends to be a model dry bias in moisture on the order of 0.5 to 1.0 g kg^{-1}. There is also a high wind speed bias in the model, which may also be a reflection of the low speed bias of these nonconventional data discussed in Carlaw et al. (2015).

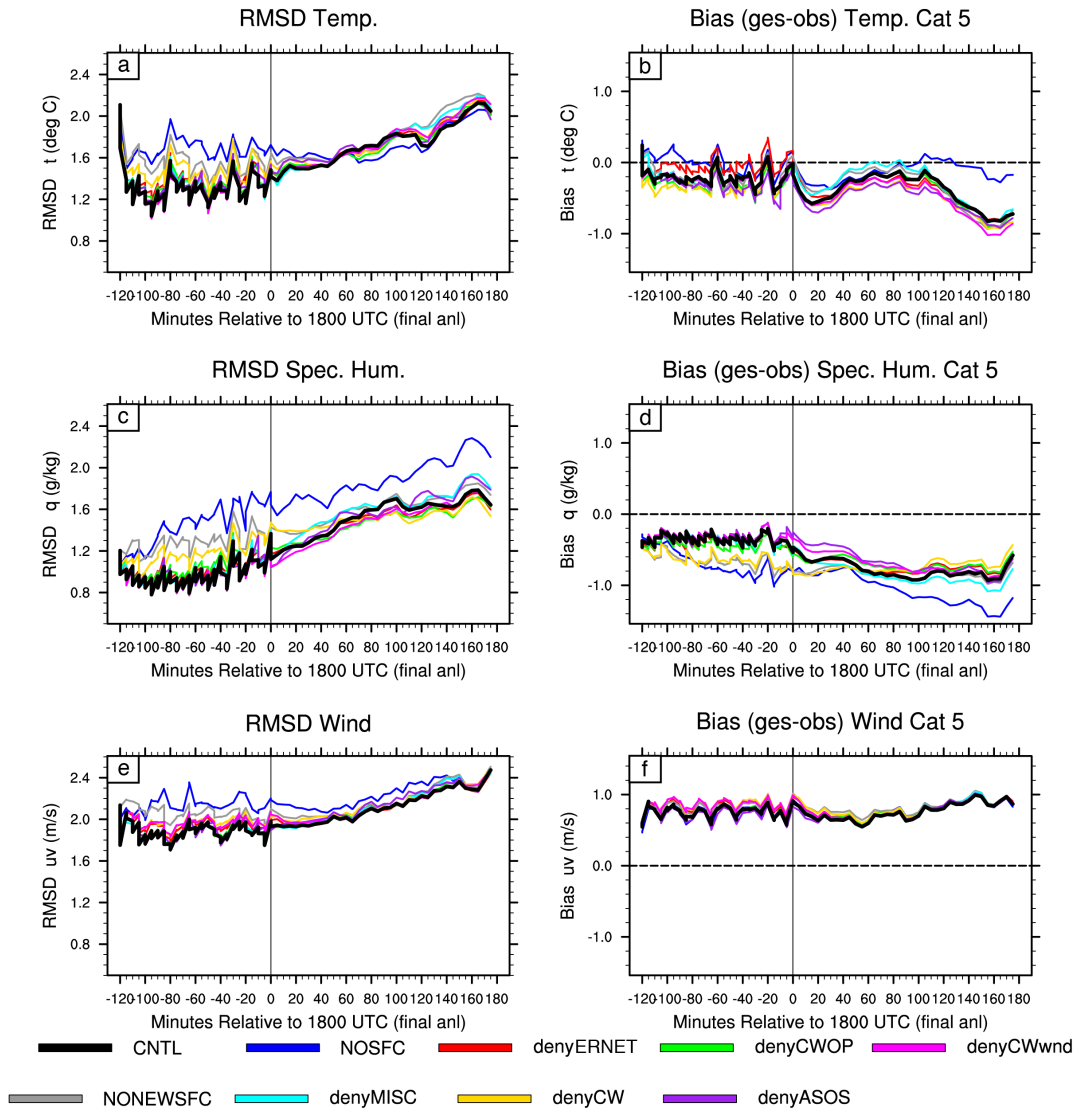

Figure 3.17. Time series root mean square difference (RMSD) and bias (model minus observations) of ensemble mean for all experiments, plotted for temperature (a-b), dewpoint temperature (c-d), and wind (e-f). Note that the bias in (f) is wind magnitude bias only.

3.4.4 Sensitivity to observation type

Carlaw et al. (2015) found the majority of impact from nonconventional observations came from thermodynamic information. On the other hand, SS15 found slightly more impact on their CI forecasts from the wind information. To find out which information has a larger benefit for this case study, experiment denyCWwnd was conducted, where only the wind observations from CWOP and ERNET are denied but thermodynamic variables are still assimilated. Compared with denyCW, denyCWwnd matches much more closely with CNTL (Figure 3.18), with only a minor reduction in ensemble probability within the first hour of the free forecast. But interestingly, the forecast of hail is actually *improved* in denyCWwnd for both 10 mm and 25 mm thresholds (Figure 3.19). In fact the 25 mm NEP has improved by 10% and shows a max probability slightly closer to the significant hail reports in Denton. So in this case study, it is clear that the thermodynamic information from the nonconventional datasets – specifically moisture – is of greater importance than the wind information.

What could be the cause of wind observations degrading the forecast? Previous studies focusing on nonconventional mesonet observations have mentioned low wind speed bias concerns of wind measurements, which is a potential source of degradation (Hilliker et al. 2010; Carlaw et al. 2015). Another source can be inferred from Figure 3.17. Denying just the wind observations from CWOP and ERNET results in an almost negligible difference in the forecast RMSD of wind – an indication that the wind observations are not adding much information. Additionally, the RMSD of specific humidity slightly improves over the

CNTL just from denying wind observations. If the wind observations do not add much information, then their assimilation is not necessary and instead will only accumulate noise in the analysis from the large number of observations being assimilated. In other words, it is possible the variability of the ensemble wind is actually smaller than the variability of the nonconventional wind observations themselves due to siting issues.

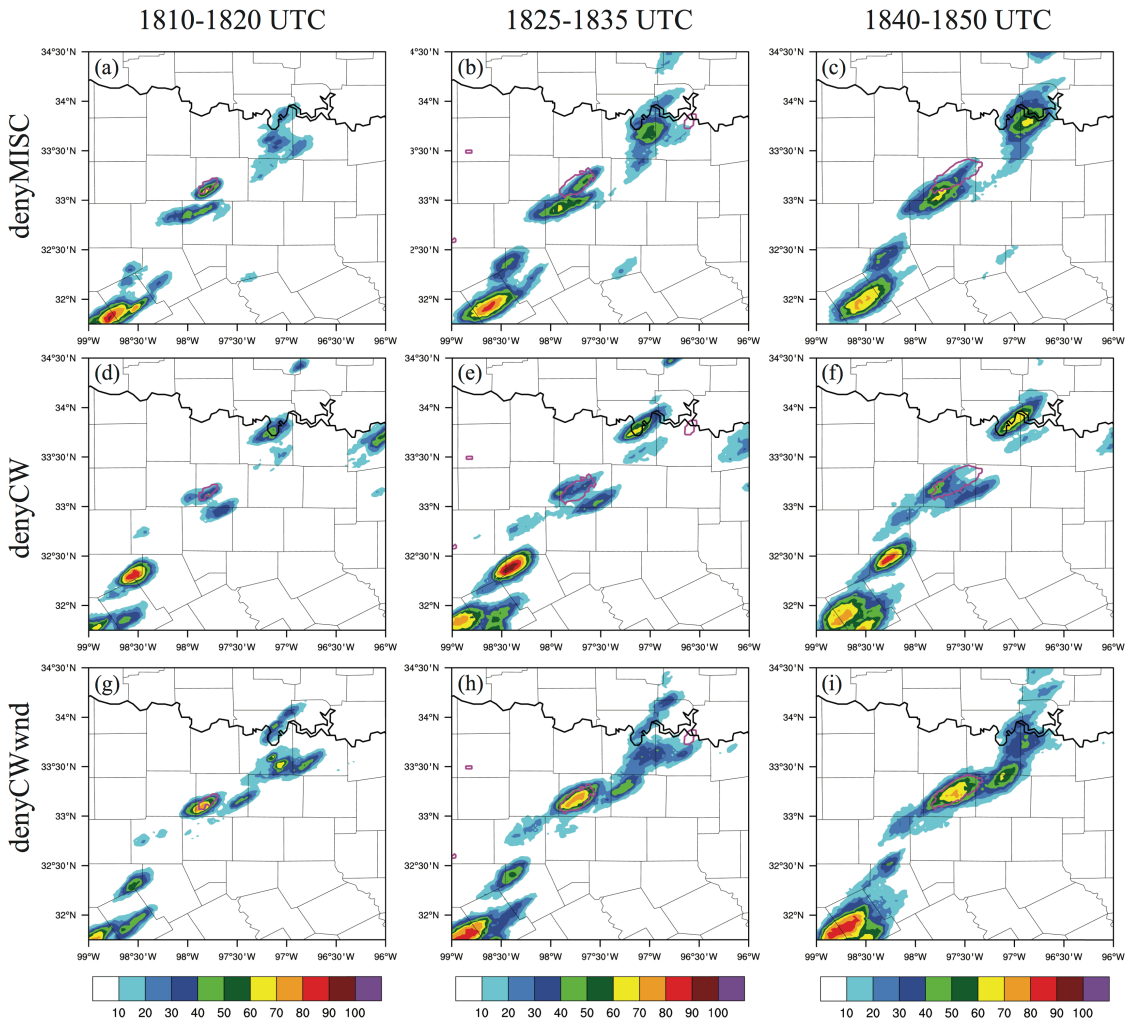

Figure 3.18. As in Figure 3.10, but for denial experiments (a-c) denyMISC, (d-f) denyCW, and (g-i) denyCWwnd.

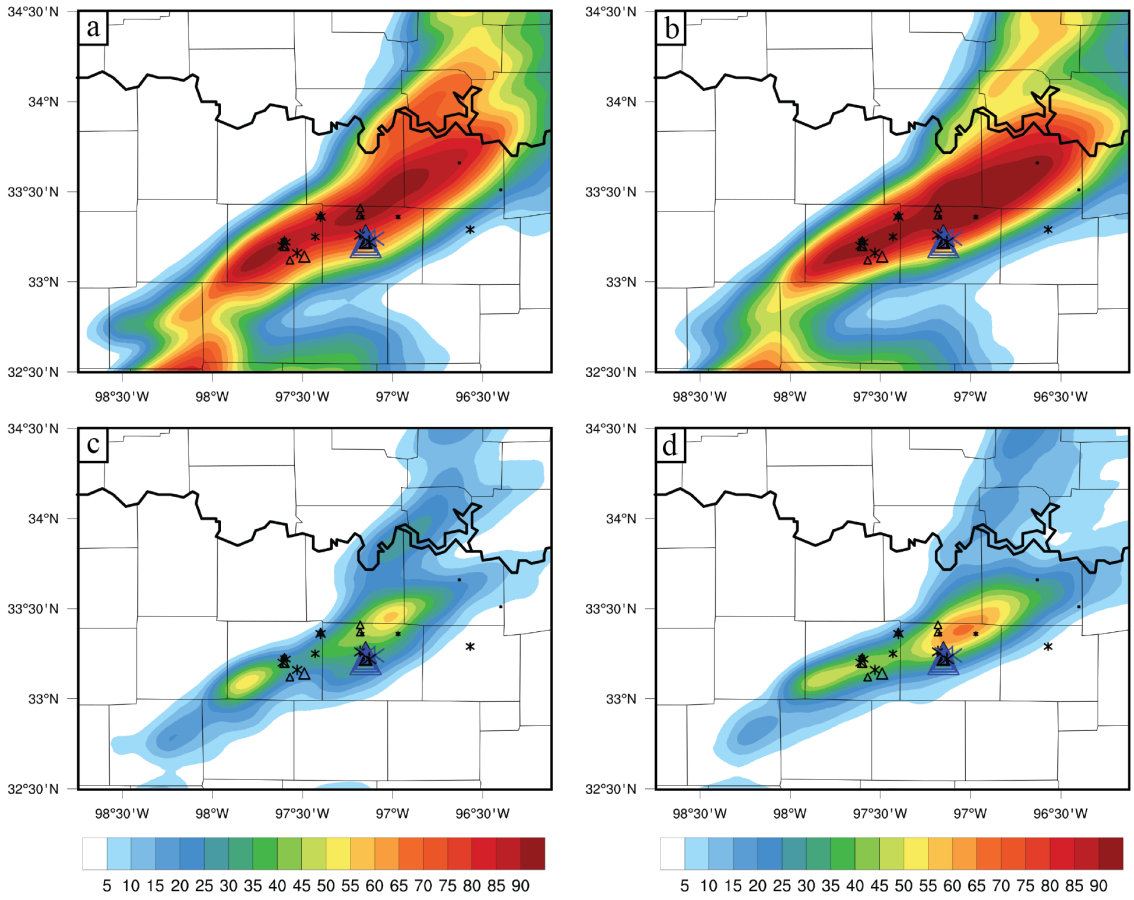

Figure 3.19. As in Figure 3.11, but comparing NEP of max hail size exceeding 10 mm (a-b) and 25 mm (c-d) for experiments CNTL (a,c) and denyCWwnd (b,d).

3.5 Summary and discussion

While the number of studies assimilating surface observations in addition to radar observations has increased, there have been very few to examine the use of surface mesonet observations on forecasts of CI. Additionally, these mesonet surface observations are typically taken as one observation source. In this study, we examined the impacts of different systems of nonconventional surface observations on the forecast of CI from an extreme hail event in the DFW testbed, 3 April 2014, using data denial experiments. Our results indicated that while

nonconventional data sources did not have a very large magnitude impact compared with conventional surface observations, they still had an important contribution to the accurate prediction of CI and subsequent hail within the ensemble. Each nonconventional dataset contained observations from different locations that modified the moisture near the CI location, leading to a cumulative effect of all networks. These important observations were all located in advantageous areas just along the dryline, bringing to mind a key result of Tyndall and Horel (2013), "location, location, location." They examined analysis impacts of mesonet observations and found those observations located in data sparse regions or in regions with sensitive local weather patterns (i.e. coastal regions) had the biggest impact. Similarly, the most important observations in this study were found within the dryline gradient, in areas where observations were relatively sparse compared to the densest part of the network in the most populous parts of DFW. In terms of the NNoN concept, dryline positions change on a daily basis so aside from a broad climatological average of dryline location it is difficult to know exactly where to site the observations, so a general approach such that observations are spread out evenly (as in the Oklahoma mesonet) is preferred. Alternatively, mobile platforms such as unmanned area vehicles (UAVs) may be useful for measuring the dryline gradient.

The high-frequency assimilation of surface observations assisted in successfully constraining model biases related to the PBL scheme, as also found in SS15. The ASOS and AWOS data largely constrained the dryline position and gradient strength, preventing the model from over-mixing the dryline and causing

the surface to be too dry in areas where CI occurred in reality. The addition of nonconventional observations helped to define the important small-scale features in drylines that are often the foci of CI. In this case, CWOP observations helped to define a small east-west kink in the dryline over Wise County, TX that allowed for the southerly flow within the moist regime to be more normal to the dryline, enhancing convergence and lift.

While the potential value of assimilating CWOP is intriguing, there are many issues that remain a challenge. Despite the positive result in CI, it was also clear that assimilating CWOP led to additional spurious convection, due to noisiness in the wind observations inducing small noisy areas of convergence where none exist. Additionally, the siting concerns of CWOP prevent it from being optimally utilized. Nearly 10% of the CWOP observations in the domain of this study were flagged with having a reported elevation significantly different than the model terrain – some may be an inaccuracy in reported elevation (such as reporting in feet rather than meters), but others could be a misreported location. Separately we have found errors in many observing systems, including Federal stations, on the order of the grid spacing used here. While the GSI automatically handled these siting issues by inflating their observation errors, it still leaves some untapped potential into fully utilizing these nonconventional observations. ERNET data do not share as much of the location and reporting concerns as CWOP, but station siting is still a big issue. We simply do not know what height these observations were located, and could only make an assumption of 2-m AGL. Since many are located on tops of buildings, it is possible that this is introducing an artificial high

bias when using those ERNET observations, as the actual height of the observation is higher than reported.

In general, the metadata for nonconventional observations needs to be improved before they can be used reliably in DA systems. This study shows that we need to treat each observation source appropriately – it is not good enough to simply lump CWOP data in with the same assumed error as Oklahoma Mesonet. Additionally, observations that we have not explicitly studied need to be examined more carefully for any sources of bias. In this work, hydrometeorological observation stations had a positive impact on the CI location but also enhanced some spurious activity. Are observations at these stations moist-biased – or have enhanced representativeness error – due to their locations next to small lakes and streams? We need to learn about observation data issues in representativeness so that it can be treated properly within the DA systems, either by removing the bias or setting more appropriate localizations and observation errors. Although this thinking leads to more complicated DA configurations with endless tuning possibilities, these fine details will determine the future utility of these high-resolution surface observations for improving high-resolution forecasts.

Despite the issues with nonconventional observations, this case study was still important in quantifying the value we can get from these surface observations in our DA systems even in the absence of radar data. They truly can give us the small-scale information that is necessary to pinpoint the correct location for CI in a dryline case. The caveat is this is just one case study, though an important high impact case. More cases need to be examined utilizing nonconventional

observations. Other ideas include evaluating a month-long quasi-real-time study or a series of high impact weather cases so that effects are not diluted by quiescent days.

Finally, this work lays the foundation of experiments to examine the application of the EFSO on the convective scale. The application of EFSO is the subject of chapter 5.

Chapter 4: Adaptive Localization for the Ensemble-based Forecast Sensitivity to Observations (EFSO) Metric using Regression Confidence Factors

4.1 Introduction

The EFSO metric is appealing because ensemble perturbations take the place of the adjoint model in estimating sensitivities. However as with any ensemble method, it suffers from sampling error, which occurs when the number of ensemble members is small compared to the degrees of freedom in a model and observing system – predominantly the case in ensemble NWP. Sampling error results in spurious correlations and can lead to filter divergence in deterministic EnKF assimilation. The problem of sampling error caused by small ensembles is a more serious issue for the EFSO metric. Localization can be applied to alleviate sampling error; however, a time-forecast component is added to the localization problem, such that a straightforward application of fixed localization techniques would not guarantee accurate impact estimates with EFSO.

This chapter explores the application of a dynamic localization method for the EFSO metric. There are two main purposes of this study. The first is to simply learn what a 'proper' localization function looks like for the EFSO impact estimate, and how it evolves with increasing forecast time. Kalnay et al. (2012) and Ota et al (2013) only considered moving the localization function, but it is possible that the shape and magnitude also need to evolve with forecast component, something that an adaptive method will be able to automatically determine. The second purpose is to test the potential effectiveness of an adaptive method on EFSO. Adaptive

localizations generally have been shown to provide more accurate assimilations at increased computational cost relative to fixed localization; in this chapter it is explored whether the application of such adaptive methods initially developed for assimilation can be extended and applied successfully to the EFSO metric. Briefly, the dynamic localization function used here is obtained from confidence factors derived using groups of ensembles, first proposed for EnKF by Anderson (2007). The focus of this chapter is to extend this group filter (GF) concept in the context of the observation impact estimate.

In section 4.2, the regression confidence factor (RCF) method is introduced. Section 4.3 describes the experiment setup. As an initial test of the method on observation impact estimates, an isentropic two-layer primitive equation model (Zou et al. 1993) under the perfect-model assumption is adopted. This model is coupled with the LETKF data assimilation system following Holland and Wang (2013). The RCF calculation settings are also explained in section 4.3. The resulting RCF localization functions are shown and applied in section 4.4 for both single-observation and full-domain observation assimilation experiments. The accuracy of the ensemble observation impact estimate using the dynamic localization is compared with that of using fixed Gaspari-Cohn localization for different observation locations (tropical vs. midlatitude), forecast length, and differing state variables. The goal of comparison to static GC is to first determine if the adaptive method is better for EFSO. More importantly, the GC function is used as a tool to provide a baseline for comparisons with RCF localization experiments, to help provide context for qualitative discussion. From this comparison, an

important relationship between localization used during assimilation and for the impact estimate was discovered and is discussed in sections 4.4.4-4.4.5. Another potential use for adaptive methods is in the ability to automatically tune GC localization. This concept is introduced and tested in section 4.4.6. A summary and discussion are given in section 4.5.

4.2 The RCF method of computing localization for EFSO

The GF method of Anderson (2007) operates using groups of ensembles to calculate regression-sampling errors in the ensembles. Assume that g groups of k ensembles are available in an assimilation system. When computing the linear regression between the state variables and observations, there are g samples of the regression coefficient, β. A weighting factor, α, is defined to minimize the expected RMS differences between all possible combinations of sample β pairs. So, α is chosen to minimize

$$\sqrt{\sum_{j=1}^{g}\sum_{i=1,i\neq j}^{g}\left(\alpha\beta_i-\beta_j\right)^2}\ . \tag{4.1}$$

A simple derivation (see Anderson 2007) leads to the following expression for α_{min},

$$\alpha_{\min}=\frac{\left[\left(\sum_{i=1}^{m}\beta_i\right)^2\Big/\sum_{i=1}^{m}\beta_i^2\right]-1}{g-1}=\frac{g-Q^2}{(g-1)Q^2+g}, \tag{4.2}$$

where Q is the ratio of the sample standard deviation to the absolute value of the sample mean of the group β's. The optimal weighting factor α_{min} is also known as

the *regression confidence factor* (RCF). A unique α_{min} can be calculated for each observation-state pair. The set of RCFs for a given observation and all state variables is called a *regression confidence envelope* and can be used directly as a localization function (Anderson 2007).

In the hierarchical filter of Anderson (2007), the regression coefficient β represents the regression between observations $\mathbf{y}_0$ and state background $\bar{\mathbf{x}}_0^b$. So β is the covariance between $\mathbf{y}_0$ and $\bar{\mathbf{x}}_0^b$ normalized by the variance of $\mathbf{y}_0$. To apply this method to EFSO, a different regression is considered to be between the analysis in observation space and the forecast, such that for each observation *l* and state variable *j,*

$$\beta_{l,j} = \frac{\left(\mathbf{H}\mathbf{X}_0^a \mathbf{X}_{l0}^{fT}\right)_{l,j}}{\left(\mathbf{H}\mathbf{X}_0^a (\mathbf{H}\mathbf{X}_0^a)^T\right)_{l,l}}. \tag{4.3}$$

The use of the RCF method for the EFSO impact metric is inherently limited by the linear regression approximation needed to compute the group β's according to (4.3) so there is an inherent limit to the forecast length at which it can be successfully applied.

4.3 Experiment design

4.3.1 The assimilation and forecast system

To evaluate and explore methods to improve the accuracy of EFSO in quantifying observation impact, experiments with a simplified primitive equation model and simulated observations were done. The dry, global, two-layer primitive

equation spectral model of Zou et al. (1993) was chosen, which has been used in several studies of perfect- and imperfect-model ensemble-based data assimilation experiments (e.g., Wang et al. 2007, 2009; Holland and Wang 2013). It is useful due to its low computational demands allowing for many experiments to be conducted. The model variables include two vertical layers of vorticity, divergence, and layer thickness coefficients. The layer thicknesses, $\Delta\pi_1$ and $\Delta\pi_2$, are described in terms of π - the Exner function. The model includes simple parameterization schemes for radiative heating and surface drag, with zonal wavenumber-2 terrain. A fourth-order Runge Kutta scheme is used for forward integration.

