

**Data Assimilation
in the
NCAR Community Atmosphere Model**

by

Justin E. S. Bagley

A thesis submitted in partial fulfillment of
the requirements for the degree of

Master of Science
(Atmospheric and Oceanic Science)

At the

UNIVERSITY OF WISCONSIN-MADISON
2008

Approved by:

(Eric DeWeaver)

Professor, Atmospheric and Oceanic Sciences

Date:

Abstract

The Community Atmosphere Model (CAM) has several well known biases that have important dynamical implications for research. Some examples of these biases include a cold temperature bias near the tropopause at high latitudes, systematic biases in the mean zonal wind, and transient momentum fluxes that are too strong. One method for investigating the impacts that these biases have is to use data assimilation techniques to reduce the bias and observe how the system reacts.

The cold polar tropopause temperature bias found in CAM is common to climate models. For the Northern Hemisphere, the bias is generally observed in the summer months. By assimilating a July 2003 set of observations in the region of this bias, we observe that reducing this bias forces the zonal jet and storm tracks to shift poleward. This finding mirrors the results of several earlier studies where the height of the polar tropopause influences the location of storm tracks and zonal jets by moving them equatorward as the height decreases, and poleward as the height increases as is predicted to occur with increased atmospheric concentrations of greenhouse gasses.

With the recent CHAMP/COSMIC missions, GPS radio occultation observations have become available. These observations have many desirable qualities including their global nature and accuracy in polar regions where quality observations are rare. In this research, we determine how strongly these observations can constrain CAM through data assimilation. We also begin to quantify the process by which these assimilated observations impact the model.

We use an ensemble adjustment Kalman filter (EAKF) provided by the Data Assimilation Research Testbed (DART) to assimilate COSMIC data into CAM for January 2007. We compare a case where only COSMIC observations are assimilated into CAM with cases where we assimilate observations from multiple sources. We find that GPS observations do improve our analysis. Further, when taken by themselves, the assimilation of GPS observations is capable of constraining CAM to a substantial degree. However, we find that water vapor, which has a strong relationship to refractivity is poorly constrained and assimilated, while dynamical quantities are more accurately assimilated.

Finally, we take a detailed look into the mechanism by which GPS observations impact the state of CAM's modeled atmosphere. By linearizing the refractivity forward operator, we use multivariate linear regression to piece apart the components of the operator. Through this process we determine how the assimilation of a GPS observation will influence the state of our system differently depending on its location in Earth's atmosphere. In particular large cancellations due to correlated temperature and pressure variations in the midtroposphere of the midlatitudes seem to reduce the potential impact of assimilated GPS observations in this location.

Acknowledgments

I cannot claim sole responsibility for the completion of this thesis. Many others have made significant contributions to this cause, at a variety of levels. First I would like to acknowledge the professors of the AOS department for providing engaging, and academically challenging courses that supplied the backbone of knowledge necessary to complete this thesis. In particular I would like to thank the faculty readers of this thesis for their thoughts on improving this document.

Another university group that made this thesis possible is my fellow grad students and friends who provided both academic support, and mental sanity.

Also, a huge thank you to Eric DeWeaver for outstanding courses, always being available for strong advice, and the many engaging conversations about the ins and outs of data assimilation and climate that we have had over the last few years.

And while it could care less about my thesis, Madison as a city also deserves a thank you for providing entertainment, and long winters that were conducive to actually getting work done.

For anyone reading this, thank you for not letting a couple years of my life sit lonely on the library shelf.

Equally important were the students of Opelousas High School for giving me perspective on priorities.

Thank you also to my family for encouragement and understanding of what I was doing.

Finally, and most importantly, thank you to my wife Elizabeth for essentially everything. You made this thesis worth doing.

Table of Contents

Abstract.....	i
Acknowledgements.....	ii
Table of Contents.....	iii
1. Introduction.....	1
2. Ensemble Kalman Filter Data Assimilation.....	5
a) <i>Overview of data assimilation and the ensemble Kalman filter</i>	5
b) <i>The ensemble Kalman filter</i>	10
c) <i>Bayesian statistics and the ensemble adjustment Kalman filter</i>	13
c) <i>Practical implementation of the ensemble adjustment Kalman filter</i>	16
3. Part 1- Evaluating Impacts of Arctic Model Bias.....	24
a) <i>Motivation for investigation of polar bias</i>	24
b) <i>Experimental setup and description of observations</i>	27
c) <i>Results</i>	30
d) <i>Conclusions on the impacts of Arctic model bias</i>	39
4. Part 2- Assimilating GPS Occultation Observations in CAM.....	42
a) <i>Introduction to GPS occultation applications to climate research</i>	42
b) <i>Description of GPS observations and experimental setup</i>	49
c) <i>GPS observations and CAM model bias results</i>	51
d) <i>Covariance structure of linearized refractivity and the state vector</i>	62
e) <i>Conclusions on the assimilation of GPS observations in CAM</i>	79
5. Conclusion.....	84
a) <i>Summary and future work</i>	84
6. References.....	86

1. Introduction

A fundamental property of every model ever built is that it has imperfections. Whether through inexact initial or boundary conditions, errors in model formulation, incomplete parameterizations, or unknown physics, error is always present and must be taken into account in every type of scientific model. Nowhere is this property of models more apparent than when investigating the complex interactions that make up the Earth system where nonlinearities can balloon small errors into large errors. Lorenz states in his seminal 1963 paper:

“When our results concerning the instability of nonperiodic flow are applied to the atmosphere, which is ostensibly nonperiodic, they indicate that prediction of the sufficiently distant future is impossible by any method, unless the present conditions are known exactly. In view of the inevitable inaccuracy and incompleteness of weather observations, precise very-long range forecasting would seem to be non-existent.”

In a weather forecast, the errors associated with initial conditions and imperfect models coupled with nonlinearities in the atmospheric system currently limit the timescale of useful prediction to the order of weeks. These errors manifest themselves differently in global climate studies where models may be run for hundreds of years and the error associated with the forecast of a model on a specific day is of minimal concern. In this case, relevant errors appear most clearly as systematic biases in the time-averaged model state. Consequently, these systematic biases can have critical implications for long-term studies.

While the statement that all models are imperfect is unequivocal, the same is true for physical observations. Every observation has some degree of error associated with it.

The errors can be introduced through instrumental deficiencies, poor retrieval schemes, or human error to name a few possibilities.

In other words, the two principal tools for understanding and investigating the Earth system are fundamentally imperfect. Hence, we are left in the unenviable position of being forced to validate flawed models with flawed observations, or conversely, evolving inexact observations forward in time or across space with imperfect models. However, this quandary does not diminish the fact that each of these tools contain a tremendous amount of useful information that can be extracted and used to help constrain and define what we know about a system.

With the increase in computing speed and the advent of satellites over the past few decades, the amount of information available to a researcher from both models and observations has increased exponentially. In the past, the connection between model and observations was mostly hidden away in the physics and parameterizations driving the model dynamics. However, in an effort to investigate the causes of model bias, reduce the discrepancy between modeled and observed states of the atmosphere, and quantify the error introduced by imperfect observations, a more intimate link between models and observations in the form of data assimilation has been developed. Data assimilation is one tool used to maximize the amount of information that can be pulled from observations and models. Data assimilation can be defined as “using all the available information, to determine as accurately as possible the state of the atmospheric (or oceanic) flow.” (Talagrand 1997) This is an extremely powerful yet limiting statement. In reality data assimilation has applications beyond state estimation ranging from parameter estimation (Annan et. al. 2005) to observation system design (Morss and

Battisti 2003). This research examines a couple of these applications. The first application this research considers is the use of data assimilation to quantify the dynamical implications of a climatological bias in the National Center for Atmospheric Research's (NCAR's) Community Atmosphere Model. The second is an assessment of how the assimilation of Global Position Satellite (GPS) radio occultation observations adjusts the zonal mean circulation of the previously mentioned model.

To begin this discussion Chapter 2 gives a brief background of data assimilation with an emphasis on the Ensemble Adjustment Kalman Filter (EAKF). This chapter will derive the basic concepts associated with data assimilation, and also illustrate how it can be practically implemented. Chapter 3 provides an investigation into how data assimilation can be used to help assess the dynamical implications of a climatological temperature bias that is common to climate models. This section will discuss the motivation for looking at the polar tropopause bias, theoretical considerations for this case, and how data assimilation can be used to quantify the impacts of this bias. The second application of assessing the impact of GPS occultation measurements will compose Chapter 4. This section discusses the application and retrieval of GPS soundings, how well assimilation of GPS occultation measurements can constrain the atmospheric circulation and how characteristics of GPS refractivity interact with the atmosphere to provide different types of information at different points in the atmosphere. Finally, Chapter 5 summarizes the findings of this research.

Through the outline given, there are several specific questions that this research attempts to answer. These questions will be expanded in detail in Chapters 3 and 4. In Chapter 3 the focus of the chapter will be on determining whether a local assimilation of

observations in the Polar tropopause can reduce a strong model bias in the Community Atmosphere Model (CAM) in the region, and how the reduction of this bias impacts the large scale circulation through secondary effects. Chapter 4, will determine how strongly assimilated GPS radio occultation observations constrain CAM, and how the variations in the model state across an ensemble determine the impact that GPS radio occultation observations have in different regions of the atmosphere. In addition, Chapter 4 will investigate the possible problem of moisture being dominant in refractivity yet hard to assimilate in models.

2. Ensemble Kalman Filter Data Assimilation

a. Overview of data assimilation and the ensemble Kalman filter

At the root of data assimilation is a desire to combine a model and observation in order to determine the state of a system as accurately as possible. To illustrate the concepts, it is useful to examine a straightforward example. To begin assume that we are looking at a temperature, T , at a single model gridpoint. There are two estimates of T , one from a global model and one from an observation, both of which have errors associated with them. The next step is finding the most likely temperature given these two pieces of information. Following Talagrand (1998), this is expressed mathematically as

$$Y_1 = T^t + v_1, \quad Y_2 = T^t + v_2. \quad (2.1)$$

Where T^t is the true temperature, Y_1 is the model estimation, Y_2 is the observation, and $v_{1,2}$ are the errors associated with each. In this case, we will assume that the expected error is zero $E(v_{1,2})=0$, i.e. the observations are unbiased. In addition, the variances are known, $v_{1,2}$, $E(v_{1,2}^2)=\sigma_{1,2}^2$. Finally, assume the errors are uncorrelated, $E(v_1 v_2)=0$. The estimate of the actual temperature which we will label T^a (analysis temperature), can be assumed to be a linear combination of the two estimates,

$$T^a = a_1 Y_1 + a_2 Y_2, \quad (2.2)$$

where the weights $a_{1,2}$ will be determined. We want our analyzed estimate of temperature to be an estimate, which minimizes the variance of the difference between itself, T^a , and

the truth, T^t . By minimizing $\sigma^2 = E[(T^a - T^t)^2]$ we can show that the weights a_1 and a_2 are given by

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \quad (2.3)$$

and the estimation error is given by

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}. \quad (2.4)$$

We can now comment on some important aspects of data assimilation. If we substitute the proper expressions for $a_{1,2}$ into $T^a = a_1 Y_1 + a_2 Y_2$, we find the most probable state of the temperature to be:

$$T^a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} Y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} Y_2 \quad (2.5)$$

Inspection of equation 2.5 reveals that if an observation has a large expected error associated with it, the observation will be weighted far less than an observation that has small errors associated with it. Equation 2.4 also provides an estimate of the error, σ , of the analyzed temperature, T^a . Since all assimilation schemes will have some error associated with them as they are assimilating non-perfect models and observations, getting a quantitative estimate of the analyzed error is extremely important.

The previous example gives a simple demonstration of the purpose of data assimilation, however in a practical application there could be thousands of variables at thousands of gridpoints that need to be assimilated with observations that may not be

directly related to the quantities we are interested in. In this multivariate case, covariance relationships must be considered in addition to simple variance from the 1-d case. In order to do this, a generalization of the concepts is necessary. To begin, it is trivial to generalize the initial equation $Y_1 = T^t + v_1$. First, expand Y_1 to be a vector \mathbf{Y} that contains all the observations, $\langle Y_1, Y_2, \dots \rangle$. Similarly, expand T^t to be a vector that contains the entire state of our system as opposed to just temperature. This will be designated \mathbf{X} , and we will refer to this as the state vector. v_1 is also expanded to a vector that represents the error associated with each observation. Finally, we need some way to connect the state variables to observations that are not directly related. We will designate the matrix that operates on the vector \mathbf{X} in such a way as to project the vector from state space to its analog in observation space, as \mathbf{H} . \mathbf{H} is known as the forward operator.

Equation 2.2 now becomes

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + v_t \quad (2.6)$$

where the subscripts indicate these quantities are all evaluated at time t . Returning to the simple example above, instead of an observation of temperature, say instead there exists an observation of radiance that is indirectly related to the temperature. To assimilate this observation, we operate on T^t with \mathbf{H} . In this case, $\mathbf{H}T^t$ transforms the state temperature into an observed radiance, and now the previously stated method can be followed exactly. At this point we need to separate the estimate from our observation from the estimate of the unanalyzed model. The unanalyzed, or background estimate, can be written as

$$\mathbf{X}_t^b = \mathbf{X}_t^t + v_t^b \quad (2.7)$$

As before, the final important piece of information required is the error covariance that is written as

$$\Sigma = \begin{bmatrix} E(\mathbf{v}^b \mathbf{v}^{bT}) & E(\mathbf{v}^b \mathbf{v}^T) \\ E(\mathbf{v} \mathbf{v}^{bT}) & E(\mathbf{v} \mathbf{v}^T) \end{bmatrix}. \quad (2.8)$$

where the superscript, T, indicates a transpose. Although the mathematics now involves matrices and vectors, the same rules are applied as in the original example. Through a series of matrix manipulations we can find the analyzed state, \mathbf{X}_t^a and the analyzed error covariance, \mathbf{P}_t^a . The expressions for these quantities are given in Kalnay (2003) as

$$\mathbf{X}_t^a = \mathbf{X}_t^b + \mathbf{P}_t^b \mathbf{H}_t^T [\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t]^{-1} (\mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_t^b), \quad (2.9)$$

$$\text{and } \mathbf{P}_t^a = \mathbf{P}_t^b - \mathbf{P}_t^b \mathbf{H}_t^T [\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t]^{-1} \mathbf{H}_t \mathbf{P}_t^b, \quad (2.10)$$

where $\mathbf{P}^b = E(\mathbf{v}^b \mathbf{v}^{bT})$, and $\mathbf{R} = E(\mathbf{v} \mathbf{v}^T)$ is the observation error covariance. It was also assumed $E(\mathbf{v}^b \mathbf{v}^T) = E(\mathbf{v} \mathbf{v}^{bT}) = 0$ (i.e. the errors between observations and modeled data are uncorrelated). The superscript “a” represents an analysis quantity formed after an observation is assimilated, and superscript “b” represents a background quantity formed before an observation is assimilated. The quantity $\mathbf{P}_t^b \mathbf{H}_t^T [\mathbf{H}_t \mathbf{P}_t^b \mathbf{H}_t^T + \mathbf{R}_t]^{-1}$ is known as the Kalman gain matrix (Anzell and Hakim 2006). Equations 2.9 and 2.10 in a slightly modified form are classified as the Kalman filter equations. If one can calculate all of the above quantities, these equations give the best linear unbiased estimate of the state and the error covariance at time, t (Kalnay 2003). However, there are a few fundamental problems that need to be addressed.