The model was run using the same parameters as in Holland and Wang (2013). To isolate the impact of sampling error, the experiments were conducted in a perfect model context. A model run with daily output over thousands of days of integration at T31 resolution served as the truth. An initial ensemble was generated by a random draw of the truth states. The assimilation-forecast cycles were run at the T31 resolution for 1000 cycles at one day *(n* = 24 h) intervals due to the long error doubling time of 3.78 days (Hamill and Whitaker 2005). The first 100 cycles were discarded from the RCF and impact calculations to allow the system to stabilize. Observations in interface height – the height between layer 1 and layer 2 – were generated from the truth by adding errors drawn from a distribution with zero mean and fixed standard deviation of 250-m, as in Wang et al. (2007). There are 362 equally-spaced observations total. These observations were assimilated into the LETKF following the settings of Holland and Wang (2013) with multiplicative and additive inflation. GC localization was applied

during assimilation with a cutoff radius of 8000 km, which was optimally tuned but, more importantly, provided a stable forecast-assimilation system for the full 1000 cycles.

4.3.2 Settings for impact estimate experiments using RCF

Figure 4.1 shows a flowchart of RCF computations. An initial set of ensembles is randomly split into g groups of k ensembles each prior to assimilation. Cycled LETKF analyses are done for each group of k ensembles separately. Following analysis, an ensemble forecast is run to some valid time t.

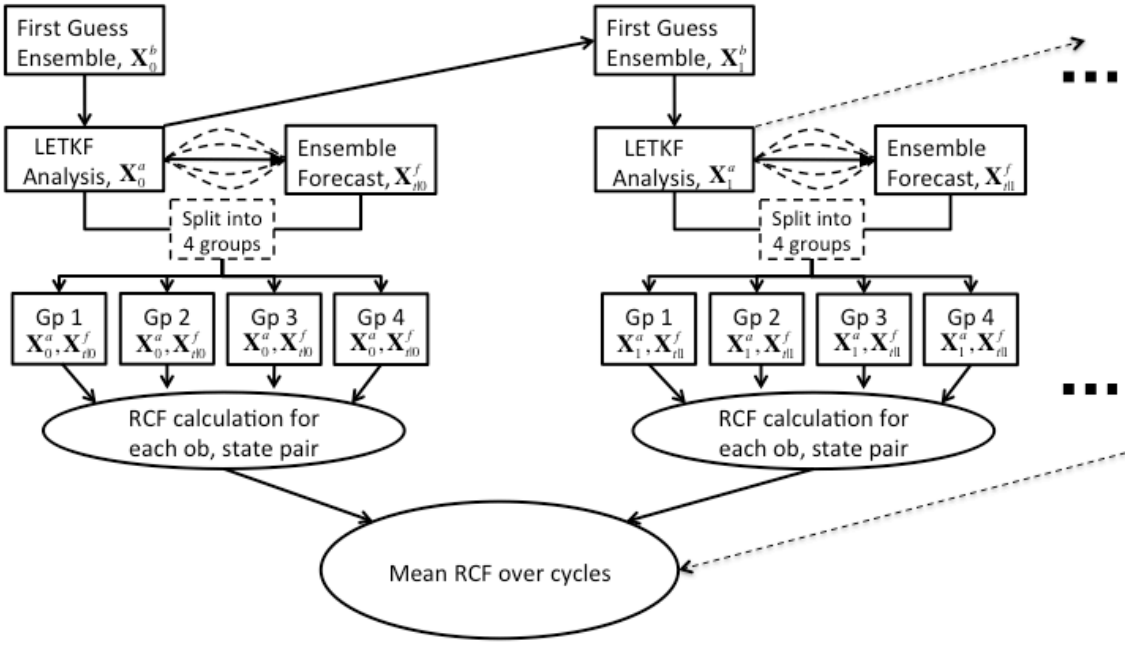

Figure 4.1. Flowchart of RCF method. After an LETKF ensemble analysis, an ensemble forecast is run to some time *t*. Then the analysis and forecast ensembles are randomly split into four groups, and for each group β is calculated according to Equation (4.3) . The RCF is then computed according to Equation (4.2) for all observation-state pairs. This RCF envelope, unique for each analysis cycle, is fed into a running average over all the analysis cycles. The mean RCF envelope then serves as the GF localization function for impact estimate experiments. For the experiments in this chapter, the mean RCF function is calculated over 900 cycles total.

For each group, a regression coefficient β is then calculated according to (4.3) from which RCF is calculated according to (4.2) and saved for each assimilation cycle. Each grid point-observation pair has a unique RCF. Since the RCF from each cycle is noisy due to a small number of groups used in the calculation, particularly for locations far away from an observation, the next step is to take the mean of the RCFs across all cycles to dampen out those effects. This results in a 'lookup table' of RCF functions for every observation. There are up to 256 ensemble members total, which are grouped randomly prior to assimilation. Initial experiments tested the sensitivity of resulting RCF functions with varying number of groups (2,4,8, and 16) using 16-member groups, and varying the number of ensembles per group (8, 16, 32, and 64) using 4 groups. For the subsequent impact experiments, RCF functions were calculated using g = 4 groups of k = 16 ensemble members each (64 ensemble members total) and applied as localization functions for ensemble impact estimates using a 16-member ensemble.

Impact experiments were conducted varying the forecast valid time t from 0 to 4 days. At $t = 0$, no forecast is run so the method gives RCF functions for the impact of the observations on the *analysis*. RCF functions were calculated first for model interface height, z_{int}, and then in terms of the state variables $\Delta\pi_1$, $\Delta\pi_2$, u_1, u_2, v_1, v_2. RCF functions for the latter are considered cross-variable impacts, which will serve to examine the effects of using the new localization method to estimate the impact of observations on unobserved variables.

Once the RCF functions were calculated, they were applied directly to the impact metric in Equation (2.17) using one of the 16 ensemble group assimilations and compared with ensemble estimates with no localization and using a static GC localization. The results were validated against the *actual impact*, which is the impact calculated directly from Equation (2.15) using the truth. Additionally, sets of single-observation experiments were conducted to better understand the results qualitatively, to aid in viewing specifically the time- and cross-variable components of the localization and the impact estimate.

4.4 Results

4.4.1 RCF localization

Prior to application of the dynamic RCF functions to the EFSO metric, we first examine the structure of the localization functions outputted from the procedure described in Figure 4.1. At the analysis time (Figure 4.2a) RCF appears to have a Gaussian-like shape to it, though there are some differences such as narrow peaks and heavier tails. The effect of averaging over 900 cycles has smoothed the functions, though some noisiness still remains particularly in the tails of the distributions. Increasing the number of ensemble members per group increases the width of the RCF function, mostly greater than 30 degrees in longitude away from the observation. The increased RCF width is because larger ensembles are less prone to noise from spurious correlations until greater distances from the observation. This result is consistent with results of Anderson

(2007, c.f. Fig. 4), as well as studies examining changes in optimal GC length scale with changing ensemble size (e.g., Hamill et al. 2001).

RCF functions for a 2-day forecast are shown in Figure 4.2b. The time-forecast dependency causes the main signal to dampen and shift downstream from the observation location. The diminished magnitude suggests less confidence in ensemble covariances at longer forecast lead times. With increasing ensemble member size per group, there is an increase in the strength of the RCF function across all longitudes. Each ensemble size is able to capture the same time-dependent shift away from the observation and generally the same shape. The ensemble size of 16 chosen for various impact experiments in following sections has a maximum

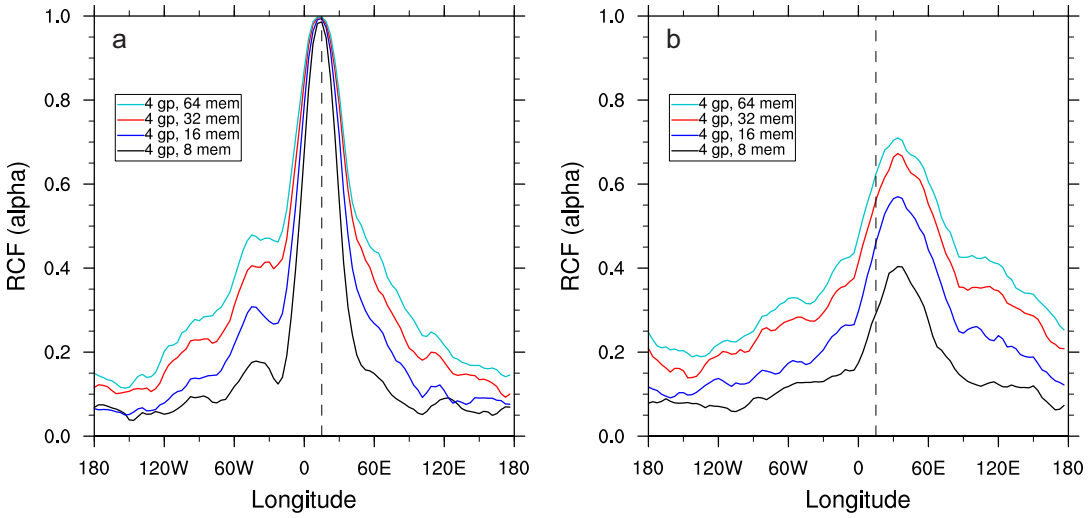

Figure 4.2. Zonal cross-sections of RCF, averaged over 900 cycles, for model interface height of a midlatitude observation located at 60°N, 15°E. The curves represent mean RCF functions calculated using g = 4 groups, with differing numbers of ensemble members per group – 8, 16, 32, and 64. The vertical dashed line represents the longitudinal location of the observation. (a) RCF function for the analysis (forecast t = 0), (b) RCF functions for a t = 2-day ensemble forecast

signal of about 0.55 for the 2-day forecast, in contrast to the maximum of 1.0 for the analysis RCF function (Figure 4.2a).

Figure 4.3 examines the sensitivity of changing the number of groups used in the RCF computation. The differences in number of groups results in small differences in the RCF functions for both the analysis and 2-day forecast. This insensitivity can be attributed to the process of taking a long-term average of RCF functions over all LETKF cycles. Interestingly, just two groups would be sufficient in capturing the general shape of the RCF function. The rest of the results consider RCF functions computed from four groups of 16 ensemble members per group.

Each observation within the domain has a unique RCF function, and because the method reveals dynamical features of the model not all observations have a Gaussian-like spatial correlation. The seven observations in Figure 4.4 each

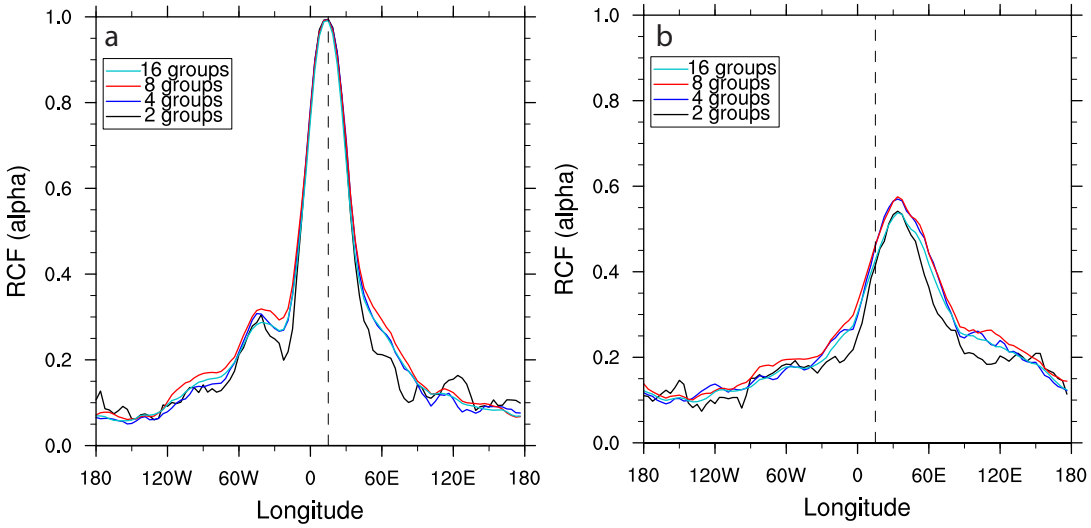

Figure 4.3. As in Figure 4.2., but for RCF functions calculated with varying numbers of groups, with 16 ensemble members per group.

show zonally-stretched RCF distributions, consistent with the predominantly zonal flow of the model. Midlatitude observations (obs. 1,2,3,6,7) tend to have a more Gaussian-like appearance, though some have a triple-peaked structure (obs. 1 and 6) associated with the strongest westerly flow. The distance between the peaks of about 40 degrees in longitude is likely a representation of underlying Rossby waves. For example, for an observation placed at a trough these additional peaks represent the adjacent ridges associated with a trough. As observations get closer to the tropics (e.g. obs. 4, 5), RCF begins to take a different, sometimes complex shape, including stretching eastward along the equator upstream of the main flow.

RCF functions from Figure 4.4 are valid for the analysis time of model interface height. As illustrated in Figure 4.5, RCF functions also reveal the time-forecast and cross-variable dynamics of the model. With increasing time from analysis to 3-day forecast, the RCF function in interface height (Figure 4.5a,d,g) shifts downstream of the observation, expands in area, and reduces in magnitude.

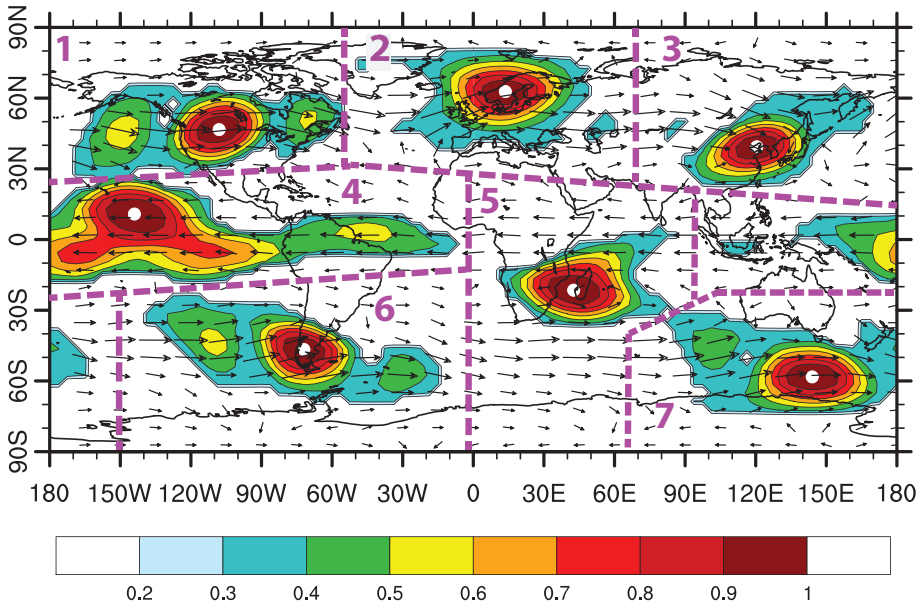

Figure 4.4. Examples of RCF functions for seven differing interface height observations (locations marked by white dots), calculated for analysis time (t =0) of model interface height. The wind vectors are a 900-cycle average of ensemble mean layer-2 wind. For plotting purposes, each observation's RCF function is displayed only for values greater than 0.3.

The shift in maximum amplitude is approximately 10° per day between analysis and 2-day forecast, which is roughly 1000-km per day. This is consistent with results of Torn and Hakim (2008) who found a 1000-km distance in their composite sensitivity 24-h patterns over Washington. In terms of layer two zonal wind, u_2 (Figure 4.5b) and meridional wind, v_2 (Figure 4.5c), the RCF functions exhibit dual peak dynamical structures. These RCF functions together mimic the shape of geostrophic adjustment correlations (e.g., Schlatter 1975) for both analysis and 1-day forecast. As forecast time increases to 2 days and longer, the RCF functions for cross-variables smooth out and lose definition in dynamical linkage, though they still show the time-forecast dependency. It is possible that the small 16-member ensemble is unable to resolve cross-variable correlations

beyond 2 days due to the weaker correlations, or other dynamical processes beyond advection occur past the 2-day lead-time forecast.

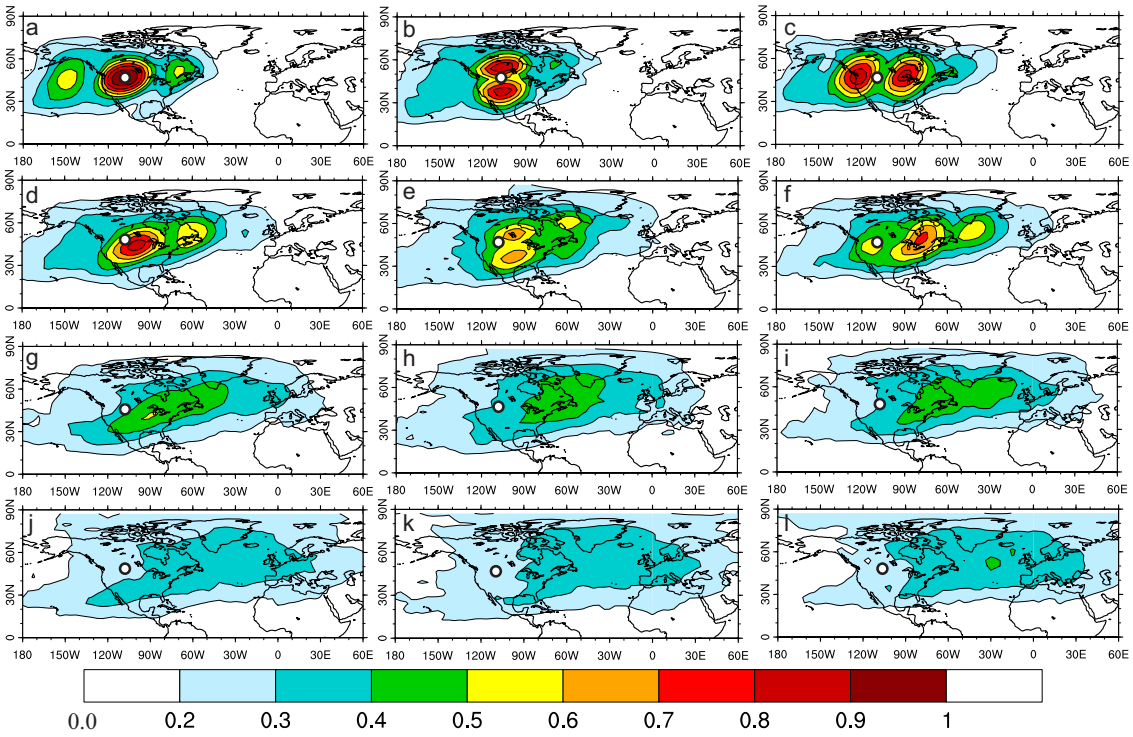

Figure 4.5. Contour plots of RCF localization functions for one interface height observation valid at various forecast lengths, (a-c) for analysis time, $t = 0$, (d-f) for t = 1-day forecast, (g-i) for t = 2-day forecast, and (k-l) for t = 3-day forecast. The first column shows RCF envelopes for model interface height, and the second and third columns show envelopes for cross-variables zonal and meridional layer-2 wind, u_{g2} and v_{g2}, respectively.

4.4.2 Single-observation impact experiment

A single-observation experiment was first conducted to examine the qualitative results of EFSO impact estimates with differing localizations, compared to the actual forecast error reduction. For the experiment, a one-day forecast of one of the all-observation LETKF analyses was chosen as the background for a single-observation analysis. This single interface height observation has an observation innovation of +500-m. Deterministic forecasts initialized from the

ensemble mean analyses and backgrounds were run to calculate actual impact according to (4.1) Ensemble forecasts of the single-observation analysis were run to calculate the ensemble estimated impact according to (4.3) with differing choices of localization. Two localizations were tested: the same optimal GC localization (8000 km) as was used during assimilation, and the dynamic RCF functions shown in Figure 4.5.

Results for one single-observation experiment are shown in Figure 4.6 for analysis, 2-day, and 4-day impact of model interface height. Note again that negative EFSO values (filled blue) imply positive impact. The actual impact at analysis (Figure 4.6a) follows flow-dependent structures of the background. Initially, three main centers are present at magnitudes above 500 m^2. The ensemble estimates (Figure 4.6b,c) are qualitatively similar, as both localizations capture these three main centers well.

The actual impact on 2-day and 4-day forecasts (Figure 4.6e,i) shows many more impact centers due to forecast error growth. These impact areas propagate along the main westerly waves predominantly located within the tightly packed interface height contours. Actual impact centers span a much greater zonal distance and can be seen at distances exceeding 8000-km from the observation. The ensemble impact estimate using static GC localization (Figure 4.6f,j) cannot capture these far away impact centers due to the limited length scale. Moreover, the magnitude of the estimated impacts of areas closest to the observation are much stronger than those shown in the actual impact, indicating the GC localization weight is too large there. Conversely, the largest area of actual positive impact in the analysis

(60ºN,75ºW) has now advected 30º and 75º eastward near the edge of the GC localization function in the 2-day and 4-day forecasts, respectively, where the localization weight is nearly zero. As a result, the GC localized impact value is underestimated compared to actual impact. These effects combined result in a global RMSE that is about the same or

greater than the RMS of the actual impact fields. The ensemble impact using dynamic RCF localization (Figure 4.6g,k) shows improved estimates, able to match the magnitudes of each center more closely with the actual error reduction. The localization functions spans a much greater distance than the GC function, with centers shifted downstream of the observation. Overall, the RMSE is much lower at less than half of global RMS of the actual impact.

This one case demonstrates both the qualitative nature of actual impact varying with forecast length, and how the RCF can outperform static GC localization and lead to improved estimates from the evolving RCF functions. In the next section, verification of the all-observation experiment is discussed to see if the RCF localization shows added overall skill in the case of homogeneous observation coverage.

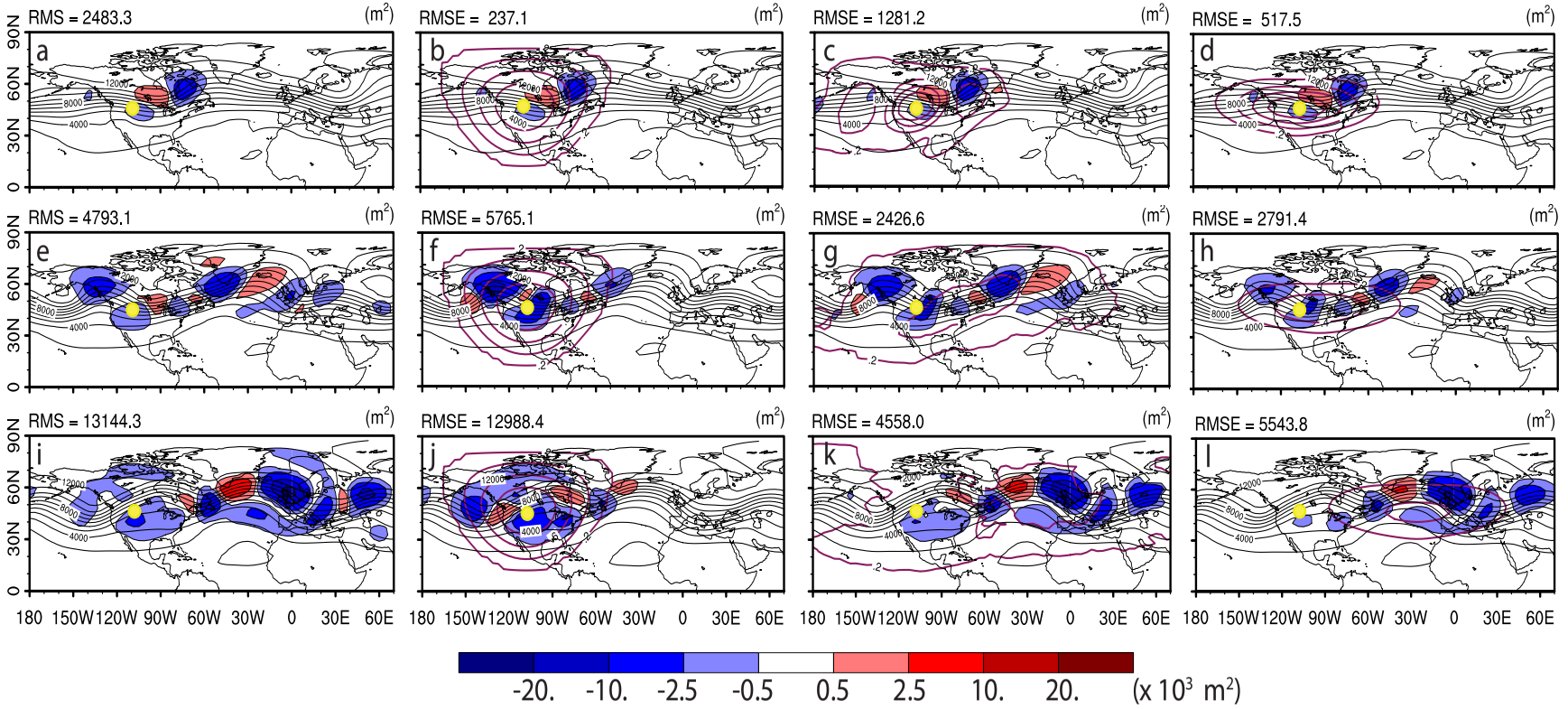

Figure 4.6. Contour maps of EFSO impact estimates from a single-observation (yellow dot) assimilation experiment. Each column, from left to right, is the actual impact (i.e. actual forecast error reduction), the ensemble-estimated impact using a static GC function with 8,000-km cutoff radius, and the ensemble-estimated impact using the RCF localization functions as pictured in Figure 4.5, and ensemble-estimated impact using automatically-tuned elliptical GC functions (see section 4.4.6), respectively. (a-d) Impacts at *analysis* (t = 0), (e-h) impacts on a t = 2-day forecast, (i-l) impacts on a t = 4-day forecast. Color-filled contours show impact values (m²), black contours are model interface height in 1000-m intervals, and magenta contour lines show the localization function applied to the impact estimate contoured in intervals of 0.2 starting at 0.2.

4.4.3 All-observation impact experiments

In all-observation impact experiments, each LETKF analysis ensemble was run for forecast lengths varied between 0 and 4 days. Actual error reduction was calculated using deterministic forecasts off the analysis mean and background fields. Ensemble-estimated impact was calculated by summing at each grid point contributions of the impact from all observations, with varying localizations applied. Figure 4.7 shows globally-averaged skill score (SS) of the time-mean RMSE of ensemble impact estimations verified against actual forecast error reduction. SS is defined as

$$SS = 1 - \frac{RMSE}{RMSE_{ref}} = 1 - \frac{\left[\sum_{k=1}^{N}\left(\Delta e^2_{actual,k} - \Delta e^2_{ens\ est,k}\right)^2\right]^{1/2}}{\left[\sum_{k=1}^{N}\left(\Delta e^2_{actual,k}\right)^2\right]^{1/2}}, \tag{4.4}$$

where N is the number of cycles in time considered (900 throughout this chapter). This SS is equivalent to that used in Kalnay et al. (2012) where the reference RMSE is equivalent to the time-mean RMS values of the actual forecast error reduction.

With increasing forecast time, for EFSO estimation of z_{int}, differences in SS emerge between GC and RCF localizations, with RCF localization showing increasingly higher skill. While the GC experiment approaches the no skill line (0.0) by day 4, the RCF experiment has a skill around 0.4, still higher than the skill of the GC experiment at day 2. At the analysis time, the SS of the fixed GC localization experiment is nearly the same as the RCF localization in impact estimates of model interface height. In terms of indirectly observed layer-2 meridional wind (Figure 4.7b), at analysis time the RCF experiment shows slightly

lower skill than the GC experiment, but with increasing forecast time the RCF experiment becomes increasingly skillful relative to the GC experiment. The slight degradation in skill at the analysis time can be attributed to the inconsistency between the localization used in data assimilation and the localization used for the EFSO estimate. This issue is discussed further in section 4.4.4.

In addition to SS using time-mean RMSE, Figure 4.8 considers SS where global-mean RMSE is calculated in (4.4) and N is now the number of grid points (4608). The result is a unique global SS for each cycle in the experiment.

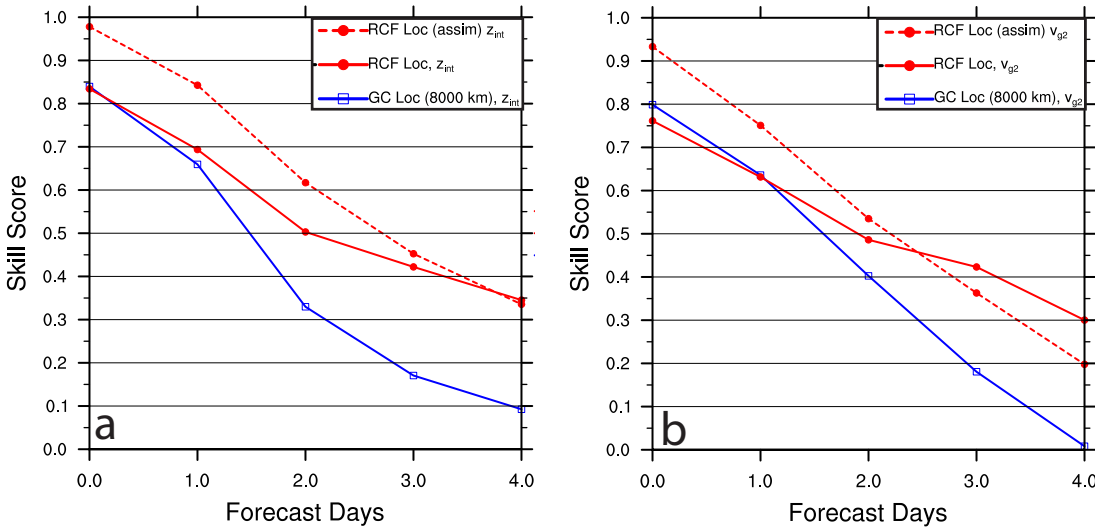

Figure 4.7. Global-average skill score of EFSO estimates verified against actual error reduction. Red lines are impact estimations using RCF localization; blue lines are estimations using static GC localization (8000 km). Solid lines valid for assimilation experiment using 8000-km GC localization during assimilation, dashed lines valid for assimilation experiment using analysis RCF functions as localization (RCF-assim, see section 4.4.5). (a) Skill scores for impact on model interface height impact, and (b) Skill scores for layer two meridional wind.

The bar chart in Figure 4.8a shows the percentage of the 900 cycles that have positive global skill for interface height EFSO. At the analysis time and 1-day forecast, both RCF and GC localization have a high percentage of cycles showing positive global SS. But as in Figure 4.7, at 2-day forecast and beyond the differences become larger. The RCF function still shows positive skill more than three-quarters of the time at the 4-day forecast length, which is still better than even the GC function at 2-day forecast impacts. Of course, positive skill may not necessarily mean much when the skill is close to zero, so Figure 4.8b shows instead percentages of cycles with SS exceeding 0.5. With the increased threshold, at 2-days and beyond the RCF still outperforms the GC by more than 20%. By 4-day forecast, with a stricter threshold the GC function is not any more favorable than a simple top-hat localization with the same length scale, whereas the RCF shows high skill for nearly a quarter of the cycles considered.

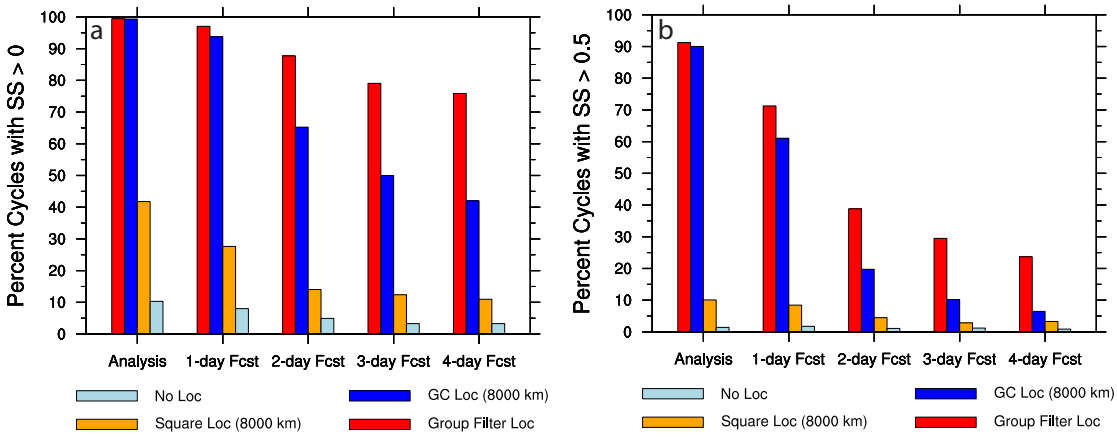

Figure 4.8. Bar charts showing percentage of cycles with skill scores of ensemble observation impact estimates greater than (a) 0.0 and (b) 0.5, for no localization (light blue), square or top-hat localization with 8000-km cutoff (orange), GC localization with 8000-km length scale (blue), and GF localization (red).

The previous results for GC localization operate under the assumption that the same localization applied during assimilation is applied towards the impact estimate. What if, instead, we were allowed to optimally tune the localization length of the GC function for each forecast length? The GC tuning of impact estimates is shown in Figure 4.9 for the analysis, 1-day and 4-day forecast lengths. At the analysis, it is verified that the same tuning (8000-km) used at assimilation also provides optimal ensemble impact estimates. With forecasts, potential improvement by tuning the GC width is relatively minor, especially compared to the improvement that the RCF localizations show (dots in Figure 4.9).

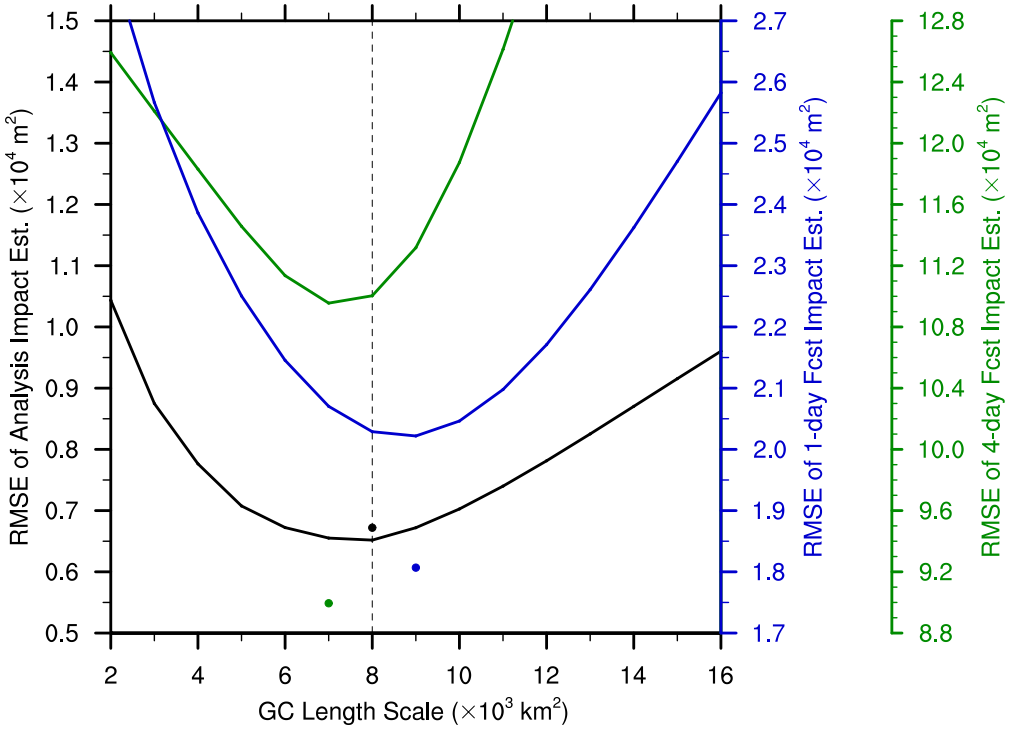

Figure 4.9. Optimal tuning of the GC length scale for EFSO at analysis (black), 1-day forecast (blue) and 4-day forecast (green), verified against actual error reduction and averaged over 48 cycles. Colored dots indicate RMSE of ensemble impacts using RCF localization, plotted vertically from its respective optimal RMSE using GC. Vertical dashed line indicates the GC length scale used for assimilation (8000 km).

A similar set of experiments varying the GC radius was conducted in Sommer and Weissmann (2014), and they too concluded that simple adjustment of the GC radius did not lead to a large improvement in accuracy of the ensemble impact method.

Finally, we examine the effect of RCF localization varying by latitude, since RCF functions at midlatitudes and near the equator are very different in shape (Figure 4.4). Here the pattern or map correlation is considered, which is the Pearson correlation calculated between the ensemble-estimated and actual impact at the same location over all 900 cycles. The result is a map of correlation values, and in Figure 4.10 the zonal average of this correlation map is shown. For 1-day forecast impact, both fixed GC and RCF show correlations well above the no localization case, as expected due to sampling error in the raw ensemble correlations. RCF shows slight improvement relative to GC at mid-latitudes (30-70^{o}) and the same or slight degradation elsewhere. At the 3-day impact (Figure 4.10b), the RCF shows improvement over the fixed GC localization for most latitudes with the largest improvement at the midlatitudes. On the other hand, GC shows low skill comparable to no localization, an indication that while localization is needed, the GC localization is too restricted and centered in the wrong spot. A distinct shape is present at both 1-day and 3-day lead times, showing high values at midlatitudes with a large dip within 30^{o} of the equator. The improvement of GC and RCF relative to no localization is much smaller in tropics than midlatitudes. In the next section we further investigate this issue.

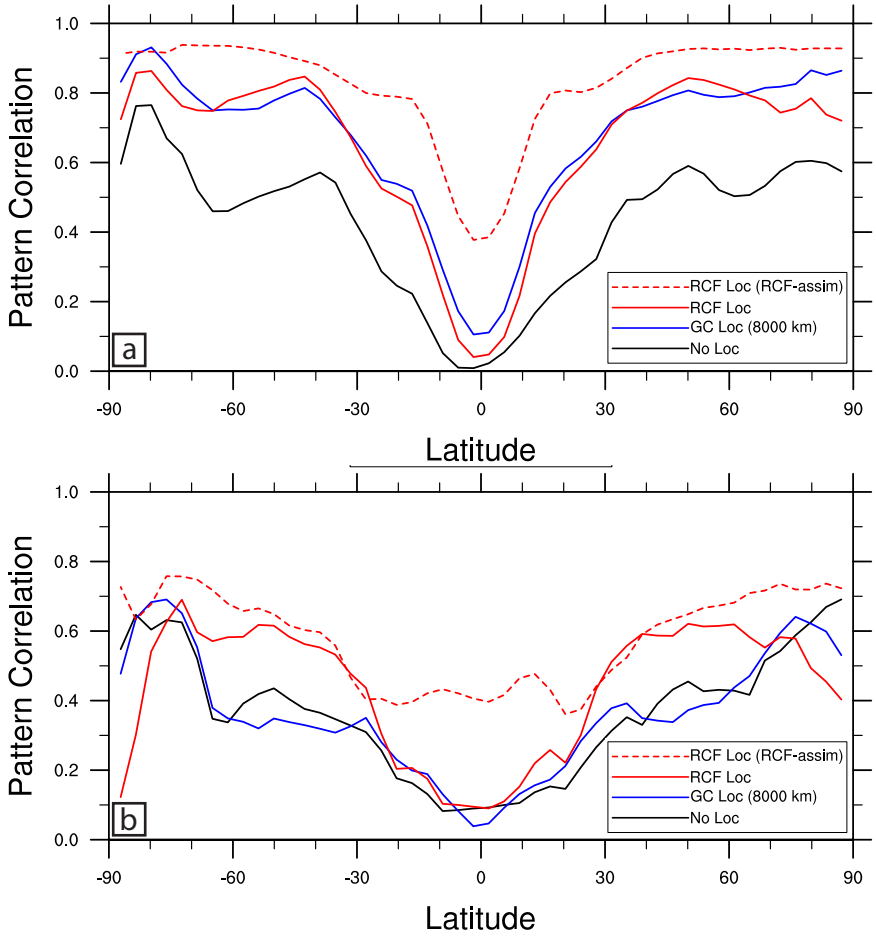

Figure 4.10. Zonally-averaged pattern/map correlation of EFSO in model interface height compared to actual impact, for no localization (black), GC localization (blue), and dynamic RCF localization (red). Dashed red line shows correlations of estimated impact using RCF localization at analysis time during assimilation (RCF-assim, see section 4.4.5 for further details) (a) For 1-day forecast impact, (b) For 3-day forecast impact.

4.4.4 Relationship between localizations for data assimilation and for EFSO

The improvement in skill in the 1-4 day forecast range is due to the time-shift dependency. Two issues with the RCF localization applied to EFSO were noted. The first is why RCF did not show improved performance at the tropics, as seen in Figure 4.10. The second, as seen in Figure 4.7, is that the cross-variable dynamical linkage revealed within RCF did not offer any improvement to skill at

analysis time. Both issues are related to the fact that localization used during assimilation was not consistent with the localization that the RCF suggested for the EFSO metric.

First, it is important to note the uniqueness of the equatorially-stretched RCF functions in the tropics (Figure 4.11). The two-layer model used here, while able to well represent realistic midlatitude baroclinic instability, is not a realistic representation for the tropics due in part to the simplifying exclusion of moisture. Hendon and Hartmann (1985) analyzed the variability of a similar dry two-layer model and noted that the tropics are dominated by internal normal modes consistent with Matsuno (1966). Additionally, these waves are equatorially trapped, with minimal activity propagating from tropics to midlatitudes. The RCFs of an equatorial observation are also equatorially trapped and show east- and westward propagating components (Figure 4.11c). So tropical RCF functions represent the internal normal mode dynamics of the two-layer model used throughout this chapter.

The effect of inconsistent localization functions for the impact estimate compared to the localization used during data assimilation can be shown within the derivation of Kalnay et al. (2012) together with a single-observation experiment for an equatorial observation using 8000-km GC localization during assimilation (Figure 4.12). A basic EnKF mean update is of the form $\overline{x}^a - \overline{x}^b = \mathbf{K}\delta\mathbf{y}$. Assuming that localization, ρ_A, was applied to the EnKF in some way, then it is contained within **K** such that it and mean update $\overline{x}^a - \overline{x}^b$ both go to 0 as ρ_A goes to 0 (i.e. no update of the background). This is evidenced by the actual impact field at

analysis time shown in Figure 4.12a, which covers the same extent as the 8000-km GC localization shown in Figure 4.11a. In Kalnay et al. (2012), the derivation of EFSO was based on substitution of post-analysis expression for gain $\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} = (k-1)^{-1}\mathbf{X}^a\mathbf{X}^{aT}\mathbf{H}^T\mathbf{R}^{-1}$ into $\mathbf{MK}\delta\mathbf{y}$ (where $\mathbf{M}$ is the tangent linear model). In order for the post-analysis formulation of gain $\mathbf{K}$ to be consistent with the EnKF update using some localization matrix ρ_A, then that localization function also has to be present in the post-analysis formation,

$$\mathbf{K} = \mathbf{P}^a_{loc}\mathbf{H}^T\mathbf{R}^{-1} \tag{4.5}$$

where $\mathbf{P}^a_{loc} = (\mathbf{I} - \mathbf{KH})\left(\rho_A \circ \mathbf{P}^b\right)$ is the "localized" analysis error covariance matrix. In other words, for the impact metric we need find a localization ρ_I so that

$$\mathbf{P}^a_{loc} = \rho_I \circ \mathbf{P}^a \tag{4.6}$$

where $\mathbf{P}^a = (k-1)^{-1}\mathbf{X}^a\mathbf{X}^{aT}$. It is not straightforward to find a localization function ρ_I to satisfy (4.6) . However taken together, (4.5) (4.6) indicate that we should expect the optimal localization choice for the impact metric ρ_I to be related to the localization used during assimilation ρ_A This is consistent with why the GC estimate (Figure 4.12b) is qualitatively close to the actual impact (Figure 4.12a) while the RCF estimate (Figure 4.12c) is not, since the GC function is the same as was used during assimilation whereas RCF is a completely different shape. For the case studied in Figure 4.11-Figure 4.12, because little activity propagates from the tropics to midlatitudes, the 8000 km GC function used during assimilation may be inappropriate for this observation, incorrectly adjusting midlatitude locations. Instead, localization consistent with RCF should be used during assimilation.