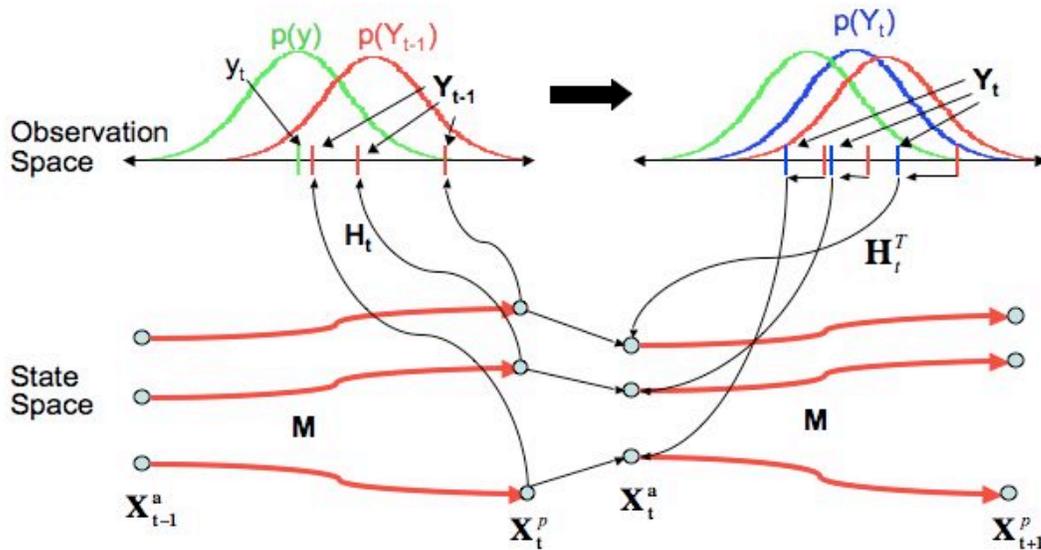


Figure 1: Schematic of the assimilation process for ensemble Kalman filters. Shows the evolution of a three-member ensemble. Process begins with the model moving the initial ensemble forward in time. Then the ensemble is transformed to observation space using the forward operator, H_t . A PDF is formed from the transformed ensemble assuming a Gaussian (red curve), and is assimilated with the observation PDF (green curve) forming the analysis PDF (blue curve). The analysis ensemble is now extracted from the analysis PDF, and is transformed back to state space. The ensemble is now updated and ready to repeat the process using the model M to move the ensemble forward in time.

The first problem is how to obtain the background estimate of the state at time, t .

The relatively simple answer is to use climatological data for the initial estimate, and then assimilate the initial observations, which gives a new estimate of the state vector \mathbf{X}^a .

Initialized from this newly analyzed state, it is trivial to integrate a model forward until new observations become available, at which time the above process is repeated. To describe this mathematically, we need a discrete representation of how the state, \mathbf{X}_t , will propagate forward in time. Following Jazwinski (1970), the evolution of the state can be written as

$$\mathbf{X}_{t+1}^b = M\mathbf{X}_t^a + \eta_t. \quad (2.11)$$

In this equation, M is the forecast model being used (in this research M will be the Community Atmosphere Model), and the second term represents the model error

(Anderson 2003). A far more difficult question is how the error covariance is determined and how it will change through time.

b. The ensemble Kalman filter

When trying to determine how the error covariance should advance in time, the simplest solution is to simply hold it constant with climatological values. This is done in the commonly used in three dimensional variational assimilation systems (3D-Var), which determines the best guess of the atmospheric state by using an adjoint model to minimize the variance of the analysis (Lorenz 1981). However, using climatological values leads to a loss of information in the form of “errors of the day”, which describe how the background error covariance changes with the flow of the atmosphere (Kalnay et al. 2006).

One example that is often mentioned of an example of how “errors of the day” can impact an assimilation is the case of synoptic-scale wave structures. Since traditional assimilation systems use climatological background error covariances, this causes the background error covariance in the midlatitudes to look like the large-scale trough-ridge circulation. In contrast, as will be discussed shortly, recent ensemble Kalman filter systems use the “errors of the day” which allows the synoptic-scale error covariances associated with the actual wave structure to be captured. This has been shown to lead to improved analyses (Zhang and Anderson 2003).

Classically, it has been shown that in Kalman theory that the error covariance evolves according to the equation

$$\mathbf{P}_t^b = \mathbf{F}_{t-1} \mathbf{P}_{t-1}^b \mathbf{F}_{t-1}^T + \mathbf{Q}_{t-1}. \quad (2.12)$$

Where \mathbf{F} is a tangent linear operator for updating the system state, and \mathbf{Q} is the model error covariance matrix (Kalman 1960). Equation 2.12 is derived from the Kolmogorov

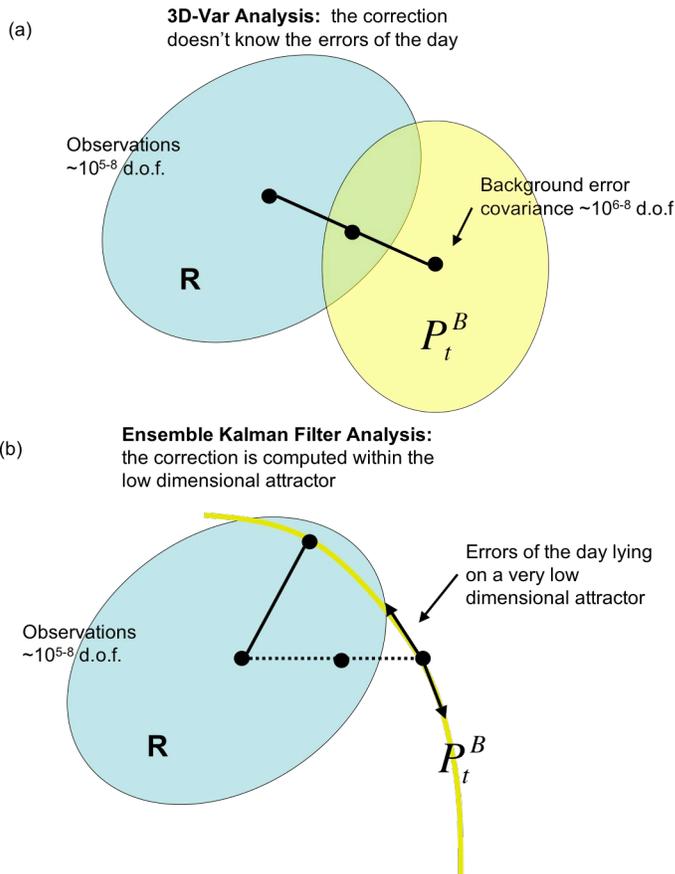


Figure 2: Schematic of the analysis estimate for typical EnKF and 3D-Var data assimilation systems given the background forecast, the background error covariance \mathbf{P}_t^B (climatological values and isotropic for 3D-Var), observations, and observation error covariance \mathbf{R} (generally diagonal for both EnKF and 3D-Var). The analysis estimate maximizes the joint probability distribution. (a) 3D-Var. (b) EnKF, where the ensemble forecast members define a subspace within which the analysis lies. The dots represent adjustments to the ensemble mean. Adapted from Kalnay et al. (2006).

equation that describes the evolution of the probability density function for the analysis state (Evensen 1997). A full derivation of the Kolmogorov equation can be found in Jazwinski (1977).

However, equation 2.12 is an approximation based on a 2nd order truncation of a Taylor series. This approximation has been shown to lead to unbounded error variance growth (Evensen 1994). There are

two other immediate problems with using the Kalman equations in this form for atmospheric data assimilation. First, \mathbf{F}_{t-1}^T is known as an adjoint model, and the development of adjoints for climate

models is non-trivial. Second, \mathbf{F}_{t-1} is on the order of $n \times n$ where n is the number of degrees of freedom for the system. For a global climate model (GCM) this can be on the

order of 10^6 . In essence, using the Kalman equations in this state require a linear tangent model and its adjoint to be run $O(n)$ times, which with current computing power is unfeasible.

To get around the limitations on the classical Kalman filter, the ensemble Kalman filter (EnKF) was originally presented by Evensen in 1994, and since then has morphed into many variations. Now, as opposed to solving the Kolmogorov equation analytically using a truncated Taylor series, it is solved using a Monte Carlo method (Evensen 1994). In practical terms, the initial state of the system is slightly perturbed many times to form an ensemble of states. These states are then each propagated forward in time by the model, as described in equation 2.11. When observations are taken and assimilated, one can calculate the error covariance matrix of equation 2.10 from statistics taken from this ensemble of states. It now becomes

$$\mathbf{P}_t^b = \frac{1}{m-1} \mathbf{X}^b \mathbf{X}^{bT} \quad (2.13).$$

where primed quantities represent deviations from the ensemble mean, and m is the number of ensemble members.

This process is shown schematically in Figure 1. In Figure 1's example, the model, M , integrates a 3-member ensemble forward in time to the point where observations are taken and assimilated (generally ensembles range in size from 20 – 200 members for atmospheric applications). The ensemble state is then transformed to observation space by the operator \mathbf{H}_t from equation 2.6. At this point, the ensemble state is represented by a Gaussian probability distribution function that is fitted to the

ensemble members. The ensemble is then assimilated with the observations, and an adjustment to the ensemble, which will be described in more detail in the next section, is made. Notably, this is the step where most of the various ensemble Kalman methods differ. Lastly, the ensemble is transformed back to model space, and the model integration is repeated for each ensemble member.

By assimilating the observations in this manner, we can continually update the error covariance statistics by performing statistical analysis on the ensemble members. The mean of the ensemble gives the best estimate of the state, and the variance of the ensemble members gives an estimate of the error covariance (Evensen 1994). In this case, the accuracy of the method is only limited by the number of ensemble members. The more ensemble members that are available, the more accurate the representation of the probability distribution function for the state of the system (Evensen 1994). This is practical because the effective dimensionality of the local flow field is small compared to the dimensionality of the model. The ensemble system for estimating the background error covariance also solves the problem of 3D-Var neglecting the errors of the day. With the EnKF, the use of a Monte Carlo analysis scheme allows for a flow-dependent background error covariance. This decreases the dimensionality of the background error and improves forecasts. This is shown schematically in Figure 2. For a more detailed comparison between EnKF and variational assimilation, refer to Kalnay et al. (2007) or Ansell and Hakim (2006). The next step is to look more closely at how the ensemble is updated at a single time, when an observation becomes available to be assimilated.

c. Bayesian statistics and the ensemble adjustment Kalman filter

All data assimilation methods can be formulated in a Bayesian framework (Talagrand 1997). In this case, the complete solution to the EnKF assimilation problem can be written as the probability of a state (\mathbf{X}_t) occurring, given all previous information (\mathbf{Y}_t) i.e. the probability distribution function $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_t)$. In the more general probabilistic case, \mathbf{X}_t is the state of the atmosphere at a given time, and \mathbf{Y}_t is the set of both information from the ensemble and physical observations of the atmosphere that are taken at or near time t (Anderson and Anderson 1999). If we take $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_{t-1})$ to be a known initial condition from which we wish to derive a future state of the system at the point in time when a new set of observations, \mathbf{y}_t , become available, the new conditional probability function will be written as $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_t)=\mathbf{p}(\mathbf{X}_t|\mathbf{y}_t, \mathbf{Y}_{t-1})$. To put the probabilistic formulation in a form directly applicable to data assimilation Anderson and Anderson (1999) uses Bayes' rule:

$$\mathbf{p}(\mathbf{X}_t | \mathbf{Y}_t) = \frac{\mathbf{p}(\mathbf{y}_t | \mathbf{X}_t, \mathbf{Y}_{t-1})\mathbf{p}(\mathbf{X}_t | \mathbf{Y}_{t-1})}{\mathbf{p}(\mathbf{y}_t | \mathbf{Y}_{t-1})}. \quad (2.14)$$

By assuming any error in observations to be uncorrelated for different observation times, the above equation can be rewritten as

$$\mathbf{p}(\mathbf{X}_t | \mathbf{Y}_t) = \frac{\mathbf{p}(\mathbf{y}_t | \mathbf{X}_t)\mathbf{p}(\mathbf{X}_t | \mathbf{Y}_{t-1})}{\mathbf{p}(\mathbf{y}_t | \mathbf{Y}_{t-1})}. \quad (2.15)$$

In this equation, the denominator acts as a normalization factor that forces the total probability of all the possible states to be equal to one. In practice this quantity does not need to be calculated. In the numerator, the first term represents the new information introduced by an observation. This probability distribution function is generally assumed to be Gaussian in order to model instrumental uncertainties. The second term represents all of the information from the ensemble of model runs up to the point of the assimilation.

This is the prior state. For most versions of the EnKF fitting a Gaussian to the ensemble, as shown in Figure 1, forms the probability distribution function (PDF) for $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_{t-1})$.

Equation 2.15 is analogous to equation 2.6 and represents the ensemble adjustment step in Figure 1.

The flavor of EnKF used in this research is the ensemble adjustment Kalman filter (EAKF). This filter was originally designed by Anderson (2003). It can be further classified as an ensemble square root filter (Tippett et al. 2003). Other filters included in this class include the ensemble transform filter (Bishop et al. 2001) and ensemble square root filter (Whitaker and Hamill 2002) among others. It differs from the original EnKF only in the methodology of the assimilation step. If it is assumed that the probability distribution functions described by $\mathbf{p}(\mathbf{y}_t|\mathbf{X}_t)$ and $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_{t-1})$ are Gaussian, then the product of those two PDFs, $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_t)$, will also be a Gaussian. $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_t)$ is the posterior or analysis PDF. In general, this is a reasonable assumption for the observation PDF, $\mathbf{p}(\mathbf{y}_t|\mathbf{X}_t)$, but not so robust for the prior model state PDF $\mathbf{p}(\mathbf{X}_t|\mathbf{Y}_{t-1})$. However, it simplifies the representation of the processes greatly. In this case, the posterior error covariance can be written as

$$\mathbf{P}^a = [(\mathbf{P}^b)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \quad (2.16)$$

and the ensemble mean state as

$$\bar{\mathbf{X}}^a = \mathbf{P}^a [(\mathbf{P}^b)^{-1} \bar{\mathbf{X}}^b + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}] \quad (\text{Anderson 2001}). \quad (2.17)$$

In many ways, the EAKF is very similar to the original EnKF. In both cases, the prior PDF is determined by fitting a Gaussian curve to the prior ensemble, and the observation. However, the EnKF takes a random draw from the observation PDF and associates it with one of the ensemble members (Evensen 1994). This is not done for the

EAKF. Instead, the EAKF uses a deterministic algorithm to shift the entire original ensemble to first have the exact mean of the posterior PDF. Next, we shift the ensemble again to have the exact variance of the posterior PDF. This is done by updating each ensemble member using

$$\mathbf{X}_i^a = \mathbf{A}^T (\mathbf{X}_i^b - \bar{\mathbf{X}}^b) + \bar{\mathbf{X}}^a, i = 1, \dots, m, \quad (2.18)$$

where \mathbf{A}^T is a linear operator selected so that the updated ensemble has the identical error covariance as equation 2.16, m is the number of ensemble members, and \mathbf{X}_i is the state vector for the individual ensemble members (Anderson 2001).