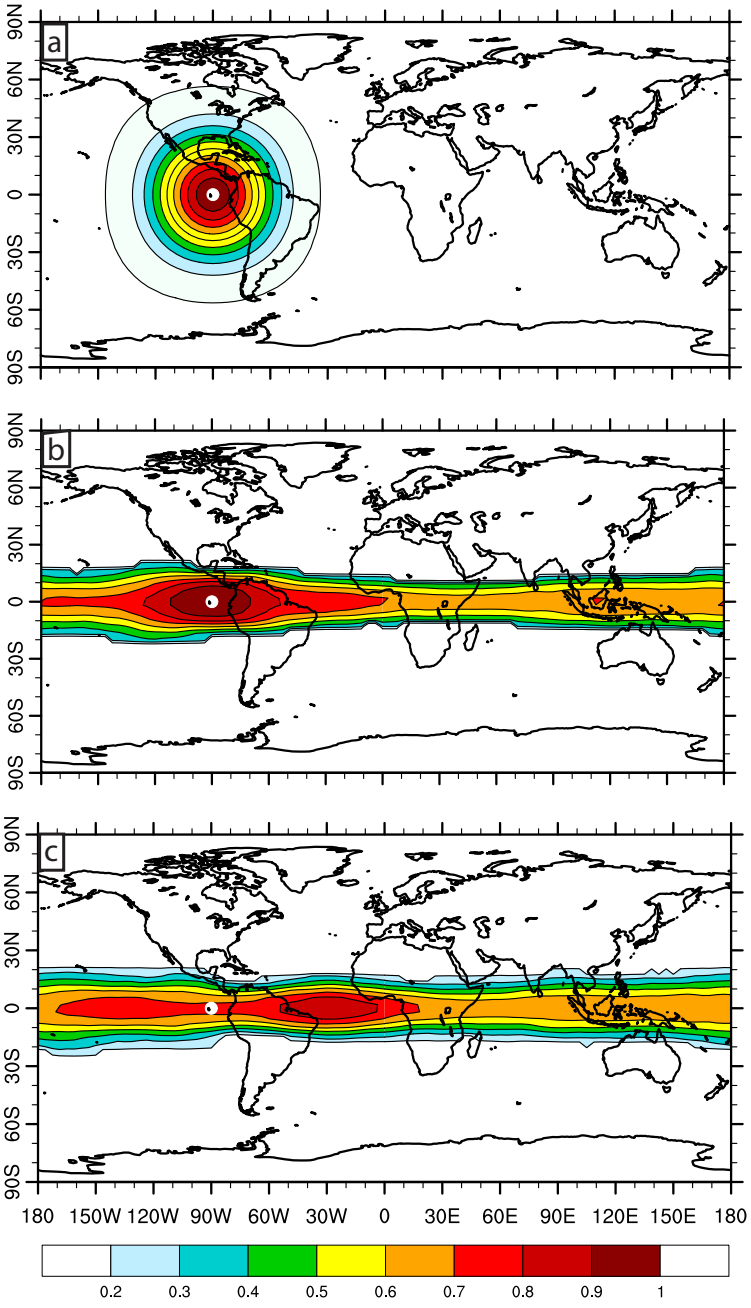

Figure 4.11. Localization functions for an equatorial observation (white dot): (a) GC localization function with 8000-km length scale (outermost ring is contoured at 0.01), (b-c) RCF functions valid for analysis (t = 0) and 2-day forecast, respectively.

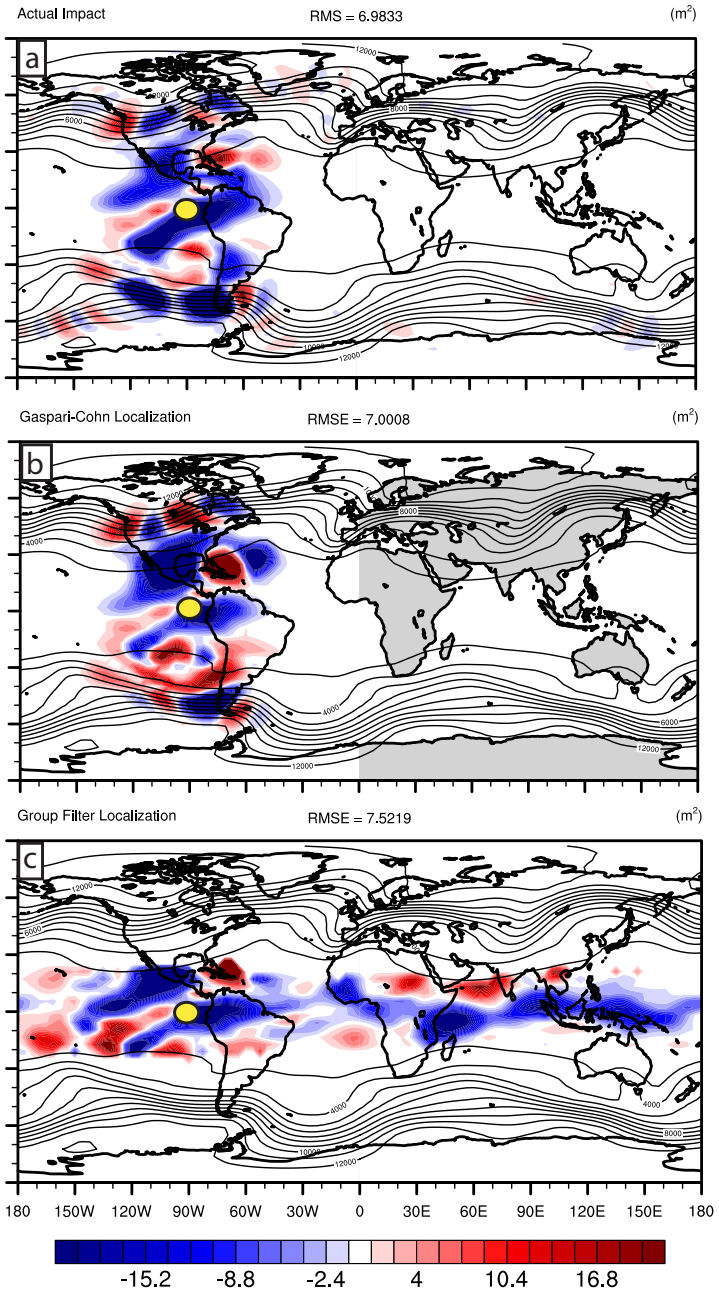

Figure 4.12. Single-observation impact experiment for an assimilated observation located at the equator (yellow dot). (a) Actual impact, or actual error reduction, of observation at analysis time, (b) Ensemble estimate of impact using GC localization (8000 km), and (c) Ensemble estimate of impact using RCF localization. Color-filled contours show impact values (m^2) and black contours are model interface height in 1000-m intervals

The implicit relationship between ρ_I and ρ_A may also explain why the dynamical linking of cross variables in the RCF function did not show improved impact estimates at the analysis (Figure 4.7), because it is simply a different shape than what was used at assimilation time. It also helps explain why the dynamic RCF method was successful at midlatitudes, because the initial RCF function at the analysis time was similar to the 8000-km GC function used during assimilation, allowing for the added benefit of the time-evolving component for forecasts.

Further discussion on this implicit relationship between localizations can be found in Appendix B.

4.4.5 All-observation experiment using RCF localization during assimilation

The inherent relationship between ρ_I and ρ_A suggests that the best use of adaptive localization for the impact metric occurs when the same adaptive method is used during assimilation. To test this hypothesis, a new all-observation assimilation experiment is conducted where the analysis RCF functions (e.g. Figure 4.4a,b,c) are used during the assimilation as localization, ρ_A. This experiment will be referred to as "RCF-assim". Impact estimates are calculated with the RCF functions applied as localization and verified against the actual forecast error reduction, as in previous experiments. Pattern correlation and SS is compared to the results of the previous experiments using GC localization during assimilation.

The SS in Figure 4.7 shows that RCF-assim yields improved skill compared to the previous application of RCF to the GC assimilation, particularly for shorter (0-2 day) forecasts. This improvement converges to the level of skill from the previous

RCF experiment at longer lead times for interface height. In terms of cross-variable v_2 impact, RCF-assim shows higher skill than the previous RCF experiment for 0-2 day forecasts, but actually becomes less skillful for 3-4 days. It is possibly an indication that RCF should also be re-calculated from the new RCF assimilation. Comparing pattern correlations of RCF-assim to the previous RCF experiment, there is clear improvement at all latitudes, including significant improvement in the tropics at both 1-day and 3-day forecasts (Figure 4.10). The results of RCF-assim confirm that the best use of adaptive localization methods for the ensemble impact metric occurs when the same adaptive method is used for localization during the assimilation.

4.4.6 Using RCF to design automatically-tuned elliptical GC localization functions

Another potential use of adaptive methods such as the RCF is its derived shifting, magnitude change, and area coverage of the localization that could be used in the future to automatically tune a GC function for localization of EFSO. For example, Figure 4.13 shows four potential tuning parameters that may be calculated from RCF and how they evolve with increasing forecast time. Given predominantly zonal flow, each parameter was tuned for observations binned by latitude. One can see, for example, the differences in how far to advect a GC localization away from an observation due to differences in the zonal wind strength at each latitude (Figure 4.13a) and the similarities at each latitude in reduction of maximum magnitude with increasing forecast time (Figure 4.13b). It can also be seen that localization area should expand with increasing forecast time,

but at different amounts depending on latitude of the observation (Figure 4.13c). Complimentary to that, the localization function should have more of a two dimensional elliptical shape, since the zonal extent is greater than the meridional extent (Figure 4.13d).

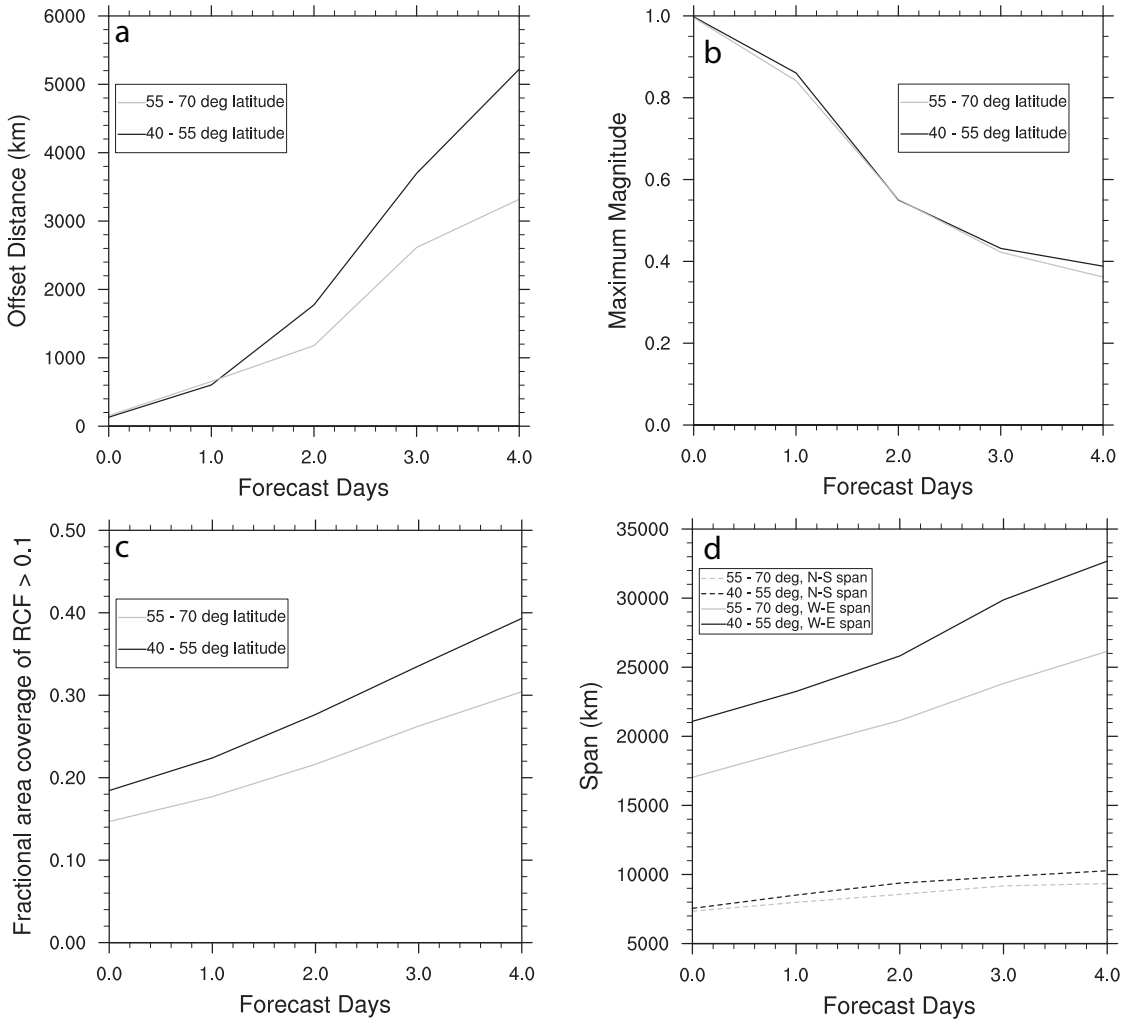

Figure 4.13. Summary plot of potential localization tuning parameters for the impact estimation as a function of forecast time, derived from RCF functions in model interface height. (a) Offset distance from observation, or shift, defined as the distance between RCF maximum and the observation, (b) Reduction in magnitude with increasing forecast time, (c) Fractional surface area coverage, (d) Maximum span, or range, of RCF values in zonal direction

(solid) and meridional direction (dashed). Values are averaged for all observations within 40 – 55 ^{o}N (black) and 55 – 70 ^{o}N (grey).

One question is how to apply these parameters with the preexisting definition of the GC function from Gaspari and Cohn (1999) to create an elliptically shaped localization. In the homogeneous or 'circular' case, the GC function has the same cutoff distance in all directions; however in the elliptical case, the cutoff distance changes based on the angle relative to the major and minor axes of the ellipse.

The steps to defining an elliptical GC localization for the impact metric are as follows:

1) Calculate parameters from the RCF functions as in Figure 4.13. Parameters needed are the zonal span, meridional span, maximum magnitude, and offset distance from observation.
2) Use offset distance to define the center of the ellipse for a given observation and forecast time.
3) At each grid point, calculate the angle θ relative to the center of the ellipse using $\theta = \arctan(\Delta y/\Delta x)$, where Δx and Δy are the zonal and meridional component distances, respectively, between the grid point and the center of the ellipse.
4) The parametric form of an ellipse is given by $x = a \cos t$; $y = b \sin t$, where a and b are major and minor axis distances, respectively, and t is the parametric angle. In this case, a and b are simply half of the zonal and meridional spans. Convert from polar angle θ to parametric angle t using

$$t = \arctan\left(\frac{a}{b}\tan\theta\right) \tag{4.7}$$

5) The cutoff distance for the GC functions as a function of angle is given by the distance formula

$$c = \sqrt{a^2\cos^2 t + b^2\sin^2 t} \tag{4.8}$$

6) Once the cutoff radius *c* is determined, the GC function can be applied directly. To adjust for reducing magnitude with forecast time, multiply the output weights from the GC function by the maximum magnitude parameter. Figure 4.14 displays the resulting elliptical GC localization functions using the parameters from Figure 4.13.

This new elliptical GC localization function was tested with the single-observation impact experiment. As shown in the rightmost column of Figure 4.6, the estimates are similar in performance to the RCF in terms of overall structure and RMSE. More precise definitions and applications of these derived parameters may yield better results, but are beyond the scope of this dissertation. Here it is presented as a proof-of-concept to demonstrate how such parameters can be calculated from RCF and applied to tune the GC function.

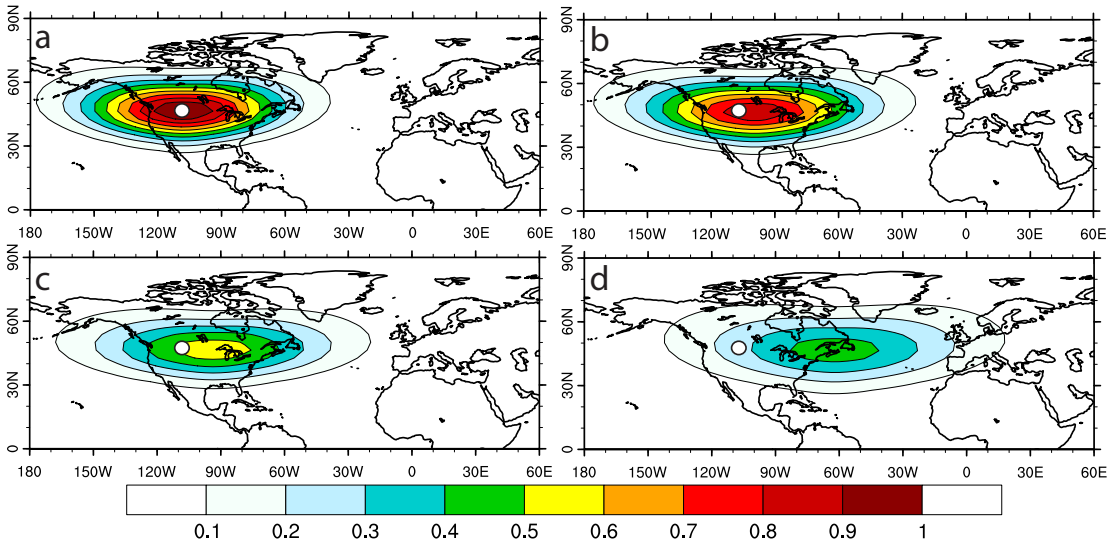

Figure 4.14. Automatically-tuned elliptical GC localization functions for (a) analysis, (b) 1-day, (c) 2-day, and (d) 3-day forecasts. The parameters used for the tuning are shown in Figure 4.13. Results of test application of these automatic GC localizations to EFSO estimate are shown in the rightmost column of Figure 4.6.

4.5 Conclusion and discussion

The real-world application of the Ensemble-based Forecast Sensitivity to Observations (EFSO) metric for evaluating observational impact needs to consider proper implementation of localization to reduce errors due to limited ensemble members. Such localization needs to consider time-forecast dependency in addition to spatial and cross-variable dependencies. Kalnay et al. (2012) and Ota et al. (2013) have shown that initial attempts to 'advect' the localization function downstream from an observation leads to more reliable and accurate impact estimates. Here we examined an alternative method, an adaptive localization that varies by location and state variable according to the underlying model dynamics. The method is based on confidence factors of grouped ensemble regression coefficients, first proposed for EnKF assimilation (Anderson 2007). In this chapter,

the method was extended to work for the ensemble-based observation impact estimate known as Ensemble-based Forecast Sensitivity to Observations, or EFSO. An envelope of *regression confidence factors,* or RCFs, for each observation and all grid points was used directly as localization. The purpose of the chapter was to explore the potential effectiveness of applying this adaptive method on the EFSO metric and to learn more about how to properly localize this impact estimate.

Results of the technique within a simple two-layer isentropic model showed the ability of the RCF method to reveal underlying dynamics between variables and the time-forecast component. Applying RCF functions for impact estimates showed overall improvement when verified against the actual forecast error reduction, especially at longer forecast lead times compared to using fixed GC localization. Single-observation experiments displayed the evolving structures of the actual forecast error reduction with increasing forecast time and showed that the dynamic localization was able to simulate the evolved error reduction much more closely than the static localization especially at longer forecast lead time. For an LETKF using all 362 observations, skill scores of EFSO estimates applying RCF localization significantly improved upon those using fixed GC for increasing forecast lead times. For example, the skill of the estimates of the impact using RCF on a 4-day forecast beats the skill of the estimate using static GC at the 2-day forecast, essentially doubling the forecast length of accurate impact estimates. This improvement was attributed mainly to the time-dependency of RCF shifting downstream, diminishing magnitude, and expanding in areal coverage.

Further diagnostics also found that optimal use of localization from dynamical methods relies on the assumption that consistent localization was used during the assimilation as well. In other words, there is a fundamental relationship between the localization applied at assimilation time and the localization used for the impact estimate. Therefore, further improvement in skill of using RCF with EFSO is possible if a consistent adaptive localization is used at the assimilation time in place of the static GC function. This was confirmed in an additional experiment using RCF localization both during assimilation and for the impact estimates. Skill scores were higher and significant improvement was made in the estimates at the topics.

Another potential utility of adaptive methods such as the RCF is to automatically tune an elliptical GC function for localization of EFSO. As a proof-of-concept, the application of RCF-derived parameters to tune the GC function shows promising results that were comparable to the RCF experiment for a single observation test.

For real-time application, a drawback of the RCF technique is computational cost of additional groups of ensemble members. To partially alleviate this issue, one could use increased ensembles only during a "training period" for a given weather regime to create RCF lookup tables offline, similar to the suggestion of Anderson (2007). Additionally, Figure 4.3 suggests that just 2 groups can yield RCF functions consistent with RCF using more groups, due to the averaging process over many cycles to reduce noise. Figure 4.3 also suggests another idea, which is to split available ensembles into subgroups without

increasing the total number of ensemble members and then to artificially "inflate" the resulting RCF functions, because with increasing groups the RCF functions retain similar shape and shift but differ only in magnitude. Alternatively, dividing existing ensembles into subgroups the RCF method can still be utilized to derive parameters for automatically tuning localizations.

Chapter 5: Application of EFSO to Convective-Scale Case Study

5.1 Introduction

In this chapter, the EFSO method will be applied to the same case study explored in chapter 3. In chapter 4, we learned that the localization function required for accurate EFSO estimates is inherently linked to the localization used during the DA, in addition to other dependencies including the time-forecast component. This adds a further layer of complexity to the proper localization application, though the dependency on the localization used during DA does constrain the problem. For example, we know that if a GC function of 200 km radius was applied for an observation, then we also know that the localization function to apply to the EFSO metric should be of similar shape and extent. This means a simple method such as the advected localization of Ota et al. (2013) may work well even for convective-scale DA.

Of course, the advected localization does not take fully into account other dependencies found using the RCF method on the two layer model in chapter 4. These dependencies include the stretching shape, the decreasing magnitude of the peak localization value, and cross-variable components. Additionally, it remains to be seen whether the assumption that advection fully captures the optimal time-forecast shift holds for more complex, nonlinear interactions found at the convective scales. For these reasons, it is important to explore the use of an adaptive method such as the RCF on the convective scale application of EFSO.

An additional consideration for the application of EFSO on the convective scale is how to choose a verification metric. Traditionally, based on the adjoint

method of Langland and Baker (2004), the formation of the actual impact metric includes an energy norm matrix **C**, such that equation (2.15) is written in the following form.

$$J = \left(\mathbf{e}_{t|0}^{T}\mathbf{C}\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}^{T}\mathbf{C}\mathbf{e}_{t|-n}\right) = \left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}\right)^{T}\mathbf{C}\left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}\right) \tag{5.1}$$

The energy norm matrix **C** is an *n* x *n* matrix of weights that converts errors in model state variables to energy in J kg^{-1}. This allows for the whole modeling system to be accounted for within the impact metric simultaneously. The most common energy norms used are the *dry* and *moist total energy norms* (Ehrendorfer et al. 1999). The total energy norm can be written in the following form.

$$TE = \frac{1}{2}\frac{1}{S}\int_{S}\int_{0}^{1}\left\{\left(u'^{2} + v'^{2}\right) + \frac{c_p}{T_{ref}}T'^{2} + \frac{R_d T_{ref}}{P_{ref}^{2}}p_s'^{2} + w_q\frac{L^2}{c_p T_{ref}}q'^{2}\right\}d\eta\, dS \tag{5.2}$$

Here u', v', T', p_s', and q' are forecast errors of zonal wind, meridional wind, temperature, surface pressure, and specific humidity, respectively. Constants c_p, R_d and L are the specific heat at constant pressure, dry air gas constant, and the latent heat of condensation per unit mass, respectively. T_{ref} and P_{ref} are constant reference temperature and pressure, respectively, commonly set to values of 280 K and 10^5 Pa. Weight w_q is a constant to define which energy norm is used (0 for dry total energy, 1 for moist total energy).

There is no requirement for a total energy norm to even be included in equation (5.1). In fact, in the previous chapter impacts were partitioned by variable, such that **C** was left off in the formula altogether. This is an important consideration for application to the convective scale because an integrated energy norm may not be appropriate to describe the impacts at these scales. In other

words, at convective scales we may be interested in only specific parts of the modeling system, such as the performance of convection described by just hydrometeor variables in the state (often represented as reflectivity). In the case of the CI event studied in chapter 3, we are mainly interested in the boundary layer and near-surface fields, given the strong dependence of CI on the resultant details of those fields. Using a total energy metric may hide some of these subtle details within a high-resolution forecast. Another example is focusing in a specific area of interest, such as the focus of impacts of observation within the DFW testbed only. For these reasons, multiple different verification metrics will be studied in addition to energy norms to see what differences exist when applying EFSO to different verifications other than integrated energy norms.

In section 5.2, the experiment setup will be discussed, including which verification metrics are used for this study and how the RCF localization is applied to this case study. In section 5.3, the resulting RCF functions are presented and analyzed. In section 5.4, the results of the application of EFSO are shown, including analysis of different localizations (static, advected, RCF-derived) and different verification metric. A summary and discussion are provided in section 5.5.

5.2 Experiment setup

The model and DA settings for this EFSO study are equivalent to the GSI-based EnKF DA and WRF forecast system described in chapter 3. Here, we will take the CNTL experiment from the 3 Apr 2014 case and use it for the EFSO

experiments. The EFSO method is the same as that used in Ota et al. (2013). It has been adapted to interface with WRF model input data.

Additionally, multiple verification metrics are tested for EFSO accuracy. Kinetic energy (KE), dry total energy (DTE), and moist total energy (MTE) norms are used as in (5.2). Note that the kinetic energy norm is made up of just the horizontal wind error components in (5.2) and setting the other components to 0. Other non-integrated metrics are used for verification testing. Given the strong dependence of CI on the features near the surface in the DFW domain (dryline characteristics), as well as the use of predominantly surface observations in the data denial study, the additional verification metrics will include near-surface model levels partitioned by state variable: p_s, u_s, v_s, T_s, q_s. These variables are defined as the first vertical mass level (k=1) above the surface with a height between 27 and 30 m AGL for the inner grid described in section 3.3.1. Finally, composite reflectivity MDBZ is tested as a verification metric, defined as the maximum value of reflectivity in the column at each grid point.

The EFSO will be computed at ten-minute forecast intervals between 0 and 120 minutes. As covered in section 2.3, the EFSO is designed to capture the impact of an analysis on the differences in deterministic forecasts initialized from the first guess and analysis means. This is a notable difference from the data denial work, which focused on analyzing the ensemble free forecast using ensemble probabilities.

The two-layer model experiments conducted statistics over 900 cycles of DA and free forecasts. The equivalent here would be to test over a significant

selection of similar convection case studies at high resolution, similar to 10 case studies used in Johnson et al. (2015). Given the application here to just a single case study, another method will be employed to get statistically meaningful results. The experiment setup described in Figure 3.6 includes 25 five-minute DA cycles over the 2-hour period from 1600 – 1800 UTC on 3 April 2014. Each DA cycle offers a relatively independent sample from which EFSO can be calculated for a given forecast verification time. In order to calculate impact for different forecast lengths from 0 to 120 minutes, two-hour free forecasts were run for each of the 25 ensemble analyses between 1600 and 1800 UTC. The EFSO metric is compared to the actual impact calculated using equation (2.15) for different verification metrics described previously. The actual impact in the case of the CNTL experiment represents the impact of *all* observations assimilated within each 5-min DA cycling window.

Multiple localization methods are applied to the EFSO metric and compared for accuracy against the actual impact. The first is static localization, applying the same GC localization as that used during DA for each observation type (see last column of Table 3.3). The next is advected localization, where the center of the GC localization is shifted proportionally to the average of the analysis and forecast horizontal wind at each vertical level, following Ota et al. (2013). The coefficient that multiplies the mean wind is chosen to be 0.75, equivalent to the optimal value found in Ota et al. (2013). A visual demonstration of 200 km GC localization and advected localization for a 30 min forecast is shown in Figure 5.1. Finally, a third localization function, the RCF or 'regression confidence factors' as introduced in

section 4.2, is applied to see if improvements can be made upon the accuracy of the EFSO metric. Details of the RCF method are discussed in the next section, with examples of different functions for different grid locations and forecast lengths.

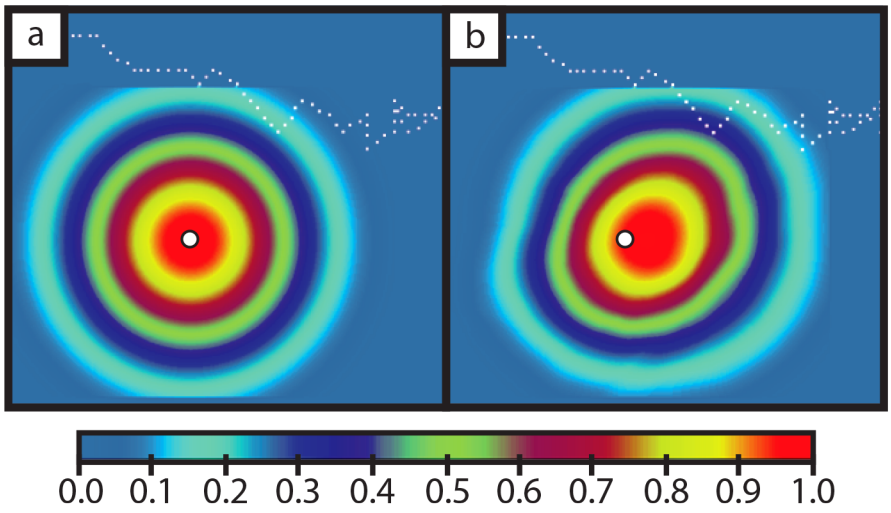

Figure 5.1. (a) Static 200-km GC localization function (b) Advected GC localization using *t* = 30 min forecast. White dot indicates location of observation.

5.3 RCF localization

5.3.1 Application

The RCF method introduced in chapter 4 is adapted for use within the GSI-based EnKF system. The method relies on independent groups of ensembles to capture the localizations for each observation. As such, the CNTL experiment was reran 3 additional times with 43 members each to create a group of 4 CNTL experiments, totaling 172 ensembles. These 3 additional ensembles were initialized by adding 3 random perturbations to the outer grid (12 km) initial conditions of each ensemble member (interpolated from SREF, GEFS, and NAM) at the beginning of the experiment, 0300 UTC 3 April 2014. Perturbations are also

added to each of the outer grid lateral boundary conditions in a similar manner. Each group is then run through the CNTL experiment setup as before (Figure 3.6). The RCF is then computed from observations assimilated at 1800 UTC for ten minute forecast intervals up to 120 minutes.

Within the two-layer model work, it was noted that the resulting RCF functions contain sampling error related to the limited number of groups used. The same is certainly true here, as can be seen in Figure 5.2a,b. To reduce noise from sampling error, RCFs across the 900 cycles were averaged together. The resulting RCF functions reflected information about the underlying dynamical correlations by model location and variable type (see Figure 4.4 and Figure 4.5). However, they did not reflect the underlying flow-dependent "errors of the day". Additionally, such an average would be difficult to accomplish at the convective scale, where cases are much more variable in nature and the observations that are assimilated vary from case-to-case and even cycle-to-cycle.

For the application to convective-scale, a different "binned" technique is used to limit the effect of sampling error in the RCF functions and more readily apply to different cycles no matter which observations are used. The model grid domain is split into latitude and longitude boxes or bins, and RCFs of all observations within each bin are averaged. Two different binning techniques are attempted here, based on the tradeoff of wanting to average a larger number of RCFs to reduce noisiness from sampling error while also maintaining important flow-dependent information valid for all observations in each box. The first is 2˚x2˚ binning, with observations for different variable types (p_s, *u, v, T, q)* averaged

separately, making for 5 different RCF functions for each bin. While this allows for treatment of different observed variables independently, the bins themselves cover a much larger area that may result in over-smoothing important flow-dependent information for sensitive high-gradient varying areas (e.g. near the dryline). The second is a 1˚x1˚ binning with RCFs averaged together from all observation variable locations and types in each bin (except surface pressure). This allows for more precise flow-dependent information, at the expense of losing variable-dependent information. It is not clear which method will yield better results, so both will be tested and compared with each other as well as static and advected methods. This binned technique is different from the averaging used in Lei and Anderson (2014), who binned RCF by radius from a given observation type.

For each observation in the EFSO estimate, it is first determined which bin the observation is located. The RCF function for that bin is directly applied as a localization function for that observation, with an applied threshold of 0.2 below which the RCF is considered zero. Additionally, the RCF is cutoff at a radius equivalent to that used during DA for each observation after re-centering using the time-shift estimated from advected localization. This process is demonstrated in Figure 5.2 for two observations and both bin types. It is motivated partially to further limit any remaining sampling error of the RCFs after averaging, but importantly to also have a localization function consistent with the results of the two-layer model study in chapter 4 (Gasperoni and Wang 2015). Since we know that localization for EFSO is implicitly related to the localization used during DA,

then it is logical to make sure the RCF for each observation also has some knowledge of the observation-dependent GC length scale used during DA. The hypothesis here is that applying a cutoff radius will add back the dependence of DA localization into the RCF localization.

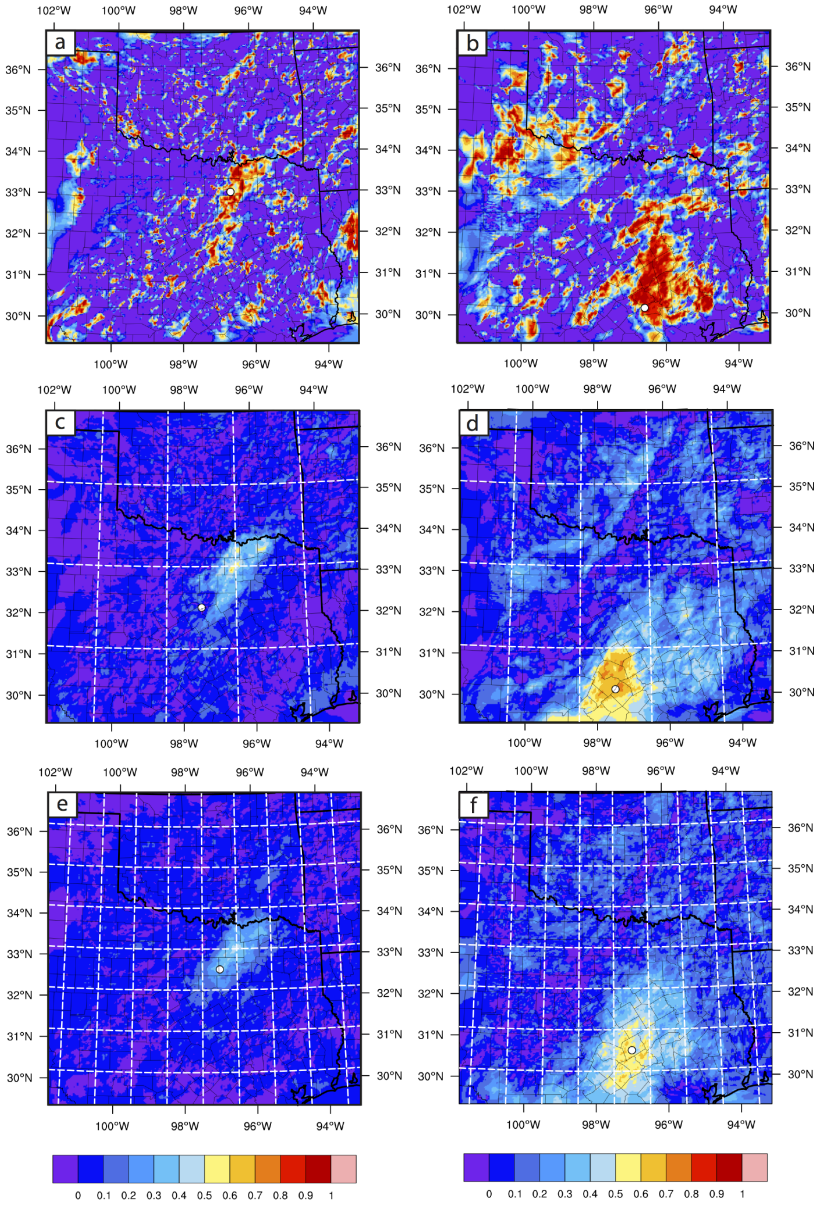

Figure 5.2. (a-b) Examples of RCF functions computed for the observation locations shown by the white dots, valid for 60-minute forecast. (c-d) 2˚x2˚ bin-averaged RCF for the bins corresponding to the observation locations in (a) and (b), respectively. White dotes correspond to the central location for each bin. (e-f) As in (c-d) but for 1˚x1˚ bins.

5.3.2 Analysis of resulting RCF functions

Figure 5.3 shows examples of RCF functions for different variable types in a given bin. There is a clear difference between dynamic wind observation RCFs and the thermodynamic RCFs both in magnitude and general areal coverage. Additionally, the RCFs in temperature and moisture both mark important underlying characteristics about the flow, tracing the warm, dry air west of the dryline and south of the cold/stationary front. The RCF for this observation is different from the two locations shown in Figure 5.2, as in general it covers a larger area than the previous examples.

The time-forecast component is also revealed within RCF functions, as can be seen in Figure 5.4. The RCF is initially centered on the bin, but advects north with the rich southerly flow in the moist air in East Texas. Additionally, the maximum magnitude diminishes, consistent with RCFs shown in chapter 4 for the two-layer model study. It should be noted, however, that these are just a few examples and the number of different RCFs is very diverse. Some RCFs maintain high magnitude or even slightly increase with forecast time (not shown).

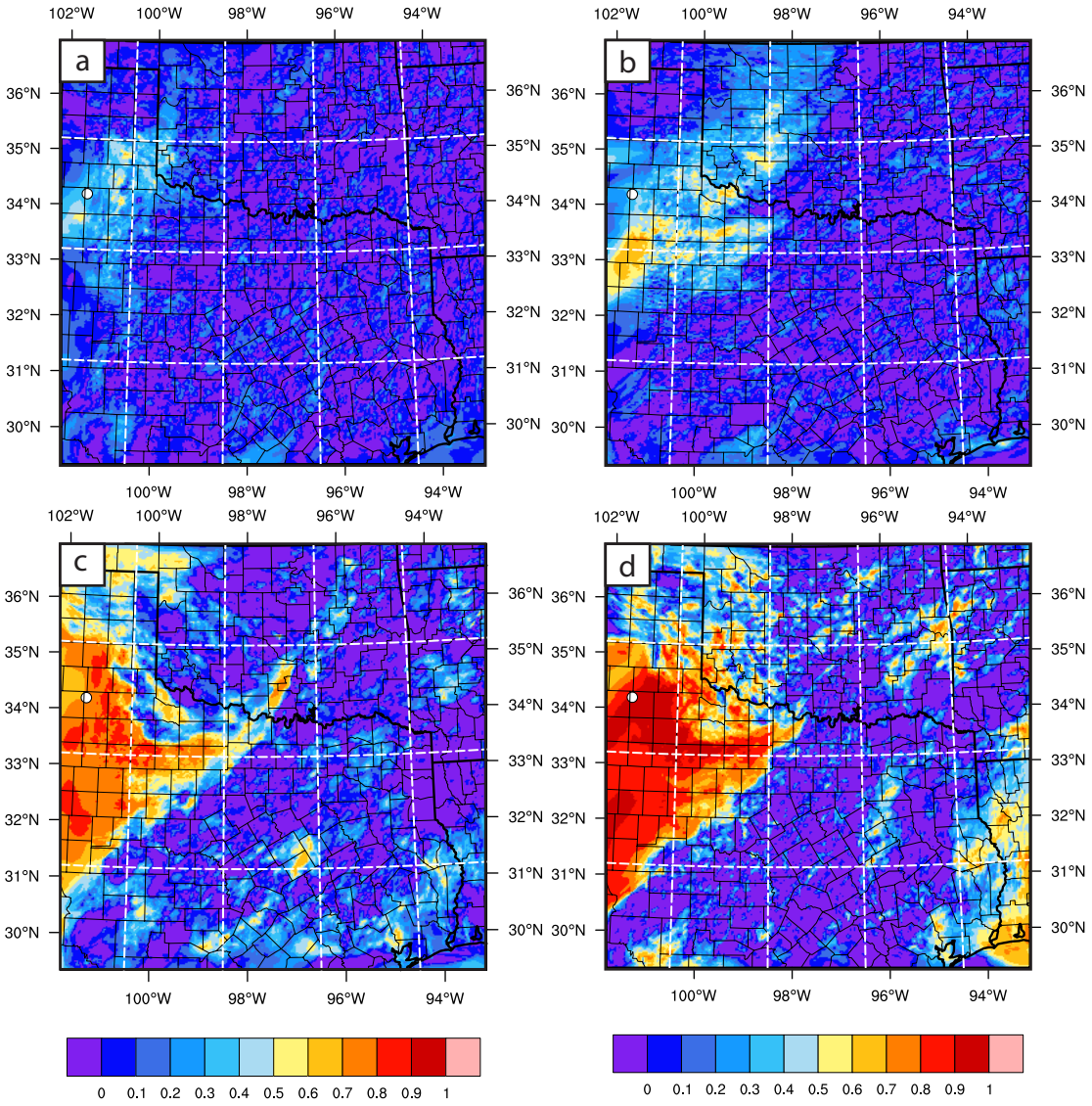

Figure 5.3. Bin-averaged RCF functions for different observation variable types: (a) zonal wind, (b) meridional wind, (c) temperature, and (d) specific humidity. White dot indicates central position of the location of the bin.

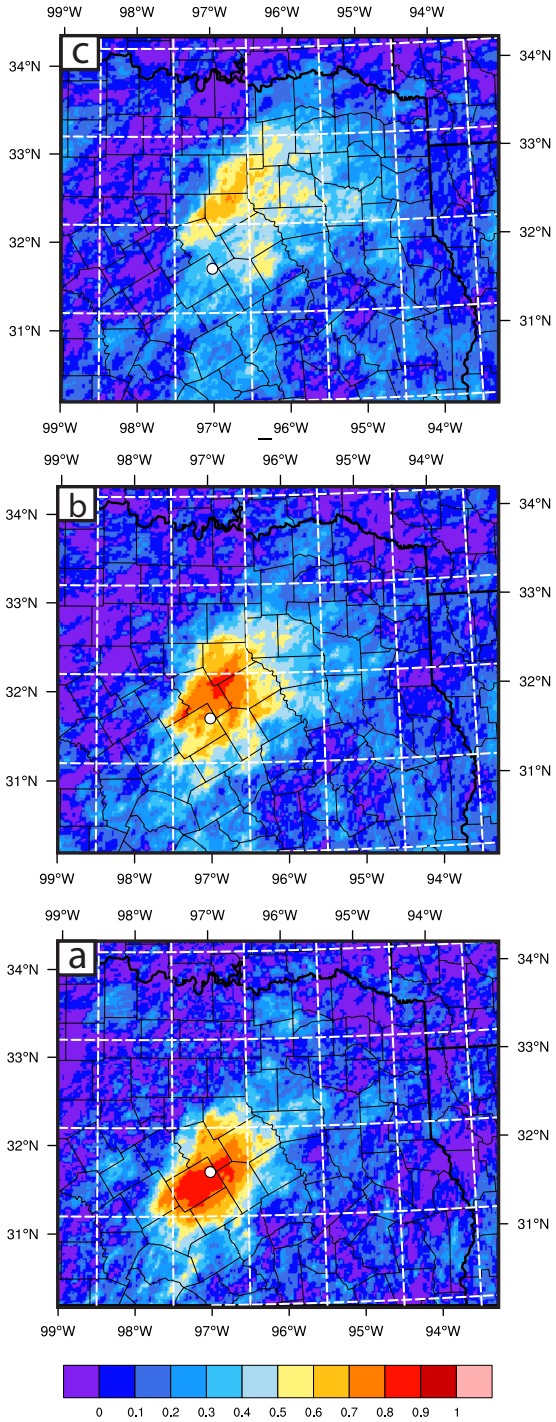

Figure 5.4. Demonstration of the time-forecast component for bin-averaged RCF function. White dot indicates central position of the location of the bin. Times shown are (a) Analysis (0-min fcst), (b) 60-min fcst, and (c) 120-min fcst.

5.4 EFSO results

5.4.1 Comparison of verification metrics

Two-dimensional patterns of impact are shown in Figure 5.5, where each grid point value is a vertical integral of energy components from the MTE formula. Initially, at analysis time, the EFSO and actual impact patterns match very closely, with some minor differences. At 60-min forecast time, the patterns begin to diverge, though the general locations of large impact are the same. At this time, convective activity is forming ahead of the dryline in Texas and the cold front in Oklahoma, which manifests as strong areas of small-scale positive and negative impacts along those areas. The EFSO metric is underestimating the impact in MTE along the cold front in Oklahoma, but captures the areas of impact at the OK-TX border and south into the DFW domain well. At 120-min forecast time, the estimate appears to have lost most of its skill except for a few small areas. Additionally, large-scale areas of negative impact appear in Oklahoma and the Texas panhandle. These areas of impact are associated with the surface pressure term in the MTE integration. The issue related to verifying pressure is discussed further in section 5.4.3.

In addition to qualitative comparison, it is important to summarize the data in a statistical sense. Given the patterns in Figure 5.5, one way to evaluate the usefulness in EFSO is to calculate the pattern correlation for the maps. This way we can determine if, regardless of potential magnitude differences, the EFSO is capturing the meteorological patterns shown in the actual impact field.

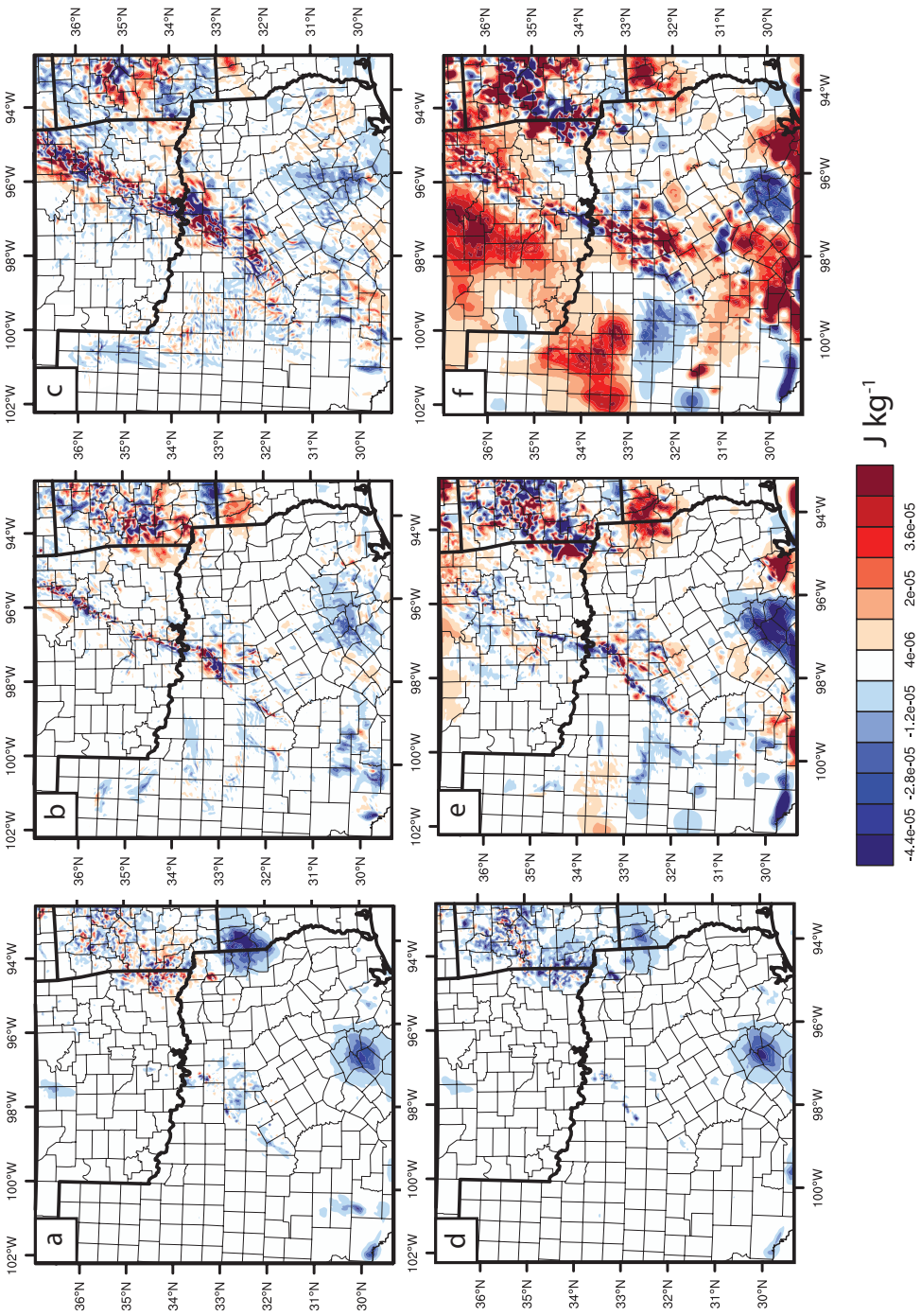

Figure 5.5. Actual impact (top row, a-c) and EFSO estimate with static localization (bottom row, d-f) in terms of moist total energy. Three times are shown: analysis t=0 (a,d), 60 min forecast (b,e), and 120 min forecast (c,f).