Some advantages of the EAKF over the EnKF are that any possible bimodality in the initial ensemble can be maintained, and random outliers are nonexistent. This derives from the fact that the EAKF only changes the mean and width of the probability distribution without changing its shape. This helps keep the ensemble from diverging, which occurs when analysis errors are underestimated and causes subsequent observations to be underweighted and have little to no impact on the assimilation (Houtekamer and Mitchell 1998). In addition, noise is not introduced through the random draws of the observation PDF. Anderson 2003 has shown that this can cause reduced accuracy in the assimilation.

d. Practical implementation of the ensemble adjustment Kalman filter

One key implication of equation 2.15 is that if observational errors are assumed to be independent (\mathbf{R} is diagonal), then the observations can be assimilated sequentially in any order (Anderson 2003). Furthermore, if \mathbf{X} is considered to consist both of the state

vector, \mathbf{x}^s , and the transformation of \mathbf{x}^s into observations space, $\mathbf{Y}=\mathbf{H}(\mathbf{x}^s)$, (i.e. $\mathbf{X}=[\mathbf{x}^s, \mathbf{Y}]$) equation 2.15 can be reframed as

$$\mathbf{p}(\mathbf{X}_t^a) = \frac{\mathbf{p}(y_t | \mathbf{X}_t^b) \mathbf{p}(\mathbf{X}_t^b)}{\text{normalization}}, \quad (2.19)$$

and the marginal distribution for \mathbf{z} is

$$\mathbf{p}_z(\mathbf{z}^a) = \frac{\mathbf{p}(y, \mathbf{z}^b) \mathbf{p}_z(\mathbf{Y}^b)}{\text{normalization}}, \quad (2.20)$$

where the subscripts in equation 2.20 indicate a marginal probability distribution function for the state in observations space (Anderson 2003). By separating the original Bayes theorem form into marginal probabilities, it becomes clear that we can update \mathbf{Y} separate from \mathbf{x}_s . We can now simplify our approach to assimilating any set of observations in several ways. First, by assuming independent observation errors we can simplify the EAKF equations by sequentially assimilating one observation at a time. As a result, equations 2.16 and 2.17 can be reformulated in reduced forms with the posterior ensemble variance and ensemble mean in observation space now being

$$\mathbf{P}^a = [(\mathbf{P}^b)^{-1} + \mathbf{R}^{-1}]^{-1} \quad (2.21)$$

$$\text{and } \bar{\mathbf{Y}}^a = \mathbf{P}^a \left(\frac{\bar{\mathbf{Y}}^b}{\mathbf{P}^b} + \frac{\mathbf{y}}{\mathbf{R}} \right) \quad (\text{Liu et al 2007}). \quad (2.22)$$

Second, by updating \mathbf{Y} separate from \mathbf{x}^s , and assuming a Gaussian distribution for the prior ensemble, Anderson (2003) assumes that update of the state vector, $\Delta \mathbf{x}^s$, which we will call the analysis increment, is simply a local least squares fit given by

$$\Delta \mathbf{x}_{j,i}^s = \frac{\sigma_{x_j, Y}}{\sigma_{Y, Y}} \Delta \mathbf{Y}_i, \quad i=1, \dots, k; \quad j=1, \dots, M, \quad (2.23)$$

where M is the number of state variables, k is the number of ensemble members, $\sigma_{Y, Y}$ is the ensemble variance of \mathbf{Y} , and $\sigma_{x_j, Y}$ is the covariance of \mathbf{x}_s with \mathbf{Y} for a given

observation. This relationship is

shown schematically in Figure

3. Equation 2.23 has several

desirable qualities. As the

variance of an observation goes

up, the impact that the

observation has on the final

ensemble goes down.

Conversely, the stronger the

covariance between a variable

in observations and the state

vector in state space, the

greater the adjustment to the

state vector. Finally, if a

variable in observation space

has a very small update when an observation is assimilated, then that observation will

also have a very small impact on the state space as well.

In practice it is now a straightforward process to assimilate an observation using the EAKF. Following Liu et al. (2007) the following steps are taken:

- 1.) Use a model to evolve an ensemble of states through time to the point that an observation is available to assimilate (equation 2.11). At this point the ensemble becomes the prior ensemble (\mathbf{X}_i^b , $i=1, \dots, k$ where k is the number of ensemble members).

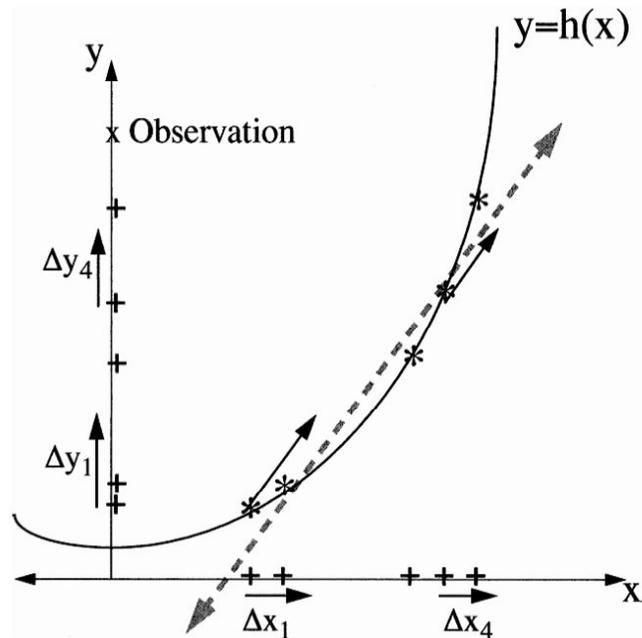


Figure 3: A schematic depicting the relationship between analysis increments for a state variable, x , and an observation y . This example uses a 5 member ensemble. The projection of the ensemble on observation space and state space are represented by the plus signs on the y and x axes respectively. The actual observation is marked with an X on the y -axis. For this example the observation is related to the state variable by a function h . The dashed line shows the least squares fit to the ensemble members. Finally, the analysis increments for the 1st and 4th ensemble members are shown. From Anderson 2003.

2.) Determine a prior estimate of the observation by applying a forward operator to each of the ensemble members ($\mathbf{Y}_i^b = \mathbf{H}(\mathbf{X}_i^b)$, $i=1, \dots, k$).

3.) Find the mean and variance for Y_i^b

$$\bar{\mathbf{Y}}^b = \frac{1}{k} \sum_{i=1}^k \mathbf{Y}_i^b, \quad (2.24)$$

$$\mathbf{P}^b = \frac{1}{1-k} \sum_{i=1}^k (\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b)(\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b)^T. \quad (2.25)$$

4.) Calculate the analysis variance and mean for the ensemble in observation space using equation 2.21 and 2.22.

5.) Update the individual ensemble members by shifting them about the ensemble mean while retaining the mean and variance from step 4. This is done using

$$\mathbf{Y}_i^a = (\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b) \sqrt{\frac{\mathbf{P}^a}{\mathbf{P}^b}} + \bar{\mathbf{Y}}^a. \quad (2.26)$$

Which is the EAKF equivalent of equation 2.18 in observation space.

6.) Find the covariance of the prior ensemble in observation space with state vector in state space and the variance of the ensemble in observation space:

$$\sigma_{x_j, Y}^b = \frac{1}{k-1} \sum_{i=1}^k (\mathbf{x}_{ij}^b - \bar{\mathbf{x}}_j^b)(\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b)^T \quad (2.27)$$

$$\sigma_{Y, Y}^b = \frac{1}{k-1} \sum_{i=1}^k (\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b)(\mathbf{Y}_i^b - \bar{\mathbf{Y}}^b)^T \quad (2.28)$$

7.) Use the results of equations 2.25, 2.27, and 2.28 to calculate the update to the ensemble in state space using a local least squares fit as given in equation 2.23 with $\Delta \mathbf{Y}_i = \mathbf{Y}_i^a - \mathbf{Y}_i^b$.

8.) Finally, the updated ensemble becomes

$$\mathbf{X}_{ji}^a = \mathbf{X}_{ji}^b + \Delta \mathbf{x}_{ji}, i=1, \dots, k; j=1, \dots, M. \quad (2.29)$$

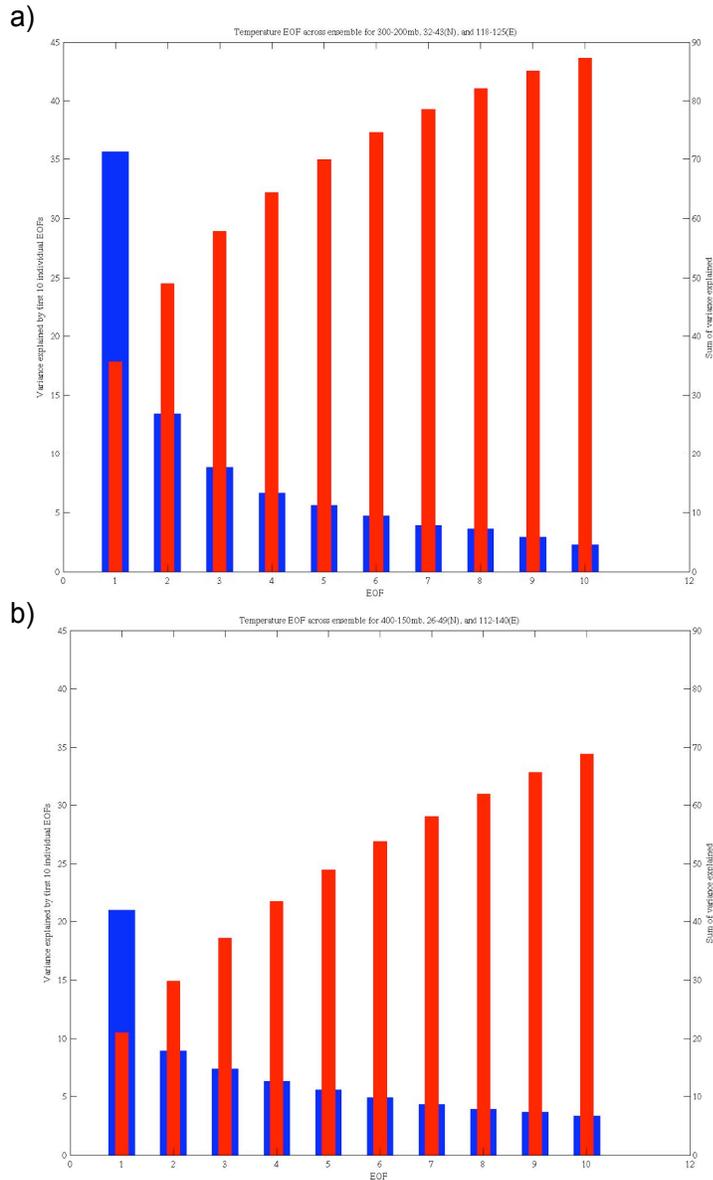


Figure 4: First ten EOF's for temperature with ensemble number as the sampling dimension. The blue lines represent the percent variance explained by the individual EOF, and the red lines indicate the cumulative percent explained. (a) is for a volume of state space from 300-200mb, 32-43N, and 118-128E. (b) is for a volume of state space from 400-150mb, 26-49N, and 112-140E.

Finally, it should be noted that in order to reduce the impact of spurious covariances from geometrically distant data points, which can frequently occur with realistic ensemble sizes, equation 2.27 is multiplied by a distant dependent factor that reduces the covariance to zero after a given distance. This effectively reduces the dimensionality that the covariance relationship of the ensemble needs to capture.

For this research, a Gaspari-

Cohn fifth order polynomial was used (Gaspari and Cohn 1999).

This relationship between dimensionality and ensemble

accuracy can be seen in Figure 4. This figure is taken from a typical DART EAKF assimilation using a 40-member ensemble. The actual system used will be described

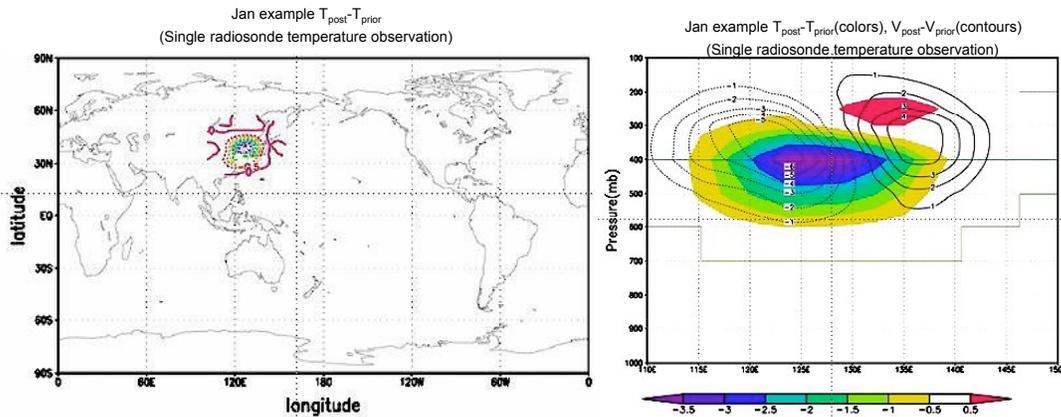


Figure 5: Left panel shows the temperature analysis increment at 400mb for a single assimilated radiosonde temperature (K) observation. The right panel shows a vertical cross section across 40°N for temperature (shaded) and meridional wind(m/s) (contours).

in more detail in Chapter 3b. Figure 4 shows the spectrum of the percentage of variance explained by the ensemble for first 10 EOF's of temperature for two different volumes of state space located near the east coast of China. These volumes were chosen to approximately represent the volume of state space in which an observation is allowed by Gaspari-Cohn distance dependent factor to impact the state vector (this will be evident in Figure 5 as well). In this case the sampling dimension is no longer time, as is normally used in EOF analysis but instead ensemble member number. It is clear that as the volume of effective space increases, the number of EOF's necessary to explain 90% of the ensemble variance increases.

When forming the analysis increment given in equation 2.2 by calculating the error covariance matrices (Eqns. 2.27 and 2.28), it is difficult to see how the increment is physically relevant, and what forces the analysis to be dynamically balanced. For example, Figure 5 shows the result of assimilating a single radiosonde temperature observation at 400mb, again using the DART EAKF system. The left panel shows the horizontal extent that the observation impacts the ensemble's temperature state. In this case, the observation was cooler than the ensemble at that

location and caused a negative analysis increment in temperature. The cooling is roughly symmetric around the location of the observation, and the analysis goes to zero far from the observation as expected due to the use of a distant dependent factor in the covariance equation. Thus if all the ensemble members in the EAKF are in thermal wind balance as given by

$$\frac{\partial \mathbf{V}_g}{\partial \ln p} = -\frac{R}{f} \hat{k} \times \nabla_p T. \quad (2.30),$$

one expects that the background error covariance between temperature and zonal/meridional wind will capture that balance.

The right panel is a vertical slice along the latitude of the observation. This panel shows both the temperature increment and how the meridional wind is updated. It is clear that on the east side of the observation there is a southerly shift in the meridional wind, and on west side of the observation there is a northerly shift. This implies that across the ensemble, the covariance structure between meridional wind and temperature is positive on the west side of the observation and negative on the right. This is anticipated by inspection the thermal wind relation of equation 2.30. This equation shows that for a positive gradient of temperature in the zonal direction, the geostrophic meridional wind should increase with height. This corresponds to the analysis increment to meridional wind on the east side of the observation. There, the observation causes a positive gradient in the analysis increment for temperature, the error covariance structure of the ensemble forces the thermal wind structure to be retained and thus accelerates the meridional wind with height. On the west side of the observation an equal but opposite gradient in the analysis increment for temperature exists and the relation also holds. This indicates that the background error covariance

has in fact captured the thermal wind structure. Thus the covariance structure of the ensemble forces the analysis increment for the state to be, at least approximately, dynamically balanced.

3. Part 1- Evaluating Impacts of Arctic Model Bias

a. Motivation for investigation of polar bias

The extratropical surface westerlies are a critical component of the climate system because of their influence on local climate, air-sea gas exchange, ocean circulation, and

other climatic impacts. Accurate

simulation of the strength,

location, and seasonality of the

surface westerlies is typically

achieved through ad hoc "tuning"

of surface and gravity wave drag,

tuning which sometimes yields

imperfect results (Robinson

1997). In the current NCAR

model the Northern Hemisphere

surface westerlies are too strong

by up to a factor of two (Yeager et

al. 2006). Also, the tuning is

resolution dependent since surface

westerlies usually shift poleward

with increased resolution (Hack et

al. 2005). Beyond these specific

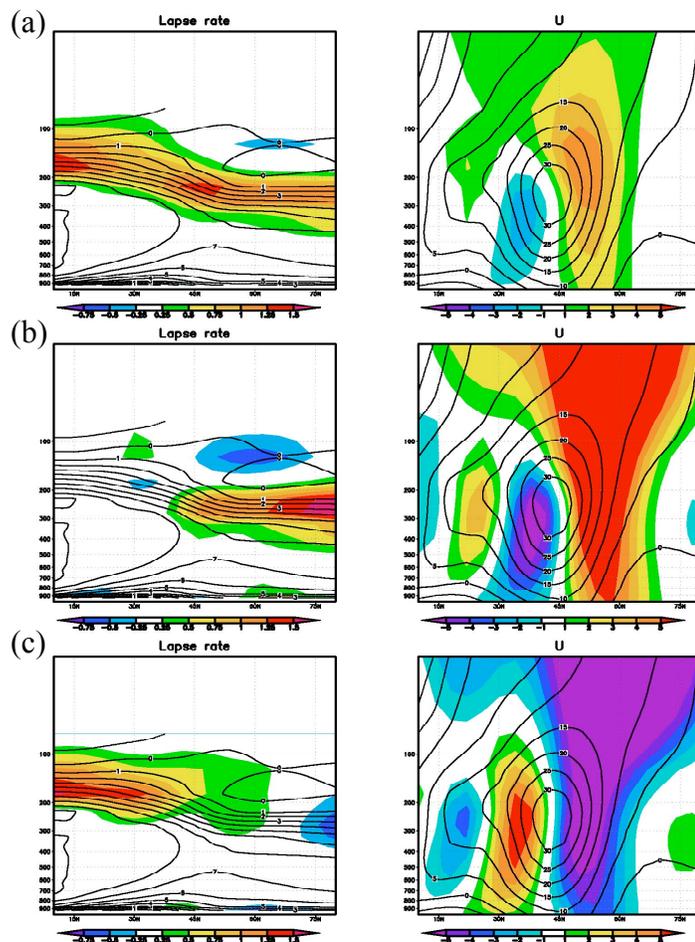


Figure 6: Left panels shows change in zonal mean lapse rate (K/km) from 20th to 21st century (shaded), and modeled 20th century lapse rate (contours) from a simple GCM for (a) globally raised tropopause, (b) tropopause raised only near the pole, and (c) only near the equator. The right panels show the corresponding images for zonal wind (m/s) in place of lapse rate. Adapted from Lorenz and DeWeaver 2007 figure 8.