Figure 5.6 shows pattern correlations averaged over all 25 DA cycles, for different forecast valid times, for energy norm metrics. The correlation starts off high at t=0 but drops off quickly for each metric. The use of advected localization does improve the correlations at 10-30 minute forecast times, but beyond that correlation is too weak (below 0.3).

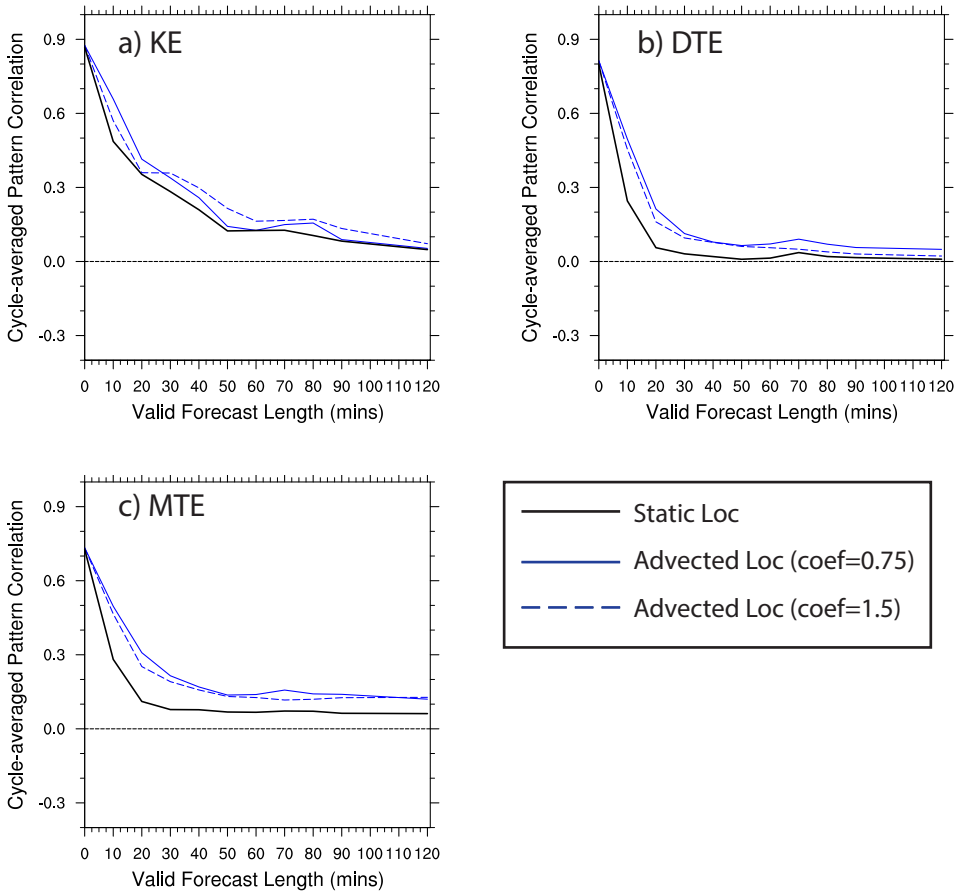

Figure 5.6. Pattern correlation of EFSO estimate compared to actual impact, averaged over the number of cycles available (25) for kinetic energy (a), dry total energy (b), and moist total energy (c). Black lines indicate static GC localization and blue lines indicate advected localization with weighting coefficient of 0.75 (solid) and 1.5 (dashed).

Correlations for surface energy metrics are shown in Figure 5.7. In contrast to energy norms, the surface variables have a slower drop off in correlation with forecast time. Using a weak relationship correlation threshold of 0.3, the EFSO for surface variables is accurate up to 50 minutes, even as high as 80 minutes for surface moisture. Additionally, surface thermodynamic variables match for longer forecasts (70-80 minutes) than either wind component (50-60 minutes).

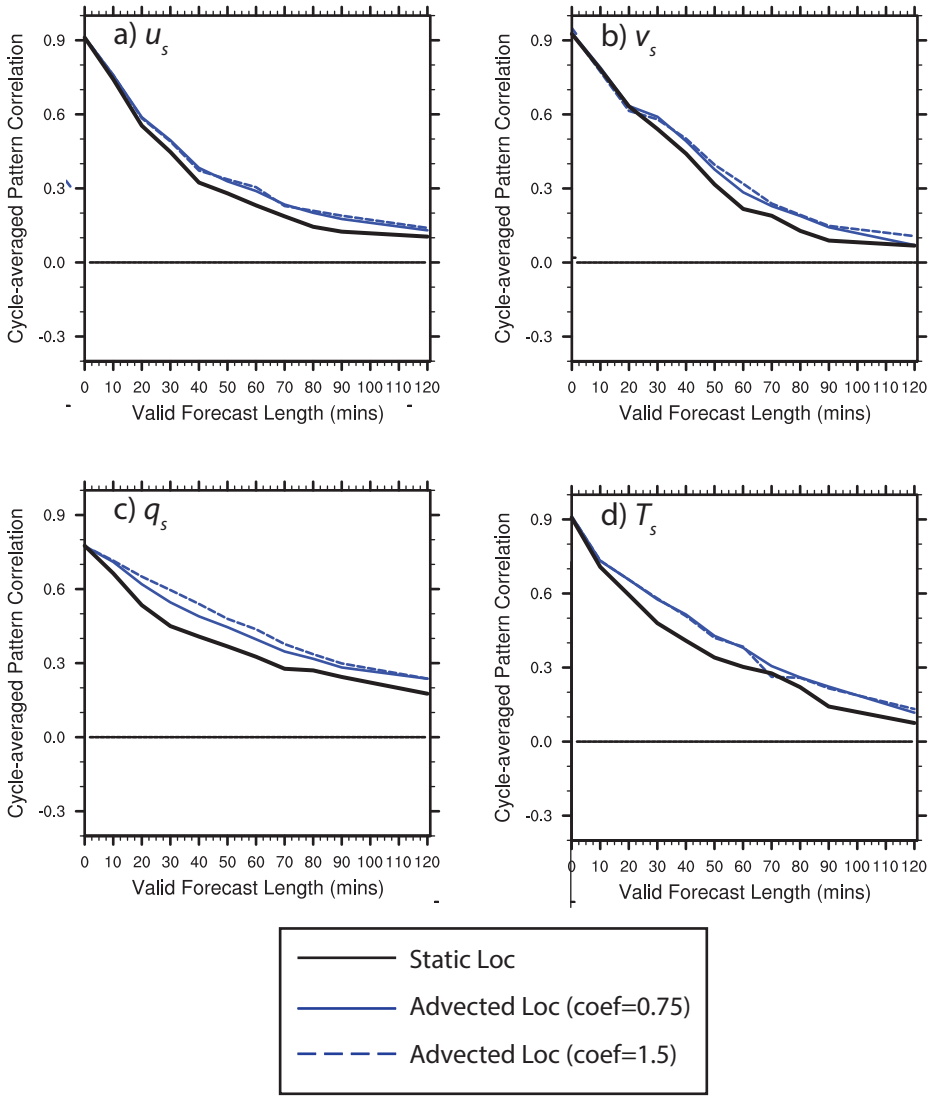

Figure 5.7. As in Figure 5.6 but for surface verification fields zonal wind (a), meridional wind (b), specific humidity (c), and temperature (d).

5.4.2 Evaluation of localization functions applied to EFSO

Correlations for EFSO with advected localizations show improvements for both energy norms and surface variable metrics (Figure 5.6Figure 5.7) The improvement is largest in the 10-30 minute forecast timeframe for energy norms, and the 20-60 minute timeframe for surface variables. The improvement is consistent as well, never falling below the performance of EFSO using static localization.

The advected localization method employs a tunable weighting factor to choose how far the localization is advected. A value greater than 1 indicates that processes other than physical advection must be contributing to the impact metric. Though most verifications show little difference in the setting of this coefficient, surface moisture shows a marked improvement when using a weight of 1.5 rather than 0.75 (Figure 5.7c). This is likely an indication that key nonlinearities are contributing to the impact metric. For surface moisture, complex boundary layer interactions ahead of the dryline could be processes contributing to impact being located further downstream than simple advection would indicate. This observation could indicate potential improvement in localizing the EFSO metric, perhaps with an adaptive method.

The pattern correlations from EFSO estimates applying RCF localization are shown in Figure 5.8. At earl lead times, the RCF underperforms static localization for all metrics. This is likely related to the finding from chapter 4, since the RCF is not the same as used during assimilation the estimate will not be as accurate. At 20-60 minute forecast times the RCF improves and is greater than static

localization for DTE, surface temperature, and surface zonal wind components, though it remains below the performance of advected localization. The exception is the period from 40-60 minutes for *u*-wind where the RCF actually outperforms the advected localization. On the other hand, for surface moisture and MTE variables the RCF remains below static localization for the duration of the forecast period.

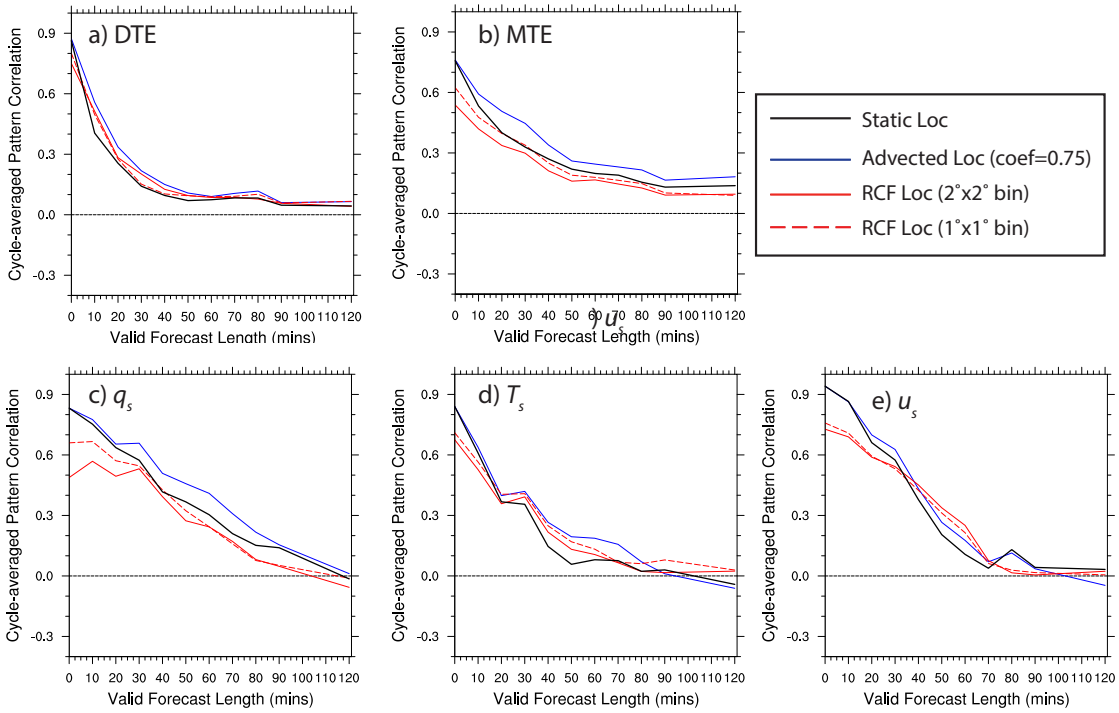

Figure 5.8. As in Figure 5.6 but with the addition of RCF localization tests using 2°x2° binning (solid red lines) and 1°x1° binning (dashed red lines). Verification metrics shown are (a) dry total energy, (b) moist total energy, (c) surface moisture, (d) surface temperature, and (e) surface zonal wind.

A subjective evaluation of EFSO for different localizations is shown in Figure 5.9, zoomed into the DFW domain. With increasing forecast lead-time, the field of actual impact also moves downstream. Each applied localization does a good job subjectively matching most of the areas. Some subtle differences in the advected localization exist, for instance the small-scale negative impact area is better captured than using static localization. The RCF function matches the main areas of positive impact well; however, the area closest to Dallas the RCF shows no negative impact as is seen in the actual field.

In terms of *u*-wind for longer lead times (60 and 120 minutes, Figure 5.10) the situation is considerably more complex. The actual impact field shows extensive small-scale areas of positive and negative impact. One such area in the western part of the domain appears to reflect convective rolls and cells found in the boundary layer. However, the EFSO is unable to capture such fine-scale details, and instead shows broad areas of positive impact. The RCF application does show a stronger hint of the rolls, but they are still too large compared to reality Figure 5.10f).

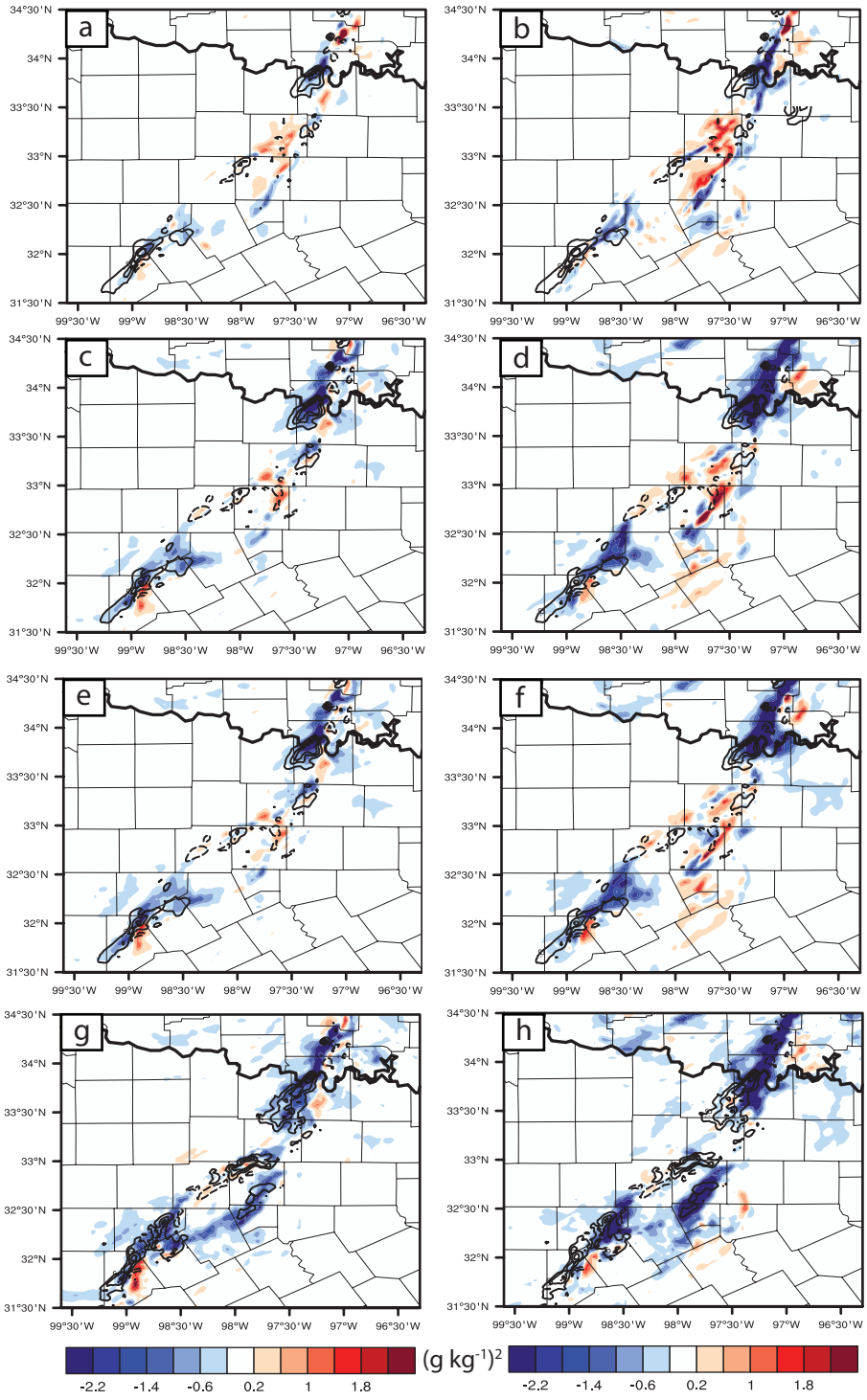

Figure 5.9. (a-b) Actual impact in surface moisture after 1800 UTC analysis. (c-d) EFSO estimates with static localization, (e-f) EFSO estimated with advected localization (coef=0.75), (g-h) EFSO estimates with 1˚x1˚ binned RCF localization. Left column is 30-min forecast, right column is 60-min forecast. Black contours show impact values at analysis time t=0.

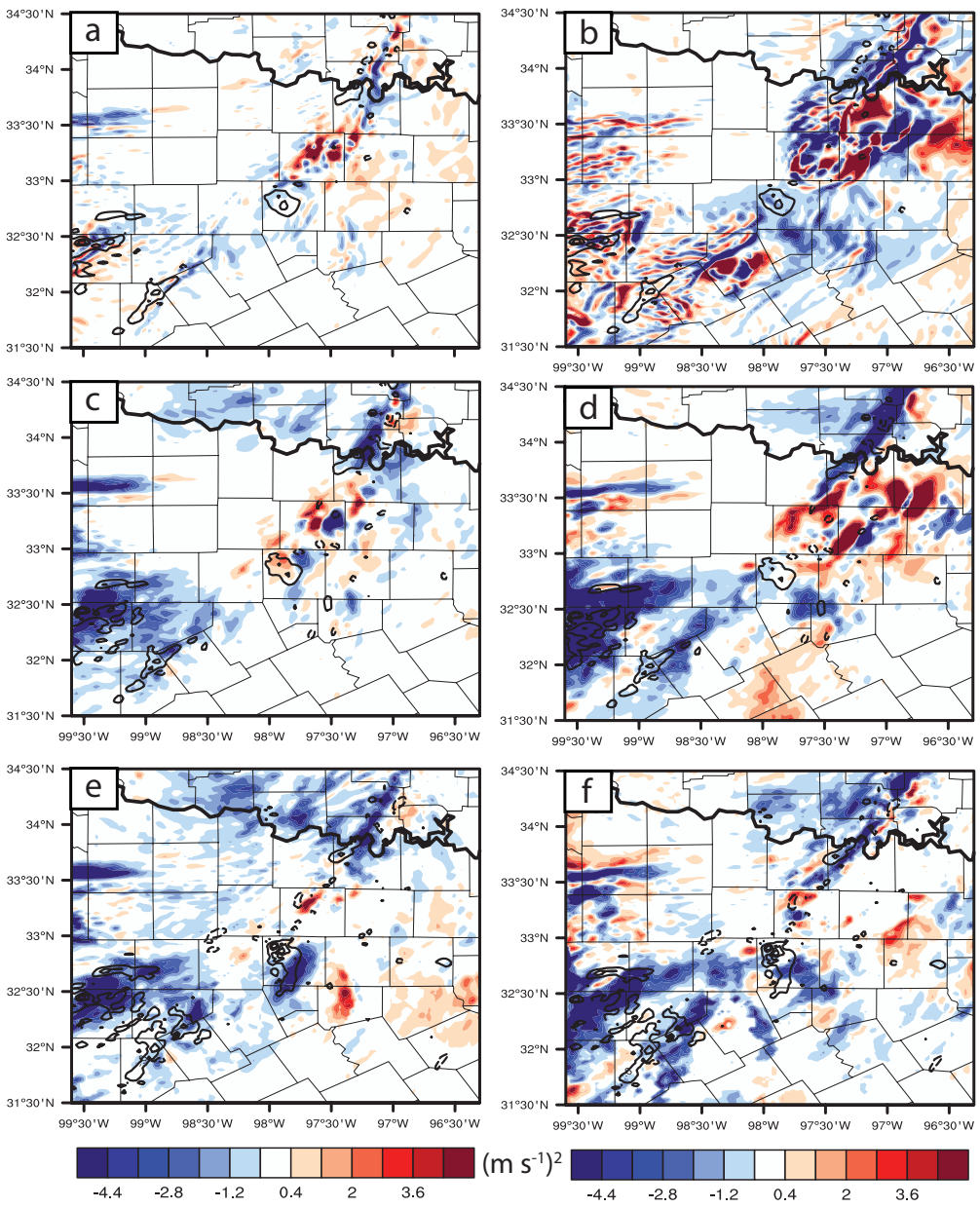

Figure 5.10. (a-b) Actual impact in surface zonal wind after 1800 UTC analysis. (c-d) EFSO estimate with advected localization, (e-f) EFSO estimate using binned RCF localization. Left column is for 60-min forecast, right column is 120-min forecast. Black contours show impact values at analysis time t=0.

5.4.3 Effect of nonlinearity on EFSO accuracy

Previous results have shown difficulty in the EFSO when significant nonlinear interactions exist in the forecast field, such as the fine scale wind features in Figure 5.10. Despite these nonlinearities, the EFSO matches favorably with the actual impact field for forecast times up to 80 minutes. The next question is can the EFSO be applied for nonlinear convective scale verification such as composite reflectivity? Figure 5.11 shows a case where it is feasible.

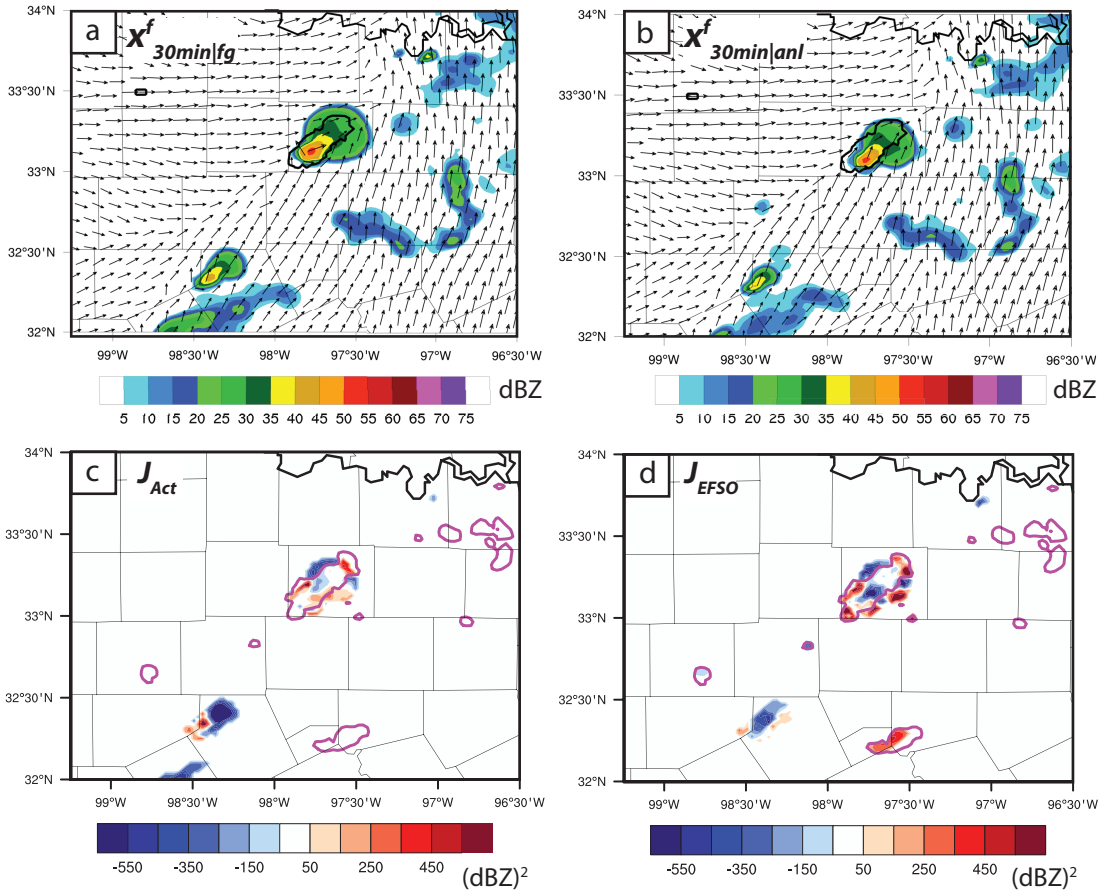

Figure 5.11. (a-b) 30-min forecast of composite reflectivity for the first guess and analysis mean, respectively. (c) Actual impact computed in terms of composite reflectivity as a verification, in units of dBZ². (d) The EFSO estimate of impact. Negative (blue) values indicate positive impact of observations.

Though the differences in the forecast storm between analysis and first guess forecasts are subjectively minor, the impact fields are consistent with the differences. The analysis forecast storm is displaced slightly south, causing a better match of the actual storm location and a slight improvement on the northern side. This area of improvement is captured in the EFSO estimate as well. Other areas of negative impact match as well. Additionally, the use of reflectivity as a verification metric can help identify areas where storms were effectively suppressed by the assimilation, such as the area in the southwest part of the plot domain. The EFSO is also capable of capturing the signal associated with removal of spurious convection.

Though the previous example shows the application of EFSO is possible for composite reflectivity, the correlation over many cycles is shown to be low (Figure 5.12b). There are several possible reasons for this low performance including sample size (less storms in the beginning of the forecast) and the enhanced nonlinearity caused by model spin up. Model spin up is also a concern for another nonlinear variable, surface pressure, where the correlation is also low and at parts even negative (Figure 5.12a).

To explore how large nonlinearity can negatively affect the EFSO metric, a single observation test with a surface observation was performed (Figure 5.13). Without an initialization technique, the effect of assimilating a single surface pressure observation is to shock the model, creating spurious gravity waves that emanate from the observation. The actual impact field thus covers a much wider area after just 5-10 minutes than the GC radius used during assimilation. Though

the EFSO captures the impact well at analysis time, even at just 5-minute forecast the EFSO is completely incapable of revealing the actual impact. Furthermore, moving localization cannot fix such a situation because the EFSO relies on linear ensemble covariances. The extent to which the EFSO is accurate relies on the accuracy of ensemble covariance between observations and a given forecast metric.

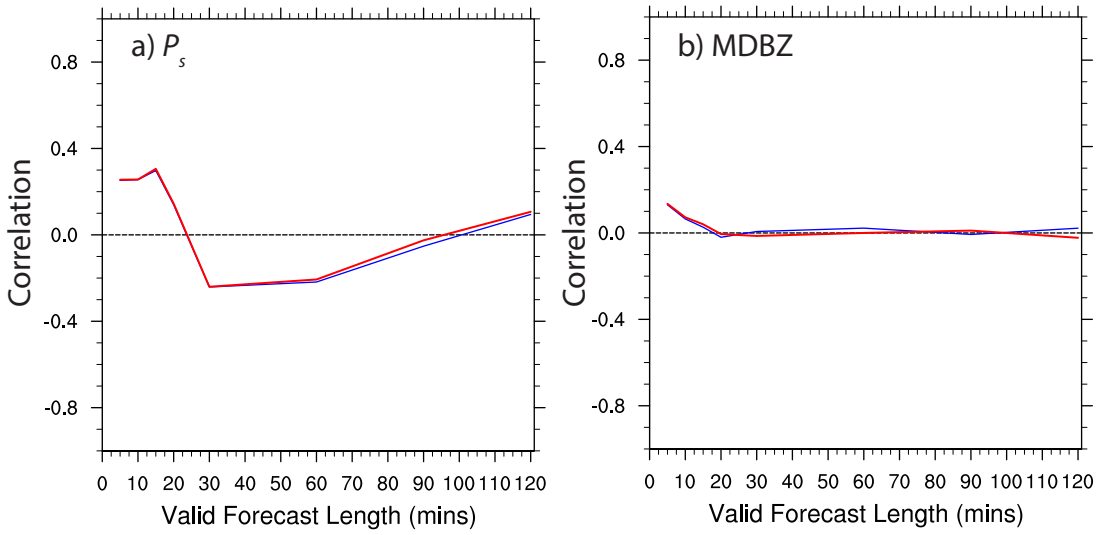

Figure 5.12. Correlation of EFSO estimate with actual impact in (a) surface pressure and (b) composite reflectivity verification metrics, shown for EFSO estimates with static (blue) and advected (red) localizations

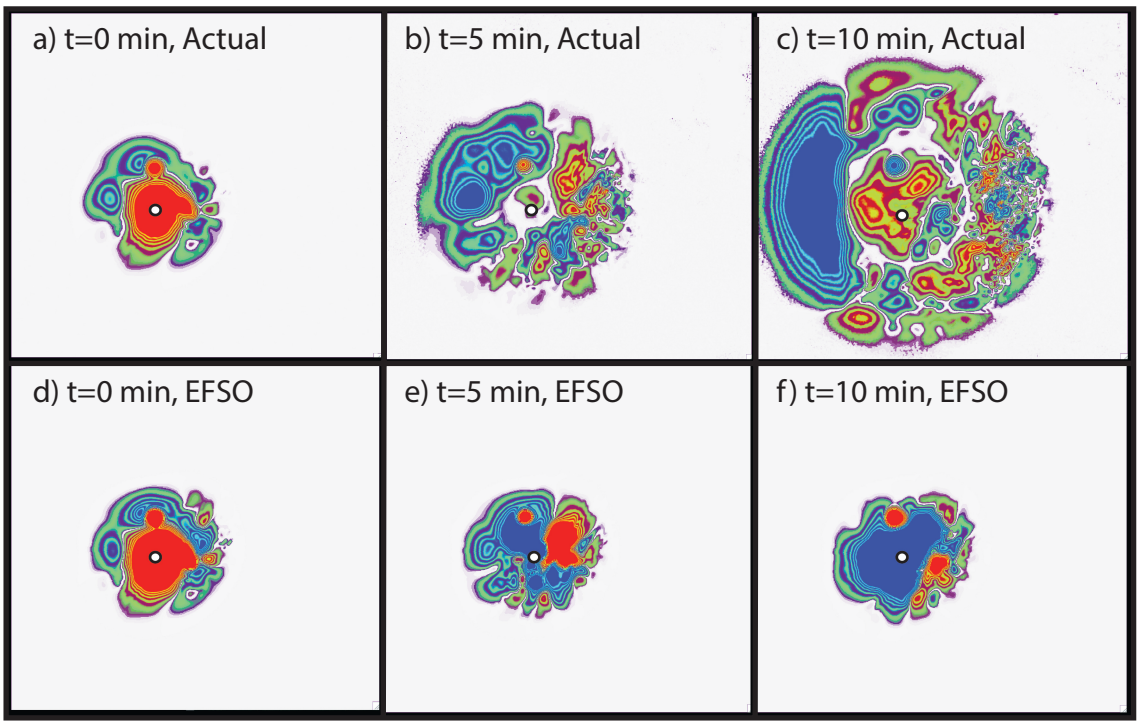

Figure 5.13. Single observation impact experiment with a surface pressure observation, verifying against surface pressure field. Top row (a-c) is the actual impact at t=0, 5, and 10 min forecasts, and bottom row (d-f) are EFSO estimates with static 200-km GC localization applied.

5.5 Summary and discussion

This chapter explored the application of the ensemble-based forecast sensitivity to observations, or EFSO, metric to a convective-scale DA and forecast system. Though one previous study, Sommer et al (2014), had applied EFSO to a convective-scale model, this is the first such study to explore the use of different verification metrics other than energy norms. These metrics included surface level variables u_s, v_s, T_s, q_s as well as composite reflectivity. Additionally, this is the first application of EFSO for high-frequency sub-hourly assimilation cycles.

The CNTL experiment from the CI case study of chapter 3 was used for the EFSO experiments. Each 5-min DA cycle in the 2-hour inner grid cycling period

allowed for 25 quasi-independent samples to compare EFSO with the actual impact. Results showed pattern correlation that dropped off quickly for energy norms. With advected localization applied, energy norms remained above a weak 0.3 correlation threshold only up until 20-30 minutes forecast. However when verifying against surface variables, it was found that the correlations dropped off much more gradually, remaining useful up until as much as 80 minutes in forecast length.

The RCF localization used in Gasperoni and Wang (2015) was adapted for the EFSO study here to see if improvements could be made to the estimate. A binning technique was used to allow for averaging to reduce sampling error. The resulting RCF functions revealed underlying flow-dependent correlations specific to the case study and represented the time-forecast component. However, the application of RCF to EFSO did not yield improved estimates. The main benefit of RCF appears to be the time-forecast component, which the advected localization already covers. Additionally RCF was not used during assimilation, and the implicit dependence on the localization function used during DA may have hindered its usefulness. Moreover, the binning technique may have smoothed out important subtle details in the forecast-time component that go beyond simple advection. Results could be improved with further tuning of the bin technique, or finding another method to average out noise while maintaining the signal. Other adaptive localization methods may be better suited to application at convective scale, such as the SENCORP method of Bishop and Hodyss (2007).

The biggest drawback to the application of EFSO to convective scales is the significant increase in nonlinear processes within model integration, which limited ensemble-size covariances will not be able to adequately represent. The EFSO can work to identify impactful areas in developing storms, but the highly nonlinear nature of storm evolution limits the accuracy of EFSO in such cases. It may be the case, however, that more predictable scenarios such as linear storm systems ahead of cold fronts and mesoscale convective systems are better suited for application of EFSO. Additionally, other convectively suitable verifications such as neighborhood methods may yield better results to avoid any double-penalty issues – though it is unclear how well the ensemble could handle covariances in that case. More study is needed in the application of the EFSO method to more diverse cases at convective scales.

Chapter 6: Conclusions

The research presented in this dissertation was motivated by the Nationwide Network of Networks vision of the National Research Council (2009). Part of the effort is to help identify mesoscale observation networks most useful for high-impact severe weather forecasts. Two main themes of research in this dissertation cover different methods to achieve such a goal. The first is straightforward data denial experiments, where all observations in the Dallas Fort Worth Testbed are assimilated in a model and experiments withholding subsets of observations help to determine the degree of impact of each different system. The second theme of research covers a new ensemble-based method – the ensemble based forecast sensitivity to observations (EFSO) method of Kalnay et al. (2012) - formulated to automatically generate observation system impacts without the need for laborious separate denial experiments. If successful, the EFSO can be applied to the DFW testbed for a month-long or seasonal monitoring of impacts in the DA system. This chapter provides a summary and synthesis of the key results of previous chapters.

On the usefulness of nonconventional surface data for CI prediction. Results from the CI case study of 3 April 2014 showed that each observational data source – conventional and nonconventional – played an important role in the successful CI prediction within the ensemble. While conventional surface data from ASOS and AWOS defined the larger mesoscale environment (i.e. dryline position), the nonconventional surface data from ERNET and WXBUG together helped constrain

the model to define a small-scale kink in the dryline where the CI occurred in reality. Though removal of one source did not completely alter the ensemble preference to initiate convection in SW Wise country, the optimal performance occurred when all sources were included within the assimilation.

The importance of metadata. The data denial work revealed a surprising dependence of resulting CI on observations relatively far away, but in a sensitive region along the dryline. These observations were also from hydrometeorology sites, which may or may not contain a moisture representativeness bias that should be taken into account into the assimilation. Additionally, many observations have unknown heights and siting quality. In the DA, the height was assumed to be 2 m AGL, but this may be highly variable and unrepresentative of the actual observation height. This had an effect on the influence of wind observations, since wind is much more dependent on the assimilation height due to the nature of friction at the surface. Poorly sited observations can also negatively affect the wind estimates. As a result, the wind information assimilated from nonconventional sources could be detrimental due to the noisiness introduced by unknown poor observations. Additionally, accurate metadata is essential to improve upon the usefulness of these nonconventional sources for the NNoN vision.

High-frequency data assimilation. The use of a long period of 5-min DA cycling helped to constrain model biases and tendencies for the CI case. The conventional

and nonconventional data together prevented the dryline from overmixing and advancing too far east. They further helped in defining the sharp gradient that the model tended to smooth out too much, further hindering the CI prediction. This result was similar to the findings of Sobash and Stensrud (2015), who found the best performance using three hours of 5-min cycling in their case study.

Application of adaptive localization to EFSO. The two-layer model work showed that an adaptive technique, in this case the RCF method based on a group filter technique, was able to capture the important time-forecast component of the localization for the EFSO metric. The improvements shown were the strongest for longer forecast time periods.

Dependence of EFSO localization on DA localization. The two-layer model work also revealed an important dependence that had not been explicitly discussed before in the literature. It was found that optimal EFSO localization depends on the localization applied during the DA period, in addition to the model time-forecast component. This was further proven by assimilating observations using analysis RCF functions such that there was consistency in localization method during DA and during EFSO estimation forecast times. This new experiment had skill scores that outperformed all other experiments, further confirming this implicit dependence. Revisiting the derivation of Kalnay et al. (2012) including the presence of localization within the Kalman gain matrix further confirmed this assumption and dependence.

Feasibility of applying EFSO at convective scales. The EFSO technique had been successfully applied to many synoptic and global scale systems. However, it had not been applied to a fully convective scale system, using not only convective scale model grid resolution but also high-frequency DA cycling with many different observation sources. The application here also critically included verification metrics more suitable for convective scale application. This was an important point stressed because energy norms may not be appropriate to describe the performance of a convective NWP system. Oftentimes the concern is with different subsets of the model grid, such as surface variables influenced heavily by the boundary layer, or metrics describing the evolution of storms such as reflectivity.

The use of reflectivity as a verification metric needs further study. It was demonstrated to work for a qualitative case study, but when compared statistically over many cycles the correlations with the actual impact were very small. Improvements may be made by using other convective-scale appropriate metrics such as neighborhood probabilities, though it is not clear how EFSO would work with a more complex, bounded variable.

Another consideration to the application of EFSO is to study more diverse cases. The CI case may limit the applicability of convective scale metrics, given its highly nonlinear evolution. More linear convective type systems such as mesoscale convective systems may be better suited for EFSO application.

Localization of EFSO at convective scales. The advected localization method from Ota et al. (2013) may likely be the most appropriate method to use for EFSO

at convective scales, given it is simple to implement and performed better than even the adaptive RCF method at convective scales. This is likely due again to the implicit dependence of localizing EFSO on the localization function applied during DA. Since the GC function is most commonly used as localization in ensemble DA systems, the easiest way to maintain this dependence is through the use of advected GC localization. Other adaptive methods may yet yield better results with EFSO, but more work is needed in order to determine where improvements can be made in localizing for EFSO.

Bibliography

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analysis. *Mon. Wea. Rev.*, **137,** 1805-1824.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129,** 2884-2903.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Phys. D,* **230,** 99-111.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140,** 2359-2371.

Anderson, J. L., and L. Lei, 2013: Empirical localization of observation impact in ensemble Kalman filters. *Mon. Wea. Rev.*, **141,** 4140-4153.

Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux, and Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite sounding data. *Mon. Wea. Rev.*, **119,** 1851-1865.

Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan*, **75,** 111-130.

Banacos, P. C., and D. M. Schultz, 2005: The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives. *Wea. Forecasting*, **20,** 351-366.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525,** 47-55.

Benjamin, S. G., B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPW-PW, METAR, and mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev.*, **138,** 1319-1343.

Bishop, C. H., and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133,** 2029-2044.

Bishop, C. H., and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus A,* **61,** 84-96.

——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus A*, **61,** 97-111.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420.

Bosse, J., D. Oberholzer, M. Nelson, and R. Ware, 2012: Thermodynamic remote sensing of the boundary layer and above, a network approach. *16th Int. Symp. Adv. Boundary-Layer Remote Sens*.

Buehner, M., 2012: Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Wea. Rev.*, **140,** 617-636.

Buehner, M., and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133,** 615-630.

Burgers, G., P. J. v. Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126,** 1719-1724.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138,** 282-290.

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135,** 239-250.

Carlaw, L. B., J. A. Brotzge, and F. H. Carr, 2015: Investigating the impacts of assimilating surface observations on high-resolution forecasts of the 15 May 2013 tornado event. *Electronic J. Severe Storms Meteor.*, **10,** 1-34.

Chadwick, R., 2014: Citizen weather observer program. Accessed 8 June 2017. [Available online at http://www.wxqa.com.]

Chen, X., K. Zhao, J. Sun, B. Zhou, and W.-C. Lee, 2016: Assimilating surface observations in a four-dimensional variational doppler radar data assimilation system to improve the analysis and forecast of a squall line case. *Adv. Atmos. Sci.*, **33,** 1106-1119.

Chen, Y., and D. Oliver, 2010: Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci.*, **14,** 579-601.

Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies, 15, NASA/TM-104606, 40 pp.

Cohn, S. E., A. d. Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126,** 2913-2926.

Coniglio, M. C., S. M. Hitchcock, and K. H. Knopfmeier, 2016: Impact of assimilating preconvective upsonde observations on short-term forecasts of convection observed during MPEX. *Mon. Wea. Rev.*, **144,** 4301-4325.

Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28,** 842-862.

Dabberdt, W. F., and Coauthors, 2005: Multifunctional mesoscale observing networks. *Bull. Amer. Meteor. Soc.*, **86,** 961-982.

Dahlia, J., 2013: The national mesonet program: Filling in the gaps. *Weatherwise*, **66 (4),** 26-33.

Developmental Testbed Center, 2016: *Gridpoint statistical interpolation advanced user's guide version 3.5.0.0.* Available at http://www.dtcenter.org/com-GSI/users/docs/index.php, 143 pp.

Earth Networks, 2017: Weather: Get smart about the weather. Accessed 1 July 2017 [Available online at https://www.earthnetworks.com/why-us/networks/weather/.]

Ehrendorfer, M., R. M. Errico, and K. D. Raeder, 1999: Singular-vector perturbation growth in a primitive equation model with moist physics. *J. Atmos. Sci.*, **56,** 1627-1648.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95,** 1335-1342.

Emerick, A., and A. Reynolds, 2011: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput. Geosci.*, **15,** 251-269.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, **99,** 10143-10162.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125,** 723-757.

Gasperoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Mon. Wea. Rev.*, **143,** 1981-2000.

Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX observation impact intercomparison experiment. *Mon. Wea. Rev.*, **138,** 4009-4025.

Grell, G. A., and D. De´ve´nyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29,** 38-31 - 38-34.

Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14,** 5233-5250.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139,** 511-522.

Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of weather and climate*, T. Palmer, and R. Hagedorn, Eds., Cambridge Press, 124-156.

Hamill, T. M., and C. Snyder, 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130,** 1552-1572.

Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.*, **133,** 3132-3147.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129,** 2776-2790.

Hamill, T. M., F. Yang, C. Cardinali, and S. J. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Mon. Wea. Rev.*, **141,** 2058-2065.

Hane, C. E., H. B. Bluestein, T. M. Crawford, M. E. Baldwin, and R. M. Rabin, 1997: Severe thunderstorm development in relation to along-dryline variability: A case study. *Mon. Wea. Rev.*, **125,** 231-251.

Hane, C. E., R. M. Rabin, T. M. Crawford, H. B. Bluestein, and M. E. Baldwin, 2002: A case study of severe storm development along a dryline within a

synoptically active environment. Part II: Multiple boundaries and convective initiation. *Mon. Wea. Rev.*, **130,** 900-920.

Hendon, H. H., and D. L. Hartmann, 1985: Variability in a nonlinear model of the atmosphere with zonally symmetric forcing. *J. Atmos. Sci.*, **42,** 2783-2797.

Heppner, P., 2013: Building a national network of mobile platforms for weather detection. *29th Conf. Environ. Inf. Process. Technol.*, A. M. Soc., Ed., 5.8.

Hilliker, J. L., G. Akasapu, and G. S. Young, 2010: Assessing the short-term forecast capability of nonstandardized surface observations using the National Digital Forecast Database (NDFD). *J. App. Meteor. and Climatol.*, **49,** 1397-1411.

Holland, B., and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **139,** 758-770.

Hotta, D., E. Kalnay, Y. Ota, and T. Miyoshi, 2017a: EFSR: Ensemble forecast sensitivity to observation error covariance. *Mon. Wea. Rev.*, **145,** 5015-5031.

Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017b: Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Mon. Wea. Rev.*, **145,** 3331-3354.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126,** 796-811.

——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129,** 123-137.

——, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131,** 3269-3289.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144,** 4489-4532.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133,** 604-620.

Hu, M., H. Shao, D. Stark, K. Newman, C. Zhou, G. Ge, and X. Zhang, 2017: *Grid-point Statistical Interpolation (GSI) user's guide version 3.6.* Developmental Testbed Center, Available at http://www.dtcenter.org/com-GSI/users/docs/index.php, 149 pp.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Phys. D.*, **230,** 112-126.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.: Atmos.*, **113,** 2156-2202.