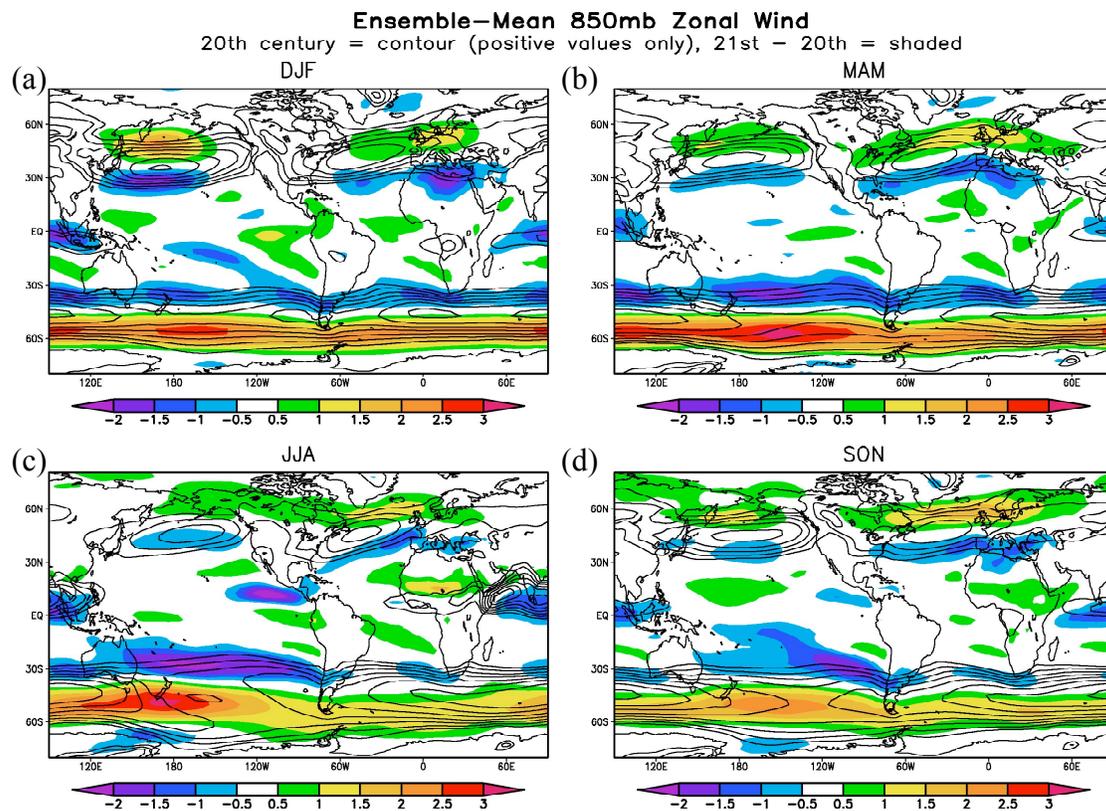


Figure 7: Ensemble-mean of IPCC 850mb climatological zonal wind (m/s) for the 20th century (contour), and the difference between 21st vs. 20th century (shaded). (a) is Winter months, (b) Spring, (c) Summer, and (d) Fall. From Lorenz and DeWeaver (2007) their figure 4.

problems, the tuning of the westerlies may obscure fundamental relationships between the strength and location of the westerlies, heat and momentum fluxes due to storm tracks and stationary waves, and other dynamical factors. In recent years, this problem has risen in prominence in the climate community due to the observed poleward shift of surface westerlies in the Southern Hemisphere (Thompson and Solomon 2002), evidence of past shifts accompanying changes in paleoclimate, and a consistently modeled poleward shift for the next century in response to anthropogenic greenhouse gas increases (Yin 2005).

Accompanying this poleward shift in surface westerlies is a corresponding shift in storm tracks and large-scale zonal wind throughout the troposphere. According to Chen et al. (2007), there have been several hypotheses presented to explain the cause of this behavior. These hypotheses include changes in temperature gradients in the stratosphere

(Polvani and Kushner 2002) and near the surface (Son and Lee 2005), increased latent heating (Frierson et al. 2006), and changes in tropopause height (Williams 2006, Lorenz and DeWeaver 2007). For the purposes of this research we will restrict ourselves to investigating only how changes in tropopause height can impact surface westerlies and zonal wind throughout the troposphere.

Using model simulations from the IPCC archive, Lorenz and DeWeaver (2007) demonstrated a close association between the radiatively forced rise in tropopause height due to global warming and the concurrent poleward shift of the surface westerlies and zonal jet. This association was reproduced in a simple "dry GCM" in which tropopause height could be externally controlled. Moreover, Lorenz and DeWeaver found that the height of the tropopause at high latitudes was the dominant driver of the position of the zonal jet and surface westerlies when compared with changing the tropopause height at lower latitudes. Their results are shown in Figures 6 and 7 for reference. This association is of interest for the simulation of surface westerlies because most current global climate models have a cold bias at the polar tropopause, which is equivalent to a high bias in tropopause height (Covey et al. 2004). Thus, we speculate that correcting the cold polar tropopause bias should result in a (possibly undesirable) southward shift of the surface westerlies and zonal jet.

This chapter will use the EAKF and an atmospheric global climate model (AGCM) for the purpose of identifying, quantifying, and analyzing the impacts of the polar tropopause bias. This will be demonstrated by selectively assimilating observations in the region of the bias, and then observing the resulting shifts in atmospheric circulation. We will begin with Section b describing the model, observations, and data

assimilation runs used in this part of the study. Section c presents the results of the experiment. Finally, Section d describes the conclusions that can be drawn from the results.

b. Experimental setup and description of observations

The tools for this study consists of three main components that include an AGCM forward model for moving the ensemble forward in time, physical observations, and the EAKF framework to assimilate observations as described above. The AGCM used for this research is the third version of the Community Atmosphere Model (CAM3). This model makes up the atmospheric component of the Community Climate System Model (CCSM3). However, we use it as a standalone model. A full scientific description of CCSM3 can be found in Collins et al. (2005). This model can be run at various resolutions with finite-volume, semi-Lagrangian spectral, or Eulerian dynamics. This research uses CAM3 exclusively with Eulerian spectral dynamics at T42 resolution. This resolution was found to have the beneficial properties of both consistent results and reduced computational cost. Also, CAM does exhibit a clear bias in Arctic tropopause temperature. Therefore, by assimilating data in the region of the bias, we can investigate the impacts of this bias.

The observations assimilated in this experiment include a subset of the observations used as input to the NCEP Global Data Assimilation System for the purpose of creating meteorological analyses and operational forecasts. Specifically this research uses radiosondes (U,V,T), surface pressure observations, conventional aircraft wind and temperature observations, ACARS (U,V,T), and satellite winds. We also use the NCEP

Reanalysis product for comparison purposes as described in Kalnay et al. (1996). The reanalysis data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <http://www.cdc.noaa.gov/>.

The EAKF framework and forward operators for the observations were provided by the Data Assimilation Research Testbed (DART). DART is a data assimilation software package being developed by the Data Assimilation Research Section (DARes) of NCAR, and is freely available from <http://www.image.ucar.edu/DARes/DART/>. It includes several data assimilation schemes beyond the EAKF, including the original EnKF, and is compatible with many common atmospheric models.

In order to analyze the impact of a cold Arctic tropopause bias in CAM, three experiments were completed. Each of these experiments used a 40-member ensemble. These experiments all covered the time period from July 1st, 2003 through July 25th, 2003. This period was chosen because the Polar tropopause bias in the Northern Hemisphere is most pronounced during the summer months, and this particular period exhibited a clear bias with respect to NCEP reanalysis as shown in Figure 8. Each of the experiments was initialized with an identical 40-member ensemble. This initial ensemble was generated by randomly pulling forty July 1st states from a 100-year climatology CAM run. This process resulted in an ensemble with a large initial spread, and a mean state that is nearly climatology. Observations were assimilated up to 100mb. The three experiments are as follows:

- Experiment 1 was a control run, and was completed assimilating no observations.

In this experiment the initial ensemble was moved forward in time using CAM.

To remain consistent with subsequent experiments, output data was collected every six hours. This experiment will be referred to as the CAM run.

- Experiment 2 assimilated all the observations described above at all locations on the globe. In this experiment all the available observations were assimilated in six-hour

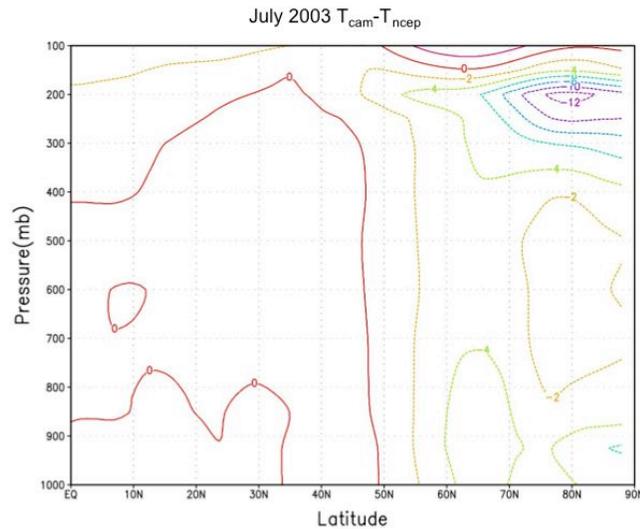


Figure 8: Average zonal mean temperature difference (degrees K) between CAM and NCEP reanalysis in the Northern Hemisphere, for July 2003.

increments. This experiment will be referred to as the full assimilation.

- Experiment 3 restricted the assimilated observations to those observations which were found north of $66^{\circ}N$ and between 300-100mb. This region was chosen to be representative of the cold polar tropopause bias shown in Figure 8. The relevant observations were also assimilated in six-hour increments. This experiment will be referred to as the Arctic tropopause assimilation.

These experiments were run and analyzed at the NERSC supercomputing site.

The runs were completed on the Jacquard machine. This computing power was necessary for both storage and processing requirements. On average, a one-day run of the full assimilation with 40 ensemble members required 10 compute nodes with two processors apiece approximately 12 hours of wallclock time to complete. Thus depending on the user load and queue time, a one-month run could take upwards of three weeks of real time to complete. Also, day outputs about 5Gb of data to be

processed and analyzed. This development of the data was done on the Davinci machine at NERSC, which provides terabytes of work space and most of the computational tools required for climatological data sets.

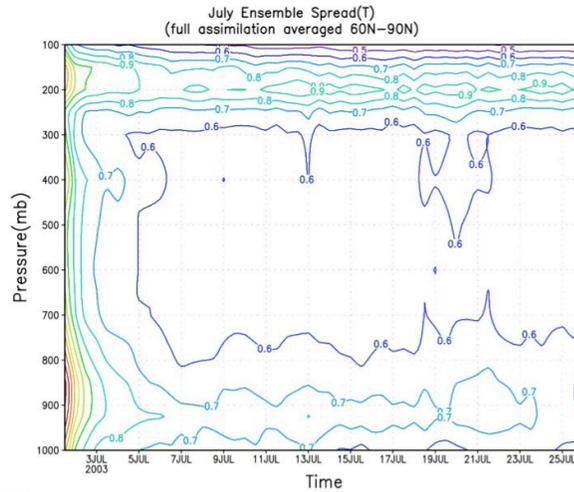


Figure 9: July 2003 ensemble spread in temperature (K) averaged from 60N-90N for full assimilation.

c. Results

When assimilating observations, the ensemble weights the impact of the observations by their uncertainty relative to the spread or uncertainty in the ensemble members. Since our initial ensemble had a very large initial spread, the first observations assimilated caused a large adjustment to the ensemble. However, even with a relatively uncertain ensemble, the shift from a nearly climatological distribution of ensemble members to a distribution that represents the best guess of the filter to the actual state of the atmosphere takes many assimilation cycles covering several days. Determining the time scale of this adjustment period is important in order to be able to effectively filter out data from the time period where the ensemble is adjusting. Also, since we have several different experiment runs as described above, this adjustment time scale could change depending on the experiment. In order to obtain consistent results the varying times need to be determined for each experiment.

One possible measure of the ensemble adjustment time is the ensemble spread. For temperature it is given by

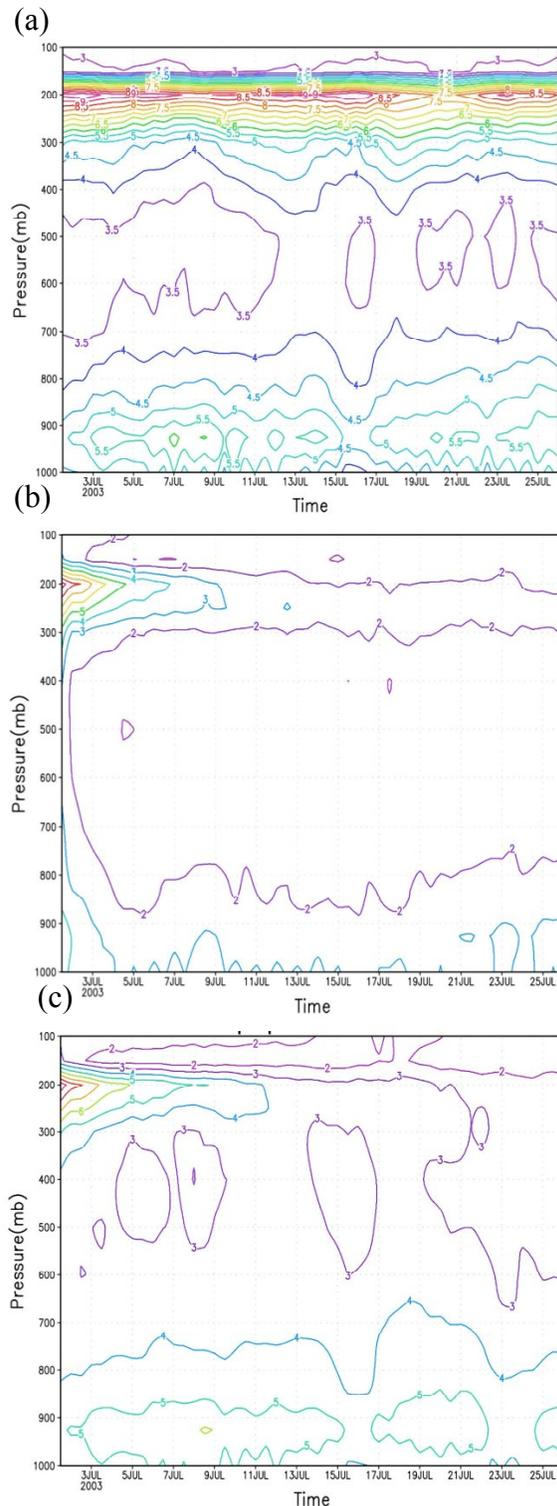


Figure 10: July 2003 RMSE in temperature (K) averaged from 66°N to 90°N for (a) CAM run, (b) full assimilation, and (c) Arctic tropopause assimilation

$$\sigma_{i,t}^T = \sqrt{\frac{1}{N_e - 1} \sum_{j=1}^{N_e} [T_{t,i}^j - \bar{T}_{t,i}]^2}, \quad (3.1)$$

where j represents an ensemble member, N_e is the ensemble size, i represents individual grid points. This metric essentially calculates the model uncertainty for temperature at a given time. This measure is particularly relevant since a combination of the ensemble spread and the uncertainty of a given observation determine the impact that an observation will have on ensemble. For the full assimilation experiment, the area average of the ensemble spread for temperature is shown in Figure 10. In this case, it appears that the ensemble spread is quickly reduced as observations are assimilated. The time scale for this reduction in ensemble spread appears to be around 5-7 days, slightly shorter than the time required to reduce the

Arctic tropopause bias. This tells us that as the ensemble reaches 5-7 days,

the model will become more strongly weighted relative to observations. As the ensemble

reaches this point the observations begin to make less extreme adjustments to the ensemble.

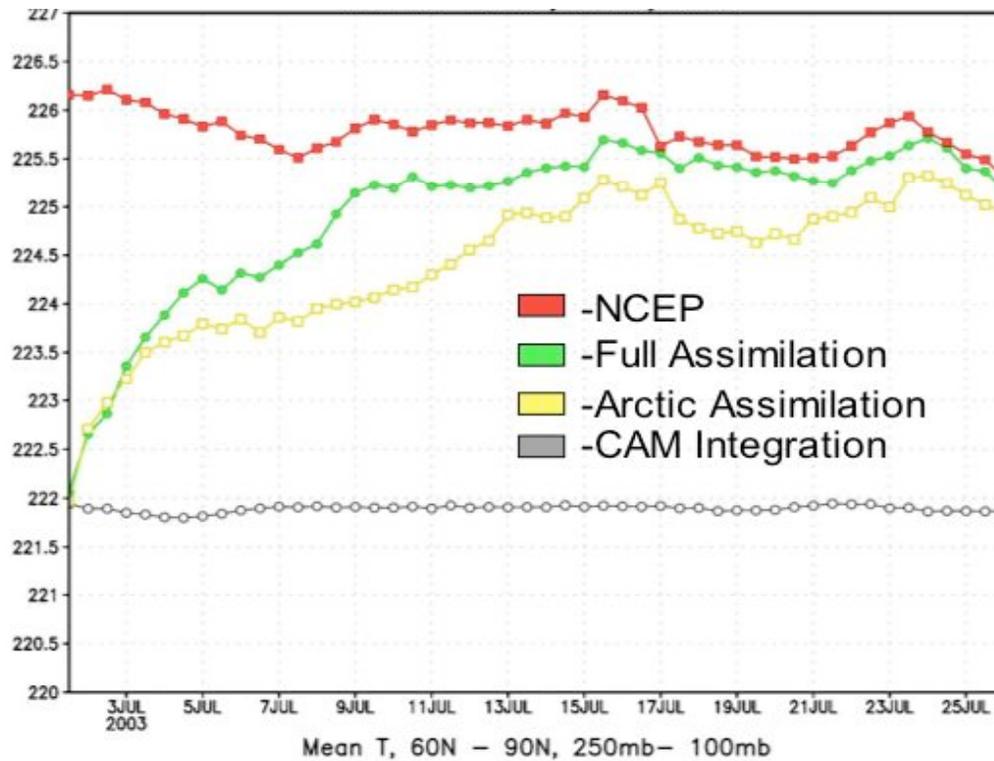


Figure 11: July 2003 mean temperature (K) averaged from 60°N-90°N and 250mb-100mb for NCEP reanalysis (red), full assimilation (green), Arctic assimilation (yellow), and CAM run (gray).

Although the ensemble spread partially determines the impact an observation may have on an ensemble, it does not necessarily describe how long it takes for a bias to be reduced by data assimilation. Figure 10 shows the area average over the Arctic of the RMSE of temperature with respect to NCEP reanalysis for each of the experiments. For the CAM run there is a clear bias throughout the month near the tropopause and throughout the troposphere. At all levels, the RMSE for temperature is consistent through time. This is as expected since no observations are being assimilated to reduce the error. In the case of the Arctic tropopause assimilation, we see a large initial bias at the tropopause level, and smaller biases lower in the atmosphere. However, as

observations are assimilated, the bias near the tropopause is slowly reduced to a smaller

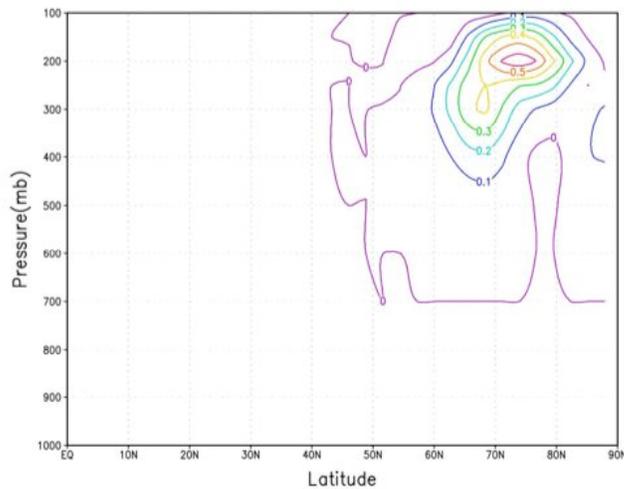


Figure 12: Zonal mean temperature analysis increment (K) for Arctic tropopause assimilation, averaged from July 13 - July 30, 2003.

consistent value. By inspection, it appears that the transition to a new stable state takes approximately 11-13 days, and the biases change minimally lower in the atmosphere. However, the RMSE in the lower troposphere in the Arctic tropopause assimilation is somewhat reduced relative to the CAM run even though no observations are being

assimilated in the region. Finally, in the full assimilation throughout the troposphere, the RMSE of temperature is strongly reduced, and the tropopause bias again takes about 9-11 days to reach a new stable state.

Figure 11 gives the averaged temperature from 66-90N and 250-100mb. From this figure it is clear that both the full assimilation and Arctic tropopause assimilation reduce CAM's cold bias, and nearly match NCEP reanalysis in this region. As expected, the full assimilation does slightly better at reducing the bias, but the Arctic tropopause assimilation is quite effective as well. Further, using Figure 10 the adjustment time for the full and Arctic tropopause assimilations can be more precisely determined. It appears that the full assimilation takes approximately 9 days to adjust, while the Arctic tropopause assimilation takes about 13 days. This gives an estimate for the initial time of a time series where both the Arctic tropopause bias and ensemble spread of our CAM

ensemble members has been reduced and are relatively stable. As a result, the remainder of this section will use data exclusively from days 13-30 to investigate the impact of a reduced CAM Arctic tropopause bias in order to minimize transient effects.

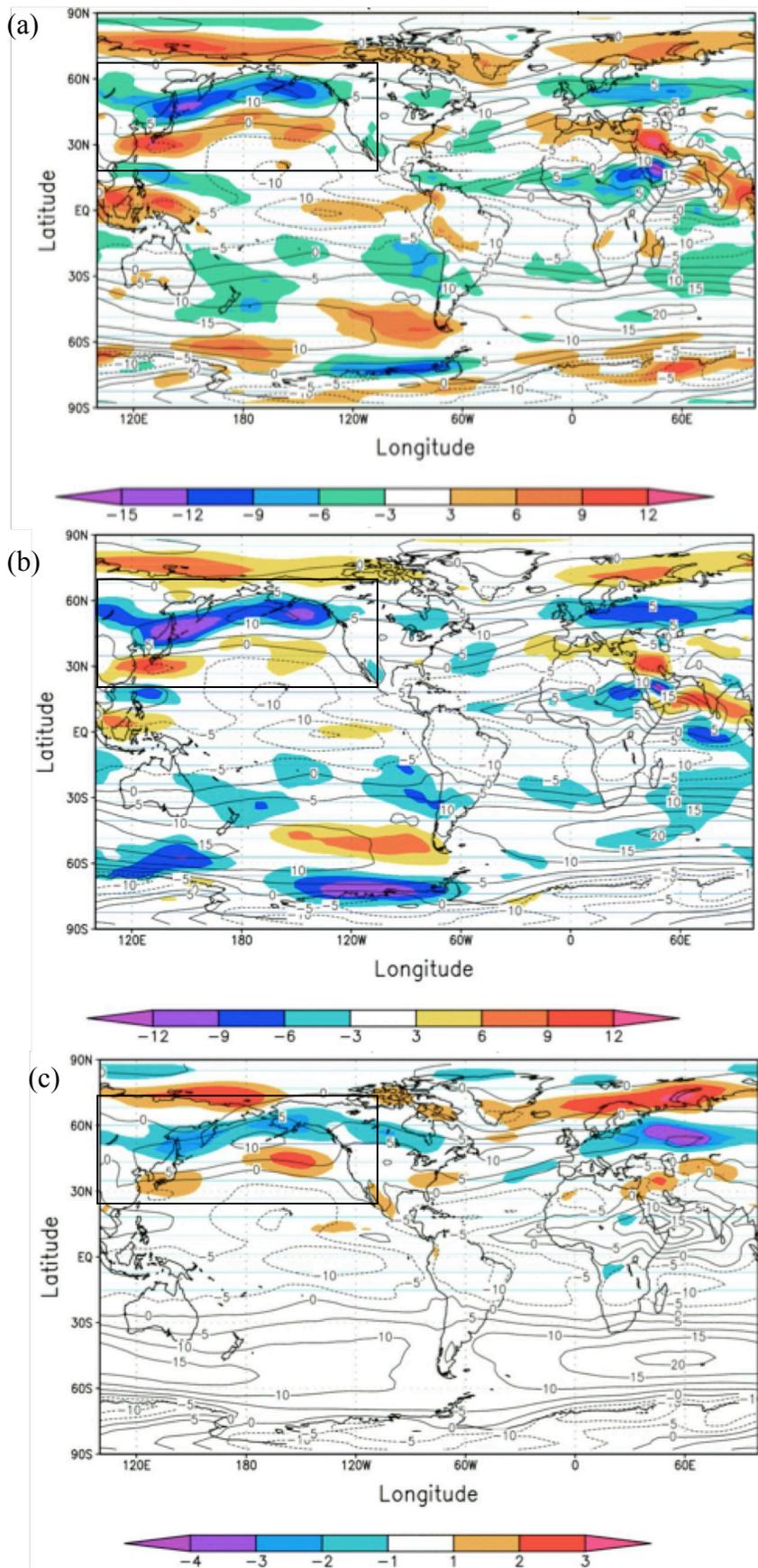


Figure 13: 850 mb zonal wind (m/s) for CAM run (contours) differences between CAM and (a) NCEP reanalysis, (b) full assimilation, and (c) arctic assimilation (shaded). The boxes highlight the region in which the storm track shifts are most clear. Averaged from July 13 – July 30, 2003.

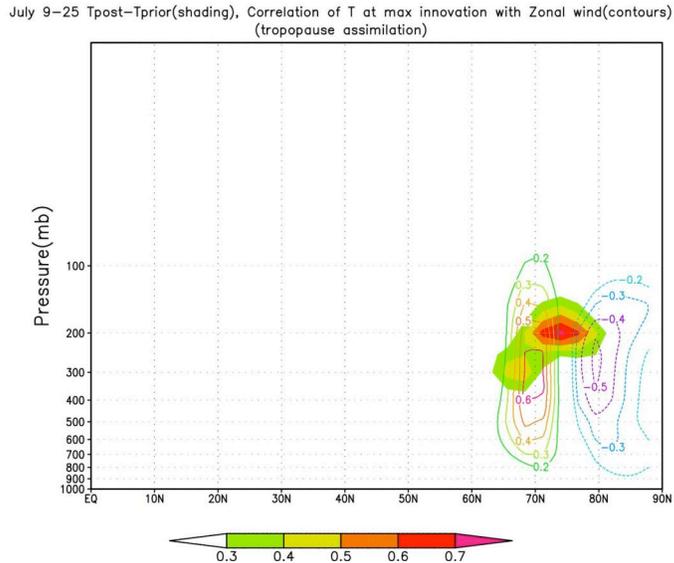


Figure 14: Zonal mean temperature analysis increment in degrees Kelvin (shaded), and zonal mean of correlation of U with temperature at location of maximum analysis increment.

When observations are assimilated into the Arctic tropopause region every six hours they effectively warm the region. Figure 12 shows the zonal average analysis increment for temperature in the Arctic tropopause run. It appears that on average, the assimilation of observations into the Arctic tropopause warms the region

by about $.5^{\circ}\text{K}$. While the analysis increment occurs instantaneously in state space, this roughly converts to a heating rate as

$$\frac{.5^{\circ}\text{K}}{\text{assimilation}} \approx \frac{.5^{\circ}\text{K}}{6\text{hours}} = 2^{\circ}\text{K}/\text{day}. \quad (3.2)$$

The application of this warming rate to the Arctic tropopause acts to keep the height of the polar tropopause lower than CAM's climatological location. Therefore, the overall impact of our Arctic tropopause assimilation is the opposite of what Lorenz and DeWeaver (2007) describe as occurring in the IPCC climate change scenarios where the tropopause rises with increasing greenhouse gas concentrations. This indicates that if the polar tropopause height and location of the zonal jet and surface westerlies are highly correlated as detailed by Kushner (2001) and Yin (2005), we would expect the resulting

dynamical adjustment of the zonal winds and low-level westerlies for the Arctic tropopause assimilation to be roughly the opposite to that of Figures 6 and 7.

We will first describe the impact of data assimilation on low-level westerlies. Storm track activity tends to be at a minimum during the Boreal Summer months, but Lorenz and Deweaver (2007) found that over the oceanic storm tracks, a signal of a shift could be found in the 850mb winds. Here we will focus on the Pacific storm track. During July 2003, the CAM run appears to place the 850mb zonal winds too far to the North with respect to NCEP reanalysis (boxed region of Figure 13a). When we assimilate data at all levels globally in our full assimilation we find that the storm track region as represented by 850mb winds shifts equatorward and closely resemble NCEP reanalysis (Figure 13b). This is to be expected since we are assimilating temperature and wind observations directly in the storm track region. Figure 13c shows the difference between the Arctic tropopause assimilation and the CAM run. In this case we see a smaller

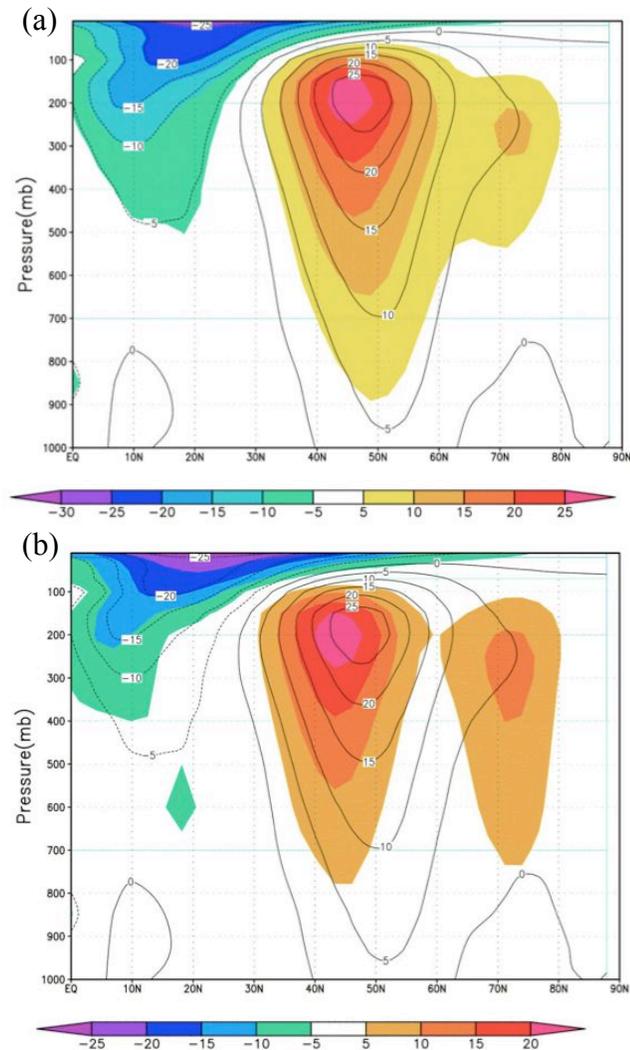


Figure 15: Northern Hemisphere CAM zonal wind (m/s) for July 13- July 30, 2003 (contours), (a) Arctic tropopause run zonal wind (shaded), and (b) full assimilation zonal wind (shaded).

equatorward shift than the full assimilation, but it is still present. This is somewhat surprising since the only place we are introducing information into the model is North of 66° and between 300-100mb, well away from the 850mb winds in this region. Additionally, for the Arctic tropopause assimilation, we see the difference between it and the CAM run goes to zero as the distance from the Arctic increases. Finally, comparing Figure 13 to Figure 7, we see that the pattern of the Pacific storm track response for the Arctic tropopause assimilation appears to be nearly opposite of what Lorenz and Deweaver found from the A2 IPCC climate scenario. This result seems to indicate that by assimilating data into the Arctic tropopause, we directly lower the height of the tropopause in the Arctic, and through a dynamical adjustment process shift the storm track region equatorward in the Pacific. However, there are features poleward of the stormtrack region (North of boxed areas in Figure 13) that indicate this may be a tripole or quadrupole response. To separate out these responses, figure 14 shows the correlation of temperature at the location of the maximum analysis increment with zonal wind across the ensemble for the Arctic tropopause assimilation. The correlation structures are located directly on the latitudes of the features poleward of the stormtrack. This suggests that these poleward features are being impacted directly by the assimilation of temperature near the tropopause, while the zonal wind shifts outside of this region, such as those in the storm track region are a result of secondary circulation effects possibly associated with tropopause adjustments.