Janjić, T., L. Nerger, A. Albertella, J. Schröter, and S. Skachko, 2011: On domain localization in ensemble-based Kalman filter algorithms. *Mon. Wea. Rev.*, **139,** 2046-2060.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122,** 927-945.

Jazwinski, A. H., 1970: *Stochastic processes and filtering theory.* 376 pp.

Johnson, A., and X. Wang, 2017: Design and implementation of a GSI-based convection-allowing ensemble data assimilation and forecast system for the PECAN field experiment. Part I: Optimal configurations for nocturnal convection prediction using retrospective cases. *Wea. Forecasting*, **32,** 289-315.

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVAR data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143,** 3087-3108.

Kain, J. S., 2004: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43,** 170-181.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82D,** 35-45.

Kalman, R. E., and R. S. Bucy, 1961: New results in linear filtering and prediction problems. *J. Basic Eng.*, **83,** 95-108.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability.* Cambridge University Press, 341 pp.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus A*, **64**.

Kepert, J. D., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135,** 1157-1176.

——, 2011: Balance-aware covariance localisation for atmospheric and oceanic ensemble Kalman filters. *Comput. Geosci.*, **15,** 239-250.

Keyser, D., 2017: Prepbufr processing at NCEP. Accessed 23 June 2017. [Available online at http://www.emc.ncep.noaa.gov/mmb/data_processing/prepbufr.doc/document.htm.]

Kunii, M., T. Miyoshi, and E. Kalnay, 2012: Estimating the impact of real observations in regional numerical weather prediction using an ensemble Kalman filter. *Mon. Wea. Rev.*, **140,** 1975-1987.

Kutty, G., and X. Wang, 2015: A comparison of the impacts of radiosonde and AMSU radiance observations in GSI based 3dEnsVar and 3DVAR data assimilation systems for NCEP GFS. *Adv. Meteor.*, **2015,** 17.

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information. *Wea. Forecasting*, **22,** 596-612.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the nrl atmospheric variational data assimilation adjoint system. *Tellus*, **56A,** 189-201.

Langland, R. H., C. Velden, P. M. Pauley, and H. Berger, 2009: Impact of satellite-derived rapid-scan wind observations on numerical model forecasts of hurricane katrina. *Mon. Wea. Rev.*, **137,** 1615-1622.

Lei, L., and J. L. Anderson, 2014: Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model. *Mon. Wea. Rev.*, **142,** 739-754.

Li, H., J. Liu, and E. Kalnay, 2010: Correction of 'estimating observation impact without adjoint model in an ensemble Kalman filter'. *Quart. J. Roy. Meteor. Soc.*, **136,** 1652-1654.

Liu, H., and M. Xue, 2008: Prediction of convective initiation and storm evolution on 12 June 2002 during IHOP. Part I: Control simulation and sensitivity experiments. *Mon. Wea. Rev.*, **136,** 2261-2283.

Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **134,** 1327-1335.

Lord, S. J., E. Kalnay, R. Daley, G. D. Emmitt, and R. Atlas, 1997: Using OSSEs in the design of the future generation of integrated observing systems. *Prepr. Vol.,*

1st Symp. Integr. Obs. Syst., Long Beach, CA, Amer. Meteor. Soc., 45-47. [Available online at http://www.emc.ncep.noaa.gov/research/osse/].

McLaughlin, D., and Coauthors, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90,** 1797-1817.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic. Technol.*, **24,** 301-321.

Miller, P. A., M. F. Barth, and L. A. Benjamin, 2005: An update on MADIS observation ingest, integration, quality control, and distribution capabilities. *21st Int. Conf. Interact. Inf. Process. Syst. (IIPS) Meteor., Oceonogr., and Hydrol.*, San Diego, CA, Amer. Meteor. Soc., J7.12. [Available online at https://ams.confex.com/ams/Annual2005/techprogram/paper_86703.htm].

Moninger, W. R., R. D. Mamrosh, and P. M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Amer. Meteor. Soc.*, **84,** 203-216.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulent closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87,** 895-912.

National Center for Environmental Information, 2014: *Storm data.* Vol. 56, No. 4, NOAA, 652 pp.

National Research Council, 2009: *Observing weather and climate from the ground up: A nationwide network of networks.* The National Academies Press, 250 pp.

——, 2012: *Urban meteorology: Forecasting, monitoring, and meeting users' needs.* The National Academies Press, 190 pp.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus*, **65A**.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, **56,** 415-428.

Pu, Z., H. Zhang, and J. Anderson, 2013: Ensemble Kalman filter assimilation of near-surface observations over complex terrain: Comparison with 3DVAR for short-range forecasts. *Tellus A*, **65**.

Roberts, R. D., and S. Rutledge, 2003: Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. *Wea. Forecasting*, **18,** 562-584.

Schenkman, A., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011a: The analysis and prediction of the 8-9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139,** 224-246.

——, 2011b: Impact of CASA radar and Oklahoma Mesonet data assimilation on the analysis and prediction of tornadic mesovortices in a MCS. *Mon. Wea. Rev.*, **139,** 3422-3445.

Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103,** 246-257.

Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A technical overview. *J. Atmos. Oceanic. Technol.*, **22,** 211-222.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25,** 263-280.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128,** 647-678.

Singh, R., S. P. Ojha, C. M. Kishtawal, and P. K. Pal, 2014: Impact of various observing systems on weather analysis and forecast over the indian refion. *J. Geophys. Res.: Atmos.*, **119,** 10,232.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227,** 3465-3485.

Snook, N., M. Xue, and Y. Jung, 2015: Multiscale EnKF assimilation of radar and conventional observations and ensemble forecasting for a tornadic mesoscale convective system. *Mon. Wea. Rev.*, **143,** 1035-1057.

Sobash, R. A., and D. J. Stensrud, 2013: The impact of covariance localization for radar data on EnKF analyses of a developing MCS: Observing system simulation experiments. *Mon. Wea. Rev.*, **141,** 3691-3709.

——, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 may 2012. *Mon. Wea. Rev.*, **143,** 3700-3725.

Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **140**, 2672-2679.

——, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus A: Dyn. Met. and Ocean.*, **68**.

Tewari, M., and Coauthors, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. *20th Conf. Wea. Anal. Forecasting/16th Conf. NWP*, Seattle, WA, Amer. Met. Soc., 11-15

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136,** 5095-5115.

Thompson, R., and J. Hart, 2017: SPC experimental sounding analysis system. Accessed 16 June 2017. [Available online at http://www.spc.noaa.gov/exper/soundings/help/index.html.]

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131,** 1485-1490.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133,** 1789-1807.

Tyndall, D. P., and J. D. Horel, 2013: Impacts of mesonet observations on meteorological surface analyses. *Wea. Forecasting*, **28,** 254-269.

Understory Weather, 2015: Understory sensor data: Weather monitoring networks accurately measuring hail and wind. [Available online at http://try.understoryweather.com/white-paper-accuracy-register/].

van Zomeren, J., and A. van Delden, 2007: Vertically integrated moisture flux convergence as a predictor of thunderstorms. *Atmos. Res.*, **83,** 435-445.

Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter – optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.*, **135,** 1055–1076.

——, 2009: A comparison of the hybrid and EnSRF analysis schemes in the presence of model errors due to unresolved scales. *Mon. Wea. Rev.*, **137,** 3219-3232.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVAR-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141,** 4098-4117.

WeatherFlow, 2017: Sodar and lidar. Accessed 8 June 2017. [Available online at http://weatherflow.com/professional-services/wind-energy/sodar-and-lidar/.]

Weckwerth, T. M., 2000: The effect of small-scale moisture variability on thunderstorm initiation. *Mon. Wea. Rev.*, **128,** 4017.

Weckwerth, T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.*, **134,** 5–22.

Weckwerth, T. M., and Coauthors, 2004: An overview of the International H2O Project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.*, **85,** 253-277.

Weissmann, M., R. H. Langland, C. Cardinali, P. M. Pauley, and S. Rahm, 2012: Influence of airborne doppler wind lidar profiles near typhoon Sinlaku on ECMWF and NOGAPS forecasts. *Quart. J. Roy. Meteor. Soc.*, **138,** 118-130.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130,** 1913-1924.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140,** 3078-3089.

Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136,** 463-482.

Xue, M., and W. J. Martin, 2006a: A high-resolution modeling study of the 24 May 2002 case during IHOP. Part I: Numerical simulation and general evolution of the dryline and convection. *Mon. Wea. Rev.*, **134,** 149–171.

——, 2006b: A high-resolution modeling study of the 24 May 2002 case during IHOP. Part II: Horizontal convective rolls and convective initiation. *Mon. Wea. Rev.*, **134,** 172–191.

Zapotocny, T. H., W. P. Menzel, J. P. Nelson, and J. A. Jung, 2002: An impact study of five remotely sensed and five in situ data types in the Eta data assimilation system. *Wea. Forecasting*, **17,** 263-285.

Zapotocny, T. H., J. A. Jung, J. F. Le Marshall, and R. E. Treadon, 2007: A two-season impact study of satellite and in situ data in the NCEP global data assimilation system. *Wea. Forecasting*, **22,** 887-909.

Zhang, X., Y. Luo, Q. Wan, W. Ding, and J. Sun, 2016: Impact of assimilating wind profiling radar observations on convection-permitting quantitative precipitation forecasts during scmrex. *Wea. Forecasting*, **31,** 1271-1292.

Zhang, Y., and D. S. Oliver, 2010: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Math. Geosci.*, **42,** 327-345.

Zou, X., A. Barcilon, I. M. Navon, J. Whitaker, and D. G. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121,** 2833-2857.

Appendix A: Derivation of EFSO

Observation impact is defined as the reduction of squared error at some forecast time *t* from observations assimilated at time 0.

$$J_{Actual} = \frac{1}{2}\left(\mathbf{e}_{t|0}^T C \mathbf{e}_{t|0} - \mathbf{e}_{t|-n}^T C \mathbf{e}_{t|-n}\right) \tag{A1}$$

Where $\mathbf{e}_{t|0} = \overline{\mathbf{x}}_{t|0}^f - \mathbf{x}_t^{tr}$ and $\mathbf{e}_{t|-n} = \overline{\mathbf{x}}_{t|-n}^f - \mathbf{x}_t^{tr}$ are the errors for analysis and first guess mean forecasts, respectively, relative to some verifying truth field. Matrix *C* is often within the formulation to define the error norm used. Equation (A1) can be rearranged into the following form.

$$\begin{aligned} J_{Actual} &= \frac{1}{2}\left(\mathbf{e}_{t|0} - \mathbf{e}_{t|-n}\right)^T C \left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \\ &= \frac{1}{2}\left(\overline{\mathbf{x}}_{t|0}^f - \overline{\mathbf{x}}_{t|-n}^f\right)^T C \left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \end{aligned} \tag{A2}$$

If we assume the forecast length is short enough such that the linear tangent model can be used, the first term can be approximated as $\overline{\mathbf{x}}_{t|0}^f - \overline{\mathbf{x}}_{t|-n}^f \approx \mathbf{M}\left(\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_{0|-n}^b\right)$. Thus, (A2) becomes

$$J \approx \frac{1}{2}\left[\mathbf{M}\left(\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_{0|-n}^b\right)\right]^T C \left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \tag{A3}$$

The first term of (A3) is tied directly to the mean update step of any deterministic EnKF; that is, we know that $\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_{0|-n}^b = \mathbf{K}\delta\mathbf{y}_0$. So, (A3) can be rewritten as

$$J \approx \frac{1}{2}\left[\mathbf{M}\mathbf{K}\delta\mathbf{y}_0\right]^T C \left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \tag{A4}$$

Within the approach of Langland and Baker (2004), the transpose of the first term is then taken, i.e. $\left[\mathbf{M}\mathbf{K}\delta\mathbf{y}_0\right]^T = \delta\mathbf{y}_0^T\mathbf{K}^T\mathbf{M}^T$. This requires the formation of the adjoint to the tangent linear model ($\mathbf{M}^T$) and data assimilation ($\mathbf{K}^T$).

In the ensemble approach of Kalnay et al. (2012), the ensemble approach to (A4) starts from Kalman gain matrix **K** and substitutes the expression $\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} = (K-1)^{-1}\mathbf{X}_0^a\mathbf{X}_0^{aT}\mathbf{H}^T\mathbf{R}^{-1}$ into (A4). Note here we are using the post-DA analysis ensemble perturbations to calculate the analysis error covariances.

$$J \approx \frac{1}{2(K-1)}\left[\mathbf{M}\mathbf{X}_0^a\mathbf{X}_0^{aT}\mathbf{H}^T\mathbf{R}^{-1}\delta\mathbf{y}_0\right]^T C\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \tag{A5}$$

Here we define $\mathbf{Y}_0^a \equiv \mathbf{H}\mathbf{X}_0^a$ as the analysis ensemble perturbations in observation space. Additionally, since we can approximate $\mathbf{M}\mathbf{X}_0^a \approx \mathbf{X}_{t|0}^f$ then (A5) becomes

$$J \approx \frac{1}{2(K-1)}\left[\mathbf{X}_{0|n}^f\mathbf{Y}_0^{aT}\mathbf{R}^{-1}\delta\mathbf{y}_0\right]^T C\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \tag{A6}$$

Finally, taking the transpose of the first term we obtain the full ensemble estimate as in Kalnay et al. (2012)

$$J_{EFSO} \approx \frac{1}{2(K-1)}\delta\mathbf{y}_0^T\mathbf{R}^{-1}\mathbf{Y}_0^a\mathbf{X}_{0|n}^{fT}C\left(\mathbf{e}_{t|0} + \mathbf{e}_{t|-n}\right) \tag{A7}$$

The benefit to the formulation of (A7) is that each term can be readily available from any deterministic EnKF. Additionally, there is no tangent linear model in the formulation; the ensemble forecast perturbations $\mathbf{X}_{0|n}^f$ could be computed from the full nonlinear model. That is, each column in $\mathbf{X}_{0|n}^f$, corresponding to each ensemble member *k*, can be calculated using $M(x_{k,t=0}^a) - \overline{M(x_{t=0}^a)}$.

Appendix B: Relationship between EnKF localization and EFSO localization

In Chapter 4, section 4.4.4, it was inferred from a single observation experiment at the tropics that when localization is applied during assimilation, ρ_A, then the EFSO localization, ρ_I, is also implicitly dependent on ρ_A. This can also be shown within the derivation of the EFSO equations. The Kalman gain **K** matrix in an EnKF is calculated as follows.

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{B1}$$

However, when localization is applied during DA, the Kalman gain matrix is instead written in the following form.

$$\mathbf{K}_{loc} = \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T\left[\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{B2}$$

For demonstrative purposes, we will assume **B**-type localization (see section 2.2.2), though the following can be shown with other types of localization in a similar manner. Depending on the type of EnKF used, the ensemble is then updated to be consistent with analysis covariance update step of the Kalman filter. With a localized Kalman gain, $\mathbf{K}_{loc}$, this step can be written as

$$\mathbf{P}^a = \left(\mathbf{I} - \mathbf{K}_{loc}\mathbf{H}\right)\mathbf{P}^b \tag{B3}$$

Note that while localization is applied to each instance of $\mathbf{P}^b$ within $\mathbf{K}_{loc}$, it is *not* applied to the final $\mathbf{P}^b$ in (B3). This is logical since we expect no update to be performed on the model state where localization goes to 0, resulting in unchanged covariances in those locations.

When ρ_A is applied to the Kalman gain, then the "localized analysis covariance" matrix, $\mathbf{P}^a_{loc}$, can be defined as

$$\mathbf{P}^a_{loc} \equiv \left(\mathbf{I} - \mathbf{K}_{loc}\mathbf{H}\right)(\rho_A \circ \mathbf{P}^b) \tag{B4}$$

Though $\mathbf{P}^a_{loc}$ has no physical meaning, it is nevertheless an important mathematical relationship relevant to the discussion here. Note that $\mathbf{P}^a_{loc} \neq \mathbf{P}^a$.

In the next step, equation (B2) will be manipulated with some basic matrix algebra into a different form. Each step is written below, starting with the (B2) expression.

$$\begin{aligned}
\mathbf{K}_{loc} &= \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T\left[\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T + \mathbf{R}\right]^{-1} \\
\mathbf{K}_{loc}\left[\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T + \mathbf{R}\right] &= \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T \\
\mathbf{K}_{loc}\mathbf{R} + \mathbf{K}_{loc}\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T &= \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T \\
\mathbf{K}_{loc}\mathbf{R} &= \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T - \mathbf{K}_{loc}\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T \\
\mathbf{K}_{loc} &= \left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T\mathbf{R}^{-1} - \mathbf{K}_{loc}\mathbf{H}\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T\mathbf{R}^{-1} \\
\mathbf{K}_{loc} &= \left[\mathbf{I} - \mathbf{K}_{loc}\mathbf{H}\right]\left(\rho_A \circ \mathbf{P}^b\right)\mathbf{H}^T\mathbf{R}^{-1}
\end{aligned} \tag{B5}$$

Note that the final expression (B5) contains the previously-defined localized analysis covariance matrix, $\mathbf{P}^a_{loc}$ from (B4), such that (B5) becomes

$$\mathbf{K}_{loc} = \mathbf{P}^a_{loc}\mathbf{H}^T\mathbf{R}^{-1} \tag{B6}$$

Equation (B6) is critical in understanding how ρ_A and ρ_I are related. The derivation in Kalnay et al. (2012) depended on the expression $\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1}$. Though true without localization, the previous steps show that it is not explicitly true when ρ_A is applied. Rather, the correct expression to use in the derivation is

(B6). Since $\mathbf{P}_{loc}^{a}$ is never calculated, we still seek to use $\mathbf{P}^{a}$ since it is readily available from the analysis ensemble. Thus there is an implicit assumption to maintain consistency in the Kalnay et al. (2012) derivation, which can be written as the following:

$$\mathbf{P}_{loc}^{a} \approx \rho_A \circ \mathbf{P}^{a} \tag{B7}$$

Returning to the EFSO derivation, then $\delta\mathbf{y}^T \mathbf{K}^T\mathbf{M}^T = \delta\mathbf{y}^T \mathbf{K}_{loc}{}^T\mathbf{M}^T$ with localization applied during DA. Equations (B6) and (B7) can be combined and plugged into $\delta\mathbf{y}^T \mathbf{K}_{loc}{}^T\mathbf{M}^T$

$$\begin{aligned}\delta\mathbf{y}^T\mathbf{K}_{loc}^T\mathbf{M}^T &= \delta\mathbf{y}^T\left(\mathbf{P}_{loc}^{a}\mathbf{H}^T\mathbf{R}^{-1}\right)^T\mathbf{M}^T \\ &\approx \delta\mathbf{y}^T\mathbf{R}^{-1}\mathbf{H}(\rho_A \circ \mathbf{P}^{a})\mathbf{M}^T \\ &= (K-1)^{-1}\delta\mathbf{y}^T\mathbf{R}^{-1}\mathbf{H}\left(\rho_A \circ (\mathbf{X}^{a}\mathbf{X}^{aT})\right)\mathbf{M}^T\end{aligned} \tag{B8}$$

From Kalnay et al. (2012), equation (B8) then becomes the following form within the final EFSO expression.

$$\delta\mathbf{y}^T\mathbf{K}_{loc}^T\mathbf{M}^T \approx (K-1)^{-1}\delta\mathbf{y}^T\mathbf{R}^{-1}\left[\rho_I \circ \left(\mathbf{Y}^{a}\mathbf{X}^{fT}\right)\right] \tag{B9}$$

Together, (B8) and (B9) reveal the implicit relationship between the two types of localization, thus the EFSO localization ρ_I is dependent upon ρ_A in addition to other dependencies (spatial, cross-variable, time-forecast component). This can be summarized in the following expression.

$$\rho_I \circ \left(\mathbf{Y}^{a}\mathbf{X}^{fT}\right) \approx \mathbf{H}\left(\rho_A \circ (\mathbf{X}^{a}\mathbf{X}^{aT})\right)\mathbf{M}^T \tag{B10}$$

The best EFSO estimates depends on finding the EFSO localization ρ_I such that (B10) is a good approximation, as well as the degree to which of the implicit assumption in (B7) is true.

Appendix C: Details of GSI observation error adjustment and quality control check

The GSI system is used prior to running the EnKF to precompute the observation innovations as well and perform various quality control (QC) and observation error adjustments. The procedures summarized here are also described in the GSI User's Guide (Hu et al. 2017) and the Advanced User's Guide (Developmental Testbed Center 2016).

The QC method is a gross error check based on the magnitude of observation innovations and the (adjusted) observation error. The user sets three parameters – *gross*, *ermax*, and *ermin* – for each type of observations. Parameters *ermin* and *ermax* define the minimum and maximum standard deviation error values for each type of observation, respectively, to limit the range of values. Thus if an observation error is below *ermin* (above *ermax*), the GSI will automatically reset the observation error to *ermin* (*ermax*). The *gross* parameter defines the standard deviation threshold for which observations are rejected based on the ratio of the observation innovations (observed-minus-background values) to the adjusted observation error for each value. For example, consider a temperature observation has an observation error of 0.5 ˚C and an associated innovation of 2.0 ˚C. The innovation-over-obserror ratio is then 2.0/0.5 = 4.0 ˚C. If the *gross* parameter for this observation were set to 3.0 ˚C, this observation would then be flagged for rejection since this ratio (4.0 ˚C) is larger than the *gross* parameter. Conversely with a more relaxed threshold of *gross* = 6.0 ˚C, this observation would be accepted for assimilation.

The GSI system also includes automatic adjustment of observation errors. There are several methods of automatic tuning and inflation of observation error (see Developmental Testbed Center 2016), though by default these options are not used and were not used within this dissertation. However, surface observation errors are automatically inflated based on discrepancies in reported pressure or height of the surface observations compared to the background grid. Surface observations are automatically placed at 2-m height above the model surface terrain for temperature, pressure, and moisture, and at 10-m for wind, consistent with the real measurement heights. Since it is unknown what height nonconventional surface observations are taken at we assumed a 2-m height for all nonconventional surface observations, *including wind*. For temperature and moisture surface observations, the observation error inflation is based upon the difference between the station observation surface pressure and the background model surface pressure. Larger differences result in increased observation error inflation. For wind and pressure observations, the observation error inflation is based upon the difference between the reported height of the observation and the height of background grid terrain at the observation location. An increased penalty is imposed upon surface observations that are found to be below the model surface.

As an example of this inflation procedure for surface observations, consider an ERNET surface moisture observation with a reported pressure of 964.0 mb. At the location of this observation, the model background has a surface pressure of 971.4 mb, resulting in a difference of 7.4 mb. Assuming the observation passes the

gross error check, the GSI adjusts this observation error from the original value that was set at 5.0% (in terms of RH) to 12.97% as a result of this discrepancy in surface pressures.

One important consideration affecting wind observations specifically is how the background field is computed at the assumed observation height. The calculation of background wind is based upon the 10-m wind read in from the background field. Note that the 10-m wind calculation occurs during WRF integration and is dependent on the chosen surface physics scheme. For surface observations above 10-m and below the first model level (approximately 25-30 m AGL), the computed background wind is a linear interpolation of the 10-m wind and first model level wind to the observation height. For surface observations below 10-m, the background wind is a linear interpolation of the 10-m wind and the wind at the surface (zero due to no-slip condition). In the case of the 2-m height assumption for nonconventional observations, the background wind is thus reduced to 0.2 times the magnitude of the 10-m wind.