When we investigate the impacts assimilating data near the Arctic tropopause have on the zonal jet, we find similar results as those on storm track location. In Figure 15a we compare the Northern Hemisphere mean zonal wind for that CAM run and for the

Arctic tropopause

assimilation. We see that the Arctic tropopause assimilation does appear to shift the jet equatorward. Also, the Arctic tropopause assimilation generates a secondary maxima in the zonal jet near the pole.

This feature does not exist in the CAM run. When we look at the full run results (Figure 15b),

we see a more distinct equatorward shift in the zonal jet and the secondary maxima.

Figure 16 shows the average analysis increment for zonal wind in the Arctic tropopause assimilation. By comparing this to Figure 15, we see that the secondary maxima feature is directly influenced by the assimilation of observations in the Arctic tropopause region. In addition, we also find that the shift in the zonal jet is due to secondary dynamical effects such as wave-mean flow interaction since the jet region lies well outside of the region where the assimilation directly impacts zonal wind. Finally, Figure 17 shows the difference between the Arctic tropopause assimilation and the CAM run. This figure clearly shows the equatorward shift in the zonal jet. By comparing Figure 17 with Figure 6, we see that our results neatly mirror the structure of those obtained by Lorenz and DeWeaver (2007) using a simple GCM with a raised Arctic tropopause. The main

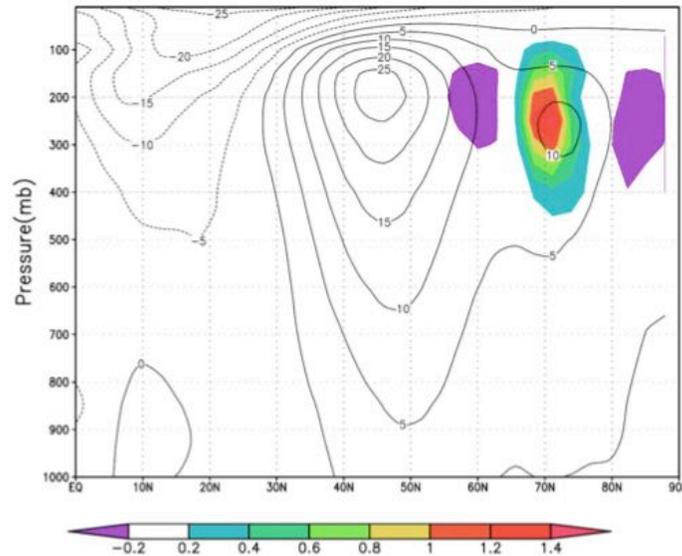
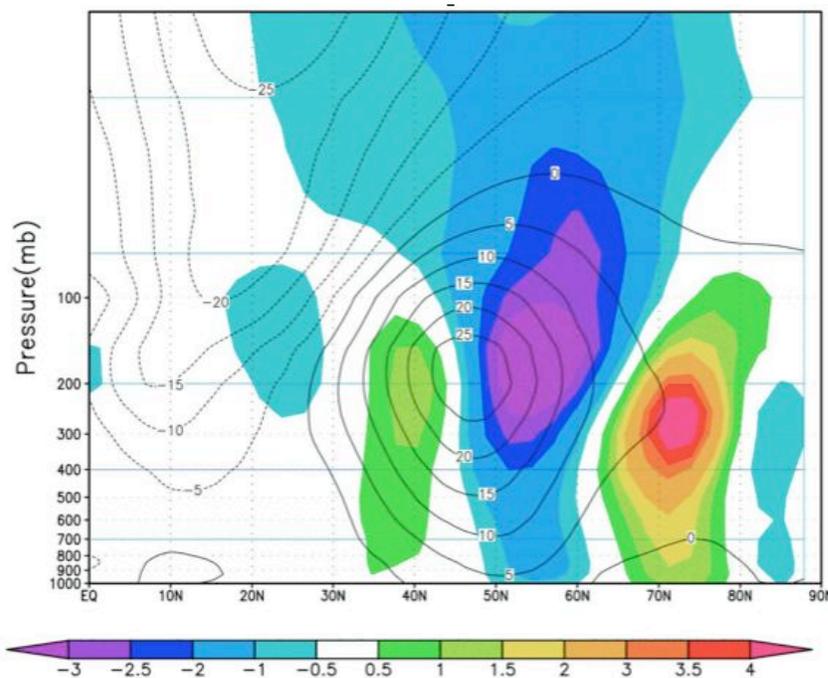


Figure 16: Mean zonal wind (m/s) for the Arctic tropopause assimilation (contours), and average analysis increment in zonal wind for the Arctic tropopause assimilation (shaded) for July 13 – July 30, 2003.

difference, is the sign change due to their experiment raising the Arctic tropopause, while our experiment lowered it.

d. Conclusions on the impacts of Arctic model bias

By inspecting the zonal mean temperature difference between CAM and NCEP reanalysis we identified a cold polar tropopause bias in CAM. By selectively assimilating only observations poleward of 66N, and between 300 and 150mb, we were able to reduce this bias significantly. In reducing this bias we effectively lowered the height of the Arctic tropopause. Through evaluation of statistical relationships found in the ensemble, we found the direct impact of assimilating polar tropopause observations to be localized near the region of the observations. However, the resulting zonal mean circulation throughout the Northern Hemisphere, was influenced by the assimilation through secondary effects. The consequences of this influence included equatorward



shifts in the extratropical storm tracks and midlatitude jet, as well as reduced model bias in the zonal mean wind field throughout the hemisphere. These results are broadly consistent with the

Figure 17: CAM run zonal mean wind (m/s) (contours) and difference between Arctic tropopause assimilation zonal mean wind and CAM zonal mean wind (shaded) for July 13 – July 30, 2003.

association between jet location and tropopause height found in the IPCC simulations and simple GCM experiments.

One important distinction between our data assimilation experiment and simple GCM experiments is that in the simple model the only forcing was tropopause temperature. In the assimilation experiment, T is not the only observation assimilated. Observations of U, V, and other variables are also included. So, we are not only specifying the temperature profile, we are also specifying the momentum profile. By doing this, we are essentially prescribing the tropopause wave structure in the Polar region. Therefore, the dynamical relationships that cause the shifts in the simple GCM or IPCC runs may not correlate with those resulting from assimilating data in a constrained region. More work is needed to determine the mechanisms by which the zonal winds shift. In addition, our work is limited by length of our time series. Longer assimilations covering multiple years are necessary to remove some of the uncertainty in determining what proportion of our results may be due to transient weather conditions that happen to exist in the month of our experiment.

In this experiment we have not only given some validation to an existing theory of how zonal winds shift in response to changing climatic variables, but have also demonstrated how data assimilation can be used to begin to tease out the dynamical impacts of model biases. By assimilating data into a restricted region of a known model bias we were able to demonstrate that the impacts of reducing this bias reached far beyond the local area. While doubtlessly there are simpler, less costly ways to achieve similar results for bias correction, this work is really a first step in using the power of data assimilation to understand the dynamical impacts of model biases. In the future some of

the more unique and powerful techniques associated with data assimilation such as sensitivity analysis and parameter estimation can be integrated to give us information that may not be easily obtainable with other methods.

4. Part 2- Assimilating GPS Occultation Observations in CAM

a. Introduction to GPS occultation applications to climate research

As discussed in the previous section, the Community Atmosphere Model (CAM) has several biases that have important dynamical implications for research. Some examples of these biases include a cold temperature bias near the tropopause at high latitudes, systematic biases in the mean zonal wind, and transient momentum fluxes that are too strong. One method for investigating the impacts that these biases have is to use data assimilation techniques to reduce the bias and observe how the system reacts. There

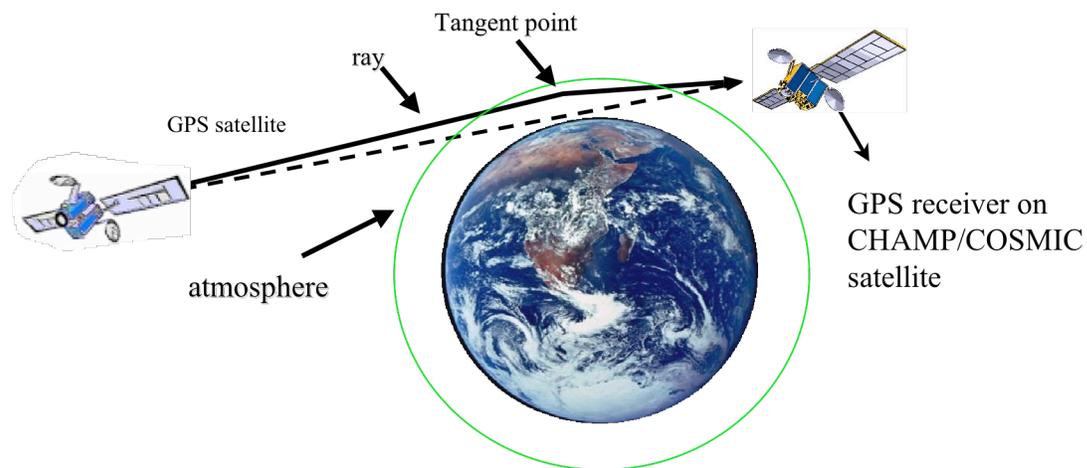


Figure 18: Schematic illustrating basic concept of GPS radio occultation observations.

are also applications for reanalysis and weather prediction systems. In fact these that are likely to be the prime users of these sets. The far-reaching applications of these data sets make understanding their properties and defining their limits essential.

There are several observation types that have a global nature such as existing satellites and, to a certain extent the radiosonde network, but each has its limitations. Over the past decade, the German Challenging Minisatellite Payload (CHAMP) and the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) missions have made GPS occultation observations available to add to the set of global observations available. This chapter will make a preliminary inspection of how the assimilation of GPS occultation measurements can constrain and impact the CAM state vector in various regions of the atmosphere, and also what the limits of assimilating these observations in CAM are.

GPS occultation observations sample the atmosphere by using low Earth-orbiting

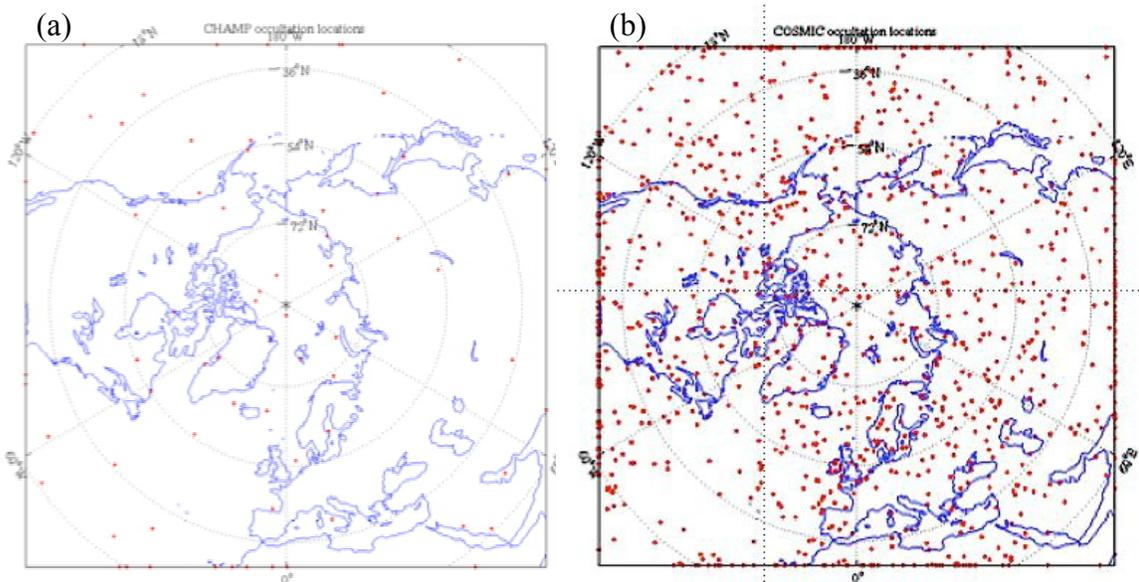


Figure 19: Red dots indicate Northern Hemisphere GPS radio occultation observation locations for (a) January 2003 and (b) January 2007.

CHAMP/COSMIC satellites to collect radio waves emitted by GPS satellites (Ware et al. 1996). As a GPS satellite rises or sets relative to a receiver, the delay of the radio wave due to the atmosphere is measured by a receiver satellite (Kursinski et al. 1997). The delay of the radio waves is caused by the refraction of the waves as the ray moves through density differences in the Earth's atmosphere as shown schematically in Figure 18. Using the Abel inversion and assuming local spherical symmetry, the refractivity and bending angle can be retrieved as described in Kuo et al. (2004). By using an assumption of local refractivity at the tangent point of the ray as shown in Figure 18, the refractivity of the atmosphere can be related to the specific humidity q , temperature T , and pressure p . This assumption generally provides good results, however large local density differences in the atmosphere can cause errors. These large local density differences generally occur in the lower troposphere due to the possible presence of hurricanes, strong convection, or atmospheric fronts (Sokolovskiy et al. 2005). There are methods that use nonlocal assumptions and take advantage of additional information from the ray path of the radio wave (Healy et al. 2006). However, these methods increase the computational cost of assimilation considerably.

Another difficulty in using GPS occultation measurements is that separating the influence of q , T , and p on refractivity can be impossible without prior estimates of the various quantities (Liu et al. 2007). However, assimilating these observations with CAM or any other atmospheric model alleviates this problem by providing an estimate of the system state prior to assimilation from short-term model forecasts as described in Chapter 2. Once we have our prior estimate of the system in state space we interpolate to observation space and estimate local refractivity using a forward operator given by

$$N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{P_w}{T^2}, \quad (4.1)$$

where N is the index of refraction, p is pressure in hectopascals, T is temperature in Kelvins, and p_w is the partial pressure of water vapor in hectopascals (Kursinski 1997). Using equation 2.26 we can then update the ensemble in observation space, and then update the ensemble in state space using equation's. 2.25, 2.27, and 2.28.

The strong connection between refractivity and T , p , and q as given in equation 4.1 implies that as refractivity observations are assimilated, updates for the above components of the state vector will occur as described in equation 2.23. However, the magnitude of the updates will vary by type and region, and depend on what atmospheric variables are controlling refractivity in a region. An important question to ask is how much of the information provided in a refractivity observation is associated with T , p , and q in various regions? Not only does the answer strongly constrain how refractivity observations will influence the state vector components in various regions, it may also have implications for the usefulness of GPS observations in many of its possible applications.

In addition to being a novel and useful application of the multibillion-dollar GPS system, the GPS radio occultation measurements have many desirable qualities. The most obvious of these properties include their global nature. As more satellites continue to be launched, the coverage of these occultation measurements has increased considerably. Also, the density of observations increases toward the poles where conventional observations can be scarce. Figure 19(a) gives the locations of GPS radio occultation measurements in the Northern Hemisphere for January 1, 2003, and Figure 19(b) shows the locations four years later. One drawback of radiosonde observations are

their reduced accuracy in Polar regions. Conversely, GPS occultations provide radiosonde-like measurements with increased accuracy in Polar regions where the ambiguity between T, Q, and p is reduced (Kursinski 1997). Another advantage is that the vertical resolution of GPS measurements is between a few hundred meters and one kilometer, which is high compared to similar satellite observations. However, the local assumption tends to smear the horizontal resolution, which causes each observation to effectively be an average measurement over several hundred kilometers (Ware et al. 1996). On the other hand, when assimilating these observations into a climate model such as CAM where the horizontal resolution is on the same order as the observation, this horizontal averaging property of GPS occultation measurements can be advantageous. Furthermore, the instruments that provide the occultation measurements are inherently stable. Once a satellite is in orbit it requires no additional recalibrations or adjustments (Ware et al. 1996). This property makes radio occultation observations particularly important to climate studies where problems associated with recalibrating or adjusting instruments are common over long time series. Finally, GPS occultation measurements are insensitive to clouds and precipitation.

The goals of this chapter are twofold. First, we begin by determining how strongly these observations can constrain CAM to match the actual atmospheric state. In order to do this the assimilation must overcome the climatological biases that exist in CAM, as we saw in chapter 3. Therefore by observing how strongly the assimilation of GPS observations reduces well-known biases in CAM we can infer the power of the assimilation to constrain the model. From equation 4.1 water vapor is directly connected to refractivity. Thus, the assimilation of a refractivity observation should make a strong

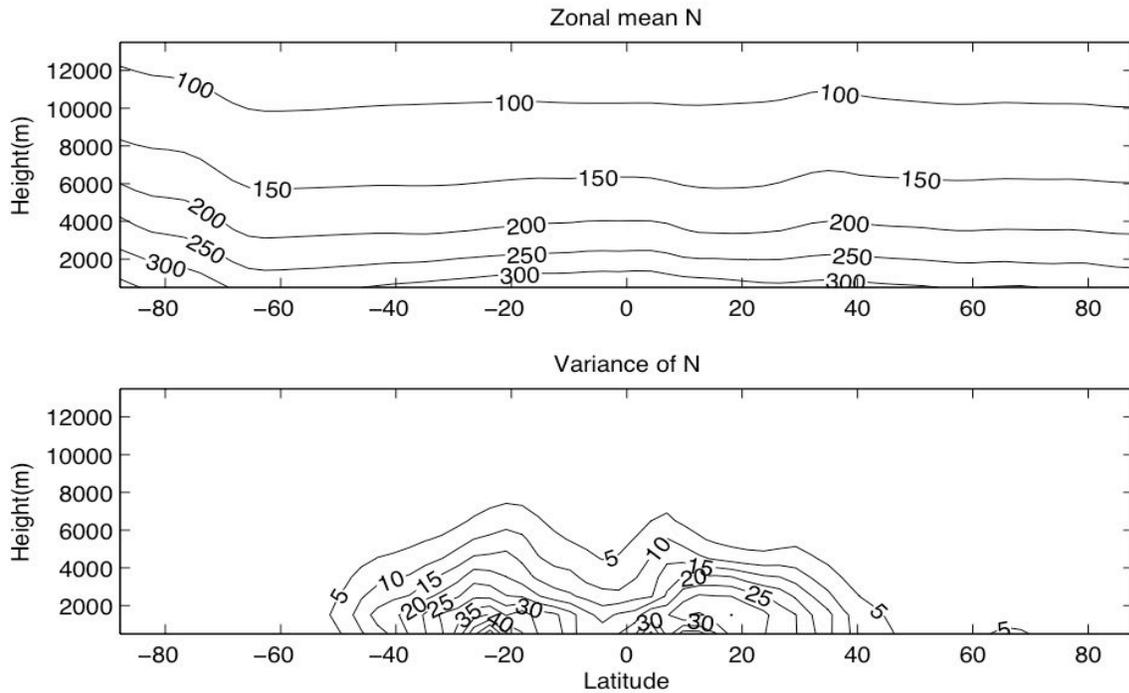


Figure 20: Zonal mean refractivity (a) and variance of refractivity (b).

constraint on water vapor in CAM. However, water vapor is known to be a difficult quantity to model, so Section c of this chapter will inspect the impact of refractivity observations on CAM's representation of water vapor closely. Second, we look closely at the structure of the GPS forward operator given by equation 4.1. By linearizing equation 4.1 we can investigate the relationship between refractivity and temperature, pressure, and specific humidity. This allows us to describe how GPS occultation measurements impact various regions of the atmosphere differently due to distinct covariance structures. Through data assimilation and understanding the process by which such assimilated observations impact CAM, we can begin to predict what power they will have in constraining different fields of our state vector in various regions of the atmosphere. In particular, there are interesting relationships between how the two terms of equation 4.1 interact. We will see that the first term dominates the mean state of

refractivity throughout the atmosphere as well its variability in the upper troposphere of the extratropics, while the second term dominates the variability in the tropics and lower atmosphere. Figures 20 and 21 show these relationships. The implications of this will be presented in Section d by considering the correlation and covariance structures of components of the state vector with refractivity in a linearized setting.

Section b of this chapter describes the GPS occultation observations and methods used in this experiment. Section c provides results describing the capability of assimilated GPS observations to constrain the state vector by overcoming CAM model bias, and represent the real atmosphere. This section will emphasize the degree to which assimilated GPS observations can impact CAM's water vapor representation. Section d is described above. Finally, section e, offers the conclusions for this experiment.

b. Description of GPS observations and experimental setup

Similar to the Arctic tropopause bias experiment of Chapter 3, this study consists of three components. We use CAM3 at T42 resolution with Eulerian dynamics as our AGCM forward model. We assimilate the same types of physical observations described in Chapter 3, as well as additional COSMIC and CHAMP GPS radio occultation observations described in more detail below. However, in this chapter we will be using observations from January 1-30, 2007. This time period was chosen due to the recent availability of COSMIC observations. Finally, the assimilation framework used is the EAKF provided by DART. We use a 40-member ensemble initialized from a climatological spread as described in Chapter 3. Observations are assimilated in 6 hour increments.

The raw CHAMP/COSMIC datasets used in this research are freely available with registration at the website of the Taiwan Analysis Center for COSMIC (<http://tacc.cwb.gov.tw>). Specifically, we use thinned level-2 refractivity atmospheric profiles. We extract the profiles from 500m-20,000m and then vertically bin and average the high resolution profiles into 22 levels spaced every 700m below 10000m and 1200m above 10000m. These observations are then assimilated up to 100mb.

In order to test the ability of assimilated GPS observations to reduce CAM model bias we completed four runs at the NERSC facilities described in Chapter 3 for the time period of January 1, 2007 to January 30, 2007. They are as follows:

- GPS assimilation: Assimilated only GPS occultation measurements globally.
- GPS+NCEP assimilation: Assimilated GPS observations and a subset of the observations used as input to NCEP reanalysis (ACARS measurements, radiosondes, satellite winds, surface pressure measurements).
- NCEP assimilation: Assimilated only NCEP input observations as described above.
- CAM control run: Assimilated no data into the model.

In addition, we use NCEP reanalysis from January 2007 for comparison purposes. The results from these runs are used to compare their representations of common forecast metrics such as 500mb geopotential height, as well as known CAM model biases. In particular, we will analyze biases in the Polar tropopause temperature, zonal wind, and transient momentum fluxes.

We also take a detailed look into the mechanisms by which GPS observations impact the state of CAM's modeled atmosphere. By linearizing the refractivity formula given by equation 4.1, we can use multivariate linear regression to piece apart the correlation structures that determine how the assimilation of a GPS observation at a given location will influence the state of our system. We will use the GPS assimilation exclusively in this process of exploring the correlation structure of refractivity with state variables.

c. GPS observations and CAM model bias results

In this section the ability of GPS observations to constrain CAM and overcome model biases is examined. We begin by looking at how well GPS observations can constrain CAM to simulate and forecast 500mb geopotential height. We then take a detailed look at how well the assimilation of GPS observations assimilation can reduce CAM biases in transient momentum flux, zonal wind, and Polar tropopause temperature.

In order to define how well an analysis matches the actual atmosphere one compares the analysis with the true state of the atmosphere. One difficulty in this process is determining what is the "truth". In many cases NCEP reanalysis is accepted as the truth. However, NCEP reanalysis has known biases of its own and is an imperfect product (Kalnay et al. 1999). Moreover, the uncertainties associated with these biases are not well defined. In this experiment we are essentially creating our own reanalysis product, albeit with limited observations for January 2007. Even though our assimilation and subsequent analysis uses a lower resolution model, fewer observations, and a less well tested assimilation system, it is still not unreasonable to hypothesize that in some

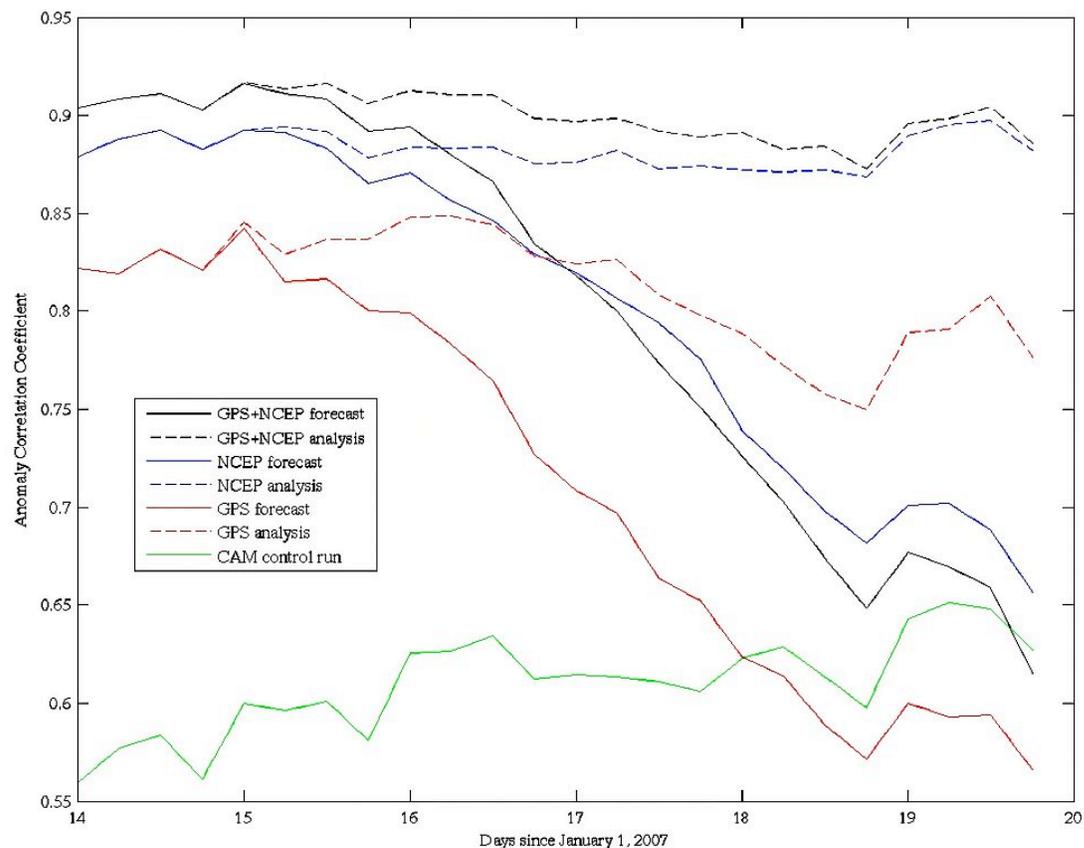


Figure 22: Anomaly Correlation Coefficient for global 500mb geopotential height from January 14 – January 20, 2007. Dashed lines indicate analysis, and solid represents forecasts. Results are given for GPS+NCEP assimilation (black), NCEP assimilation (blue), GPS assimilation (red), and the CAM control run (green).

regions our system may better represent portions of the actual atmosphere on a given day than NCEP reanalysis. With that caveat in mind, we will assume NCEP reanalysis to represent the true state of the atmosphere. It will be with respect to NCEP reanalysis that we will define the skill with which our assimilations represent the state of the atmosphere by constraining CAM and overcoming model bias.

One common way to define the skill with which an analysis matches the actual atmosphere on a given day is to compare the analysis 500mb geopotential height with the true 500mb geopotential height. The metric often used to accomplish this is the anomaly

correlation coefficient (ACC). There are several variations on this formula, one that is commonly used for ECMWF forecasts, and we will use here is:

$$ACC_{x,\hat{x}} = \frac{E[(\mathbf{X} - \mathbf{C})(\hat{\mathbf{X}} - \mathbf{C})]}{\sqrt{\{E[(\mathbf{X} - \mathbf{C})^2]E[(\hat{\mathbf{X}} - \mathbf{C})^2]\}}} \text{ (Persson and Grazzini 2005). (4.2)}$$

Usually, this equation is used in with \mathbf{X} representing a forecast and $\hat{\mathbf{X}}$ an analysis.

However we will use this equation with \mathbf{X} representing an assimilation run, $\hat{\mathbf{X}}$ being the “truth” run/NCEP reanalysis, and \mathbf{C} is the climatological 500mb geopotential height field for January. This formula gives us a measure of how closely our assimilation’s deviations from climatological 500mb geopotential height correlate with those of NCEP reanalysis. In applying this formula we also initialized a 6-day forecast for each of our runs on January 14, 2007. In these cases \mathbf{X} becomes the forecast run, and $\hat{\mathbf{X}}$ remains the NCEP reanalysis field. The ACC’s from January 14, 2007 through January 20, 2007 for the various month-long runs and their associated forecast are displayed in Figure 22.

There are several important points to emphasize from Figure 22. First, unlike classical ACC plots our plot does not smoothly depart from a value of close to one. Instead our various runs start anywhere from approximately .55-.9, and the forecast runs are not smooth as they deviate from their initial values. This is an artifact of using NCEP reanalysis as our “truth.” In classical applications of the ACC the truth is considered to be an analysis run using all available information, and the forecasts are generated using the same model, analysis system, and initial conditions with slightly different sets of observations which cause the forecasts to deviate only slightly from each other. In our case, we are using massively different sets of observations to initialize each run, and our forecast model is not the same as that used for NCEP reanalysis. This combination of

factors causes the ACC for each of our runs to be clearly distinct, and most likely decorrelate more quickly than expected. However, one advantage of calculating the ACC using this method is that it does give us some measure of the skill of our analyses/forecast relative to the accepted values provided by NCEP reanalysis. Starting with the CAM control run we see that it has an initial ACC of about .55, and stays relatively constant although it does increase slightly over the forecast time. This compares fairly well with .6 being the approximate accepted value of a forecast with no useful forecast skill. It can be shown mathematically that .5 is the ACC for a model state with no correlation to the analysis (Persson and Grazzini 2005). Figure 22 also gives a reference baseline to determine when the forecasts for the other runs have lost their skill. By inspection of Figure 22, this appears to occur around 4 days.

Another important point to take away from Figure 22 is a comparison of the analysis runs relative to NCEP reanalysis. Since the GPS analysis starts this period with an ACC of about .83, it appears that GPS observations can reduce the error of CAM's representation of 500mb geopotential height. However, relatively more information is gained from assimilating the observations that make up the NCEP analysis (\sim .88 initial ACC). Finally, the combination of assimilating both the GPS observations and NCEP observations forces CAM to give the best representation of 500mb geopotential height relative to NCEP reanalysis (GPS+NCEP analysis \sim .91 initial ACC).

While a general skill score as given in Figure 22 is useful for determining the power of our system to reproduce the daily variations that make up the weather, it could be argued that understanding a model's long-term biases is more important for studying the dynamics of climate. As we saw in Chapter 3, data assimilation systems can be a

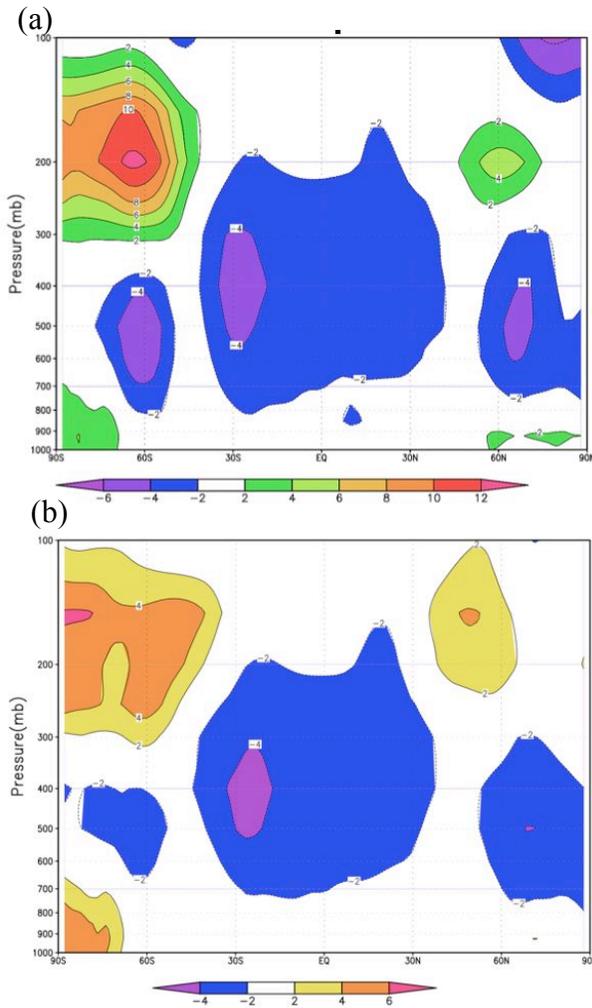


Figure 23: Difference between NCEP reanalysis zonal mean temperature (K) and (a) CAM run zonal mean temperature, (b). GPS assimilation zonal mean temperature averaged from January 10 - January 30, 2007.

this common bias shifts to the Southern Hemisphere's Polar region with similar implications (Covey et al. 2004). Figure 23b gives the difference between NCEP reanalysis and our GPS assimilation. From this figure it appears that the assimilation of GPS occultation measurements alone is capable of reducing the Polar tropopause bias, but not eliminating it. However, in regions where biases are not so clear, such as the tropics, the GPS assimilation does not correct temperature very strongly. The next

useful tool in understanding the impacts of climate model bias. With GPS radio occultation measurements being a new product that provides quality observations in previously poorly observed regions, such as much of the Southern Hemisphere and over oceanic regions (Kursinski 1997), it is important to determine how strongly their assimilation will interact with CAM's major biases.

Figure 23a shows the zonally averaged temperature bias for CAM with respect to NCEP reanalysis for the Southern

Hemisphere. In Chapter 3 we investigated a large temperature bias near the Arctic tropopause during the month of July. During the boreal winter months,

section will take a closer look at the varying regions where GPS occultation observations impact different components of the state vector.

We discussed in Chapter 3 that biases in temperature and zonal wind may have dynamical connections. Figure 24a gives difference between NCEP reanalysis and CAM's representation of the zonal wind for January 2007. Here we find a similar result to those of Chapter 3, with the zonal jet for the summer hemisphere being shifted toward the equator relative to NCEP

reanalysis as evident by the negative values between 45-30S, positive values between 60-45S and the jet being located near 45S. Figure 24b shows the difference between NCEP reanalysis zonal wind and the GPS assimilation's zonal wind. This shows that the GPS assimilation effectively eliminates CAM's errors in the troposphere below about 400mb.

Finally, it is also clear that as GPS observations are assimilated, the biases in the zonal winds are reduced. Unlike the temperature

field this adjustment appears to be more uniform throughout the troposphere, although

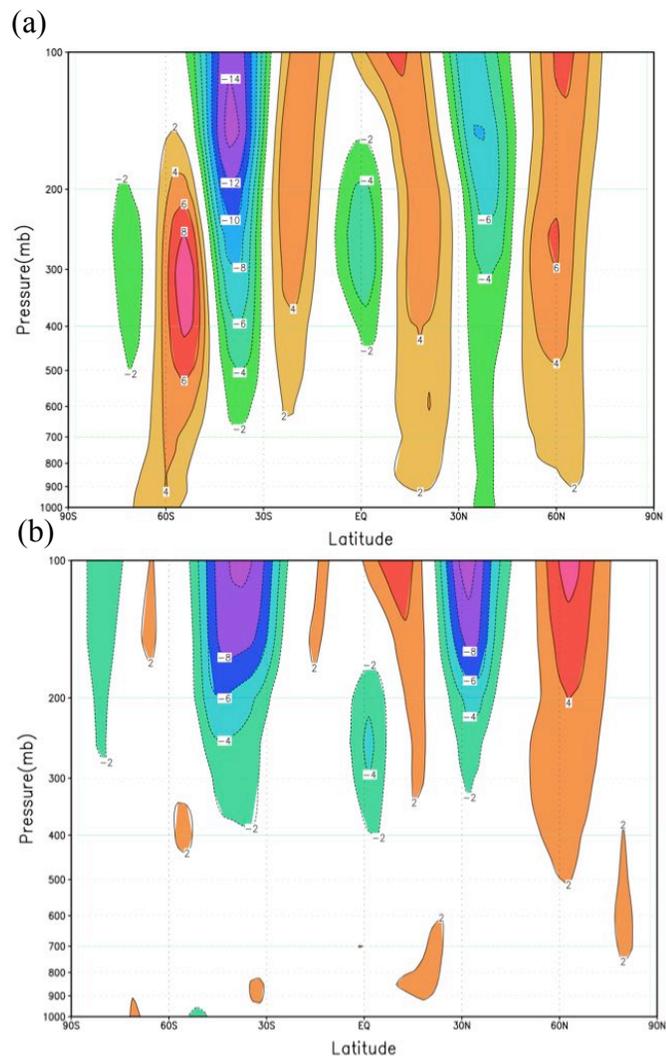


Figure 24: Difference between NCEP reanalysis zonal wind (m/s) and (a) CAM run zonal wind, (b) GPS assimilation zonal wind averaged from January 10 - January 30, 2007.

strong biases still exist at upper levels. However, This may be due to artificially limiting the assimilation and impact of observations to levels below 100mb.

By replacing partial pressure of water vapor with specific humidity equation 4.1 can be rewritten as

$$N = 77.6 \frac{p}{T} + 5.997 * 10^5 \frac{pq}{T^2}. \quad (4.3)$$

Where q is specific humidity in (kg/kg), T is temperature in kelvins, p is pressure in hectopascals, and N is the same refractivity index as equation 4.1. This form is useful because the variables in equation 4.3 are all easily accessible diagnostic variables in CAM. This form clearly demonstrates that variations of specific humidity across the ensemble will correlate with changes in refractivity through the second term of equation 4.3. This indicates from equation 2.23 that assimilating observations of refractivity should strongly influence and constrain CAM's representation of specific humidity. Figure 25a shows the zonal mean difference in specific humidity between CAM and NCEP. In the tropics above the boundary layer, it appears that CAM is too wet in the zonal mean, and this extends to a lesser extent out to the subtropics and midlatitudes. Figure 25b gives the difference between the GPS assimilation and NCEP reanalysis. Clearly, figures 25a and 25b appear very similar. Unfortunately this indicates that the assimilation is not constraining CAM's water vapor field very strongly. Finally, Figure 21c shows the mean analysis increment between the posterior and prior ensemble means. From this figure, the assimilation of GPS occultation measurements actually slightly moistens the tropics, and slightly dries the subtropics. This appears to be a strange result

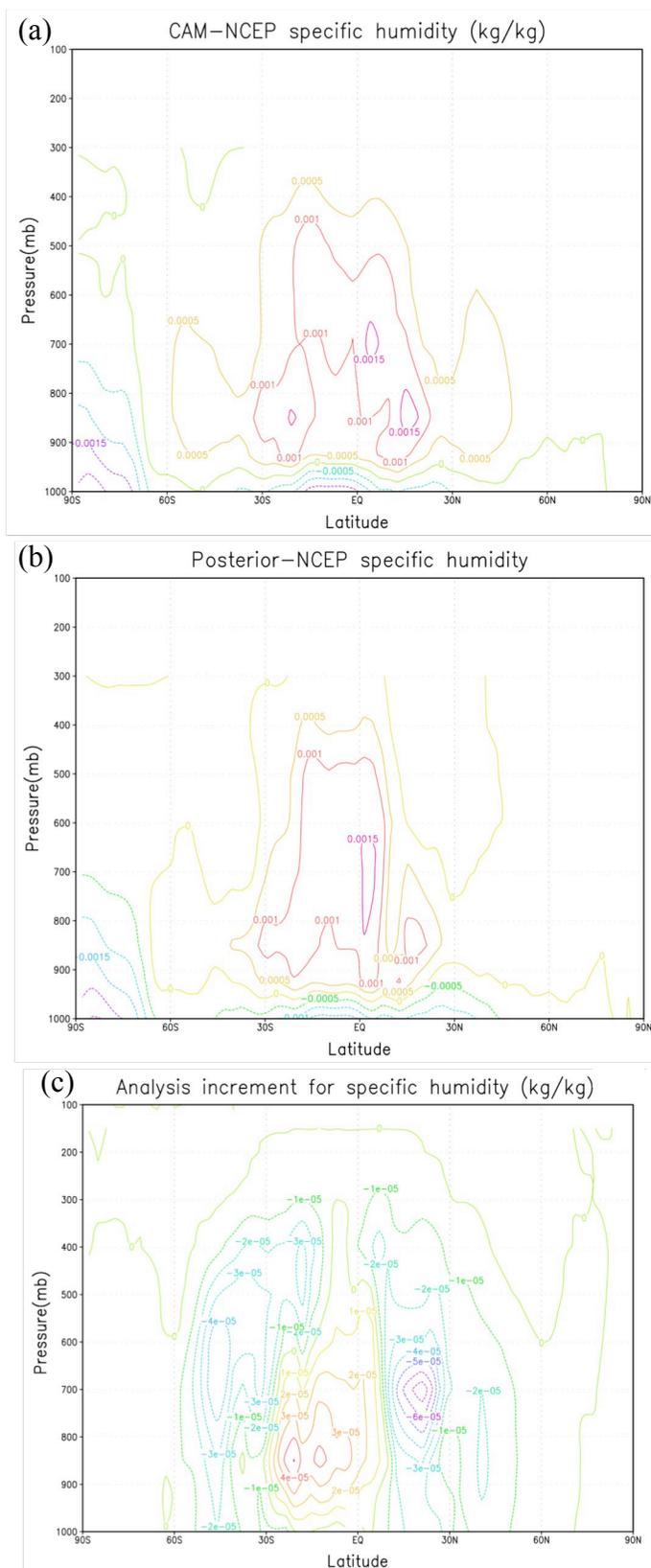


Figure 25: Zonal mean difference in specific humidity between CAM and NCEP (a), Posterior GPS assimilation and NCEP (b), and analysis increment (c) in kg/kg.

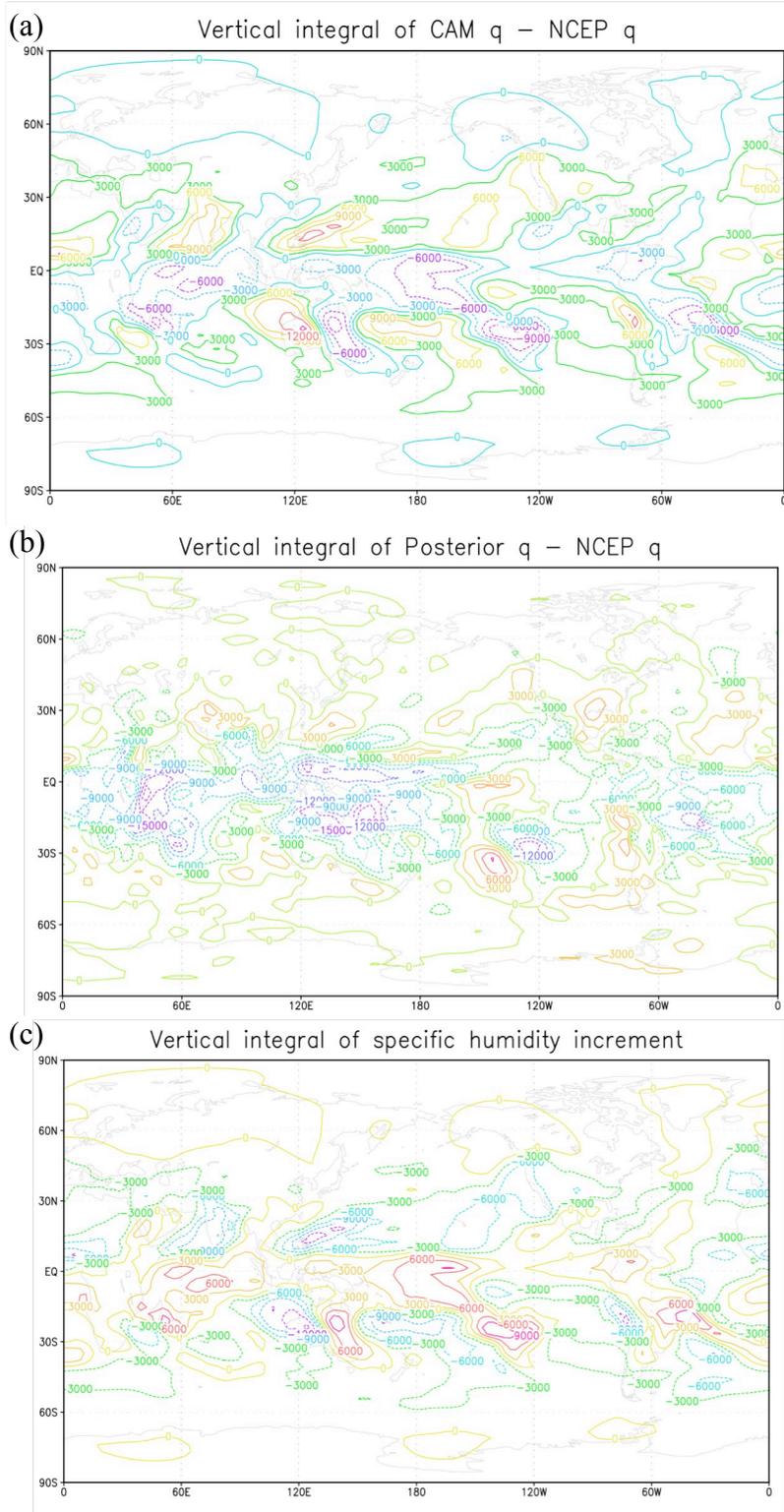


Figure 26: Difference in the vertical integral of specific humidity (kg/kg) between CAM and NCEP (a), Posterior GPS assimilation and NCEP(b), and analysis increment(c).

given figure 25a. Figure 26 shows the vertical integral of specific humidity or total precipitable water in the x-y plane. Figure 26c shows that CAM is integrating some information of specific humidity. However, the analysis increment is relatively small compared to the biases shown in Figure 26a. There are several questions that this result raises. The first most obvious question is why is CAM's water vapor field so weakly constrained by GPS observations? One possibility is that the assimilation is rejecting the specific humidity updates as being too far from the model state. However, more research is needed to understand this. The second question is what impact does the moisture that is assimilated have on the model state? Is the moisture simply precipitated out by the model physics, or does it remain in the modeled atmosphere? Again, additional research is needed to clarify these questions.

The final bias this research will inspect is CAM's meridional transient momentum flux. This measure has important climatic consequences for the location and strength of synoptic activity as well as the regional balances of energy and momentum. Figure 27a shows the transient meridional momentum flux ($[u'v']$) at 250mb for the CAM control run (black), GPS assimilation (yellow), and NCEP reanalysis (green). This product is formed taking the average value from January 10, 2007 – January 30, 2007 (The first 10 days were ignored to reduce errors due to assimilation adjustment), and subtracting the instantaneous value of U and V at each time step to form u' and v' . The product of u' and v' is then calculated at each time step, averaged in time, and finally averaged zonally. We find similar results to DeWeaver and Bitz (2005, their Figure 11). The CAM control run's transient meridional momentum flux is nearly twice that of NCEP reanalysis in the midlatitudes. Additionally, there are small biases in the latitude at which the sign of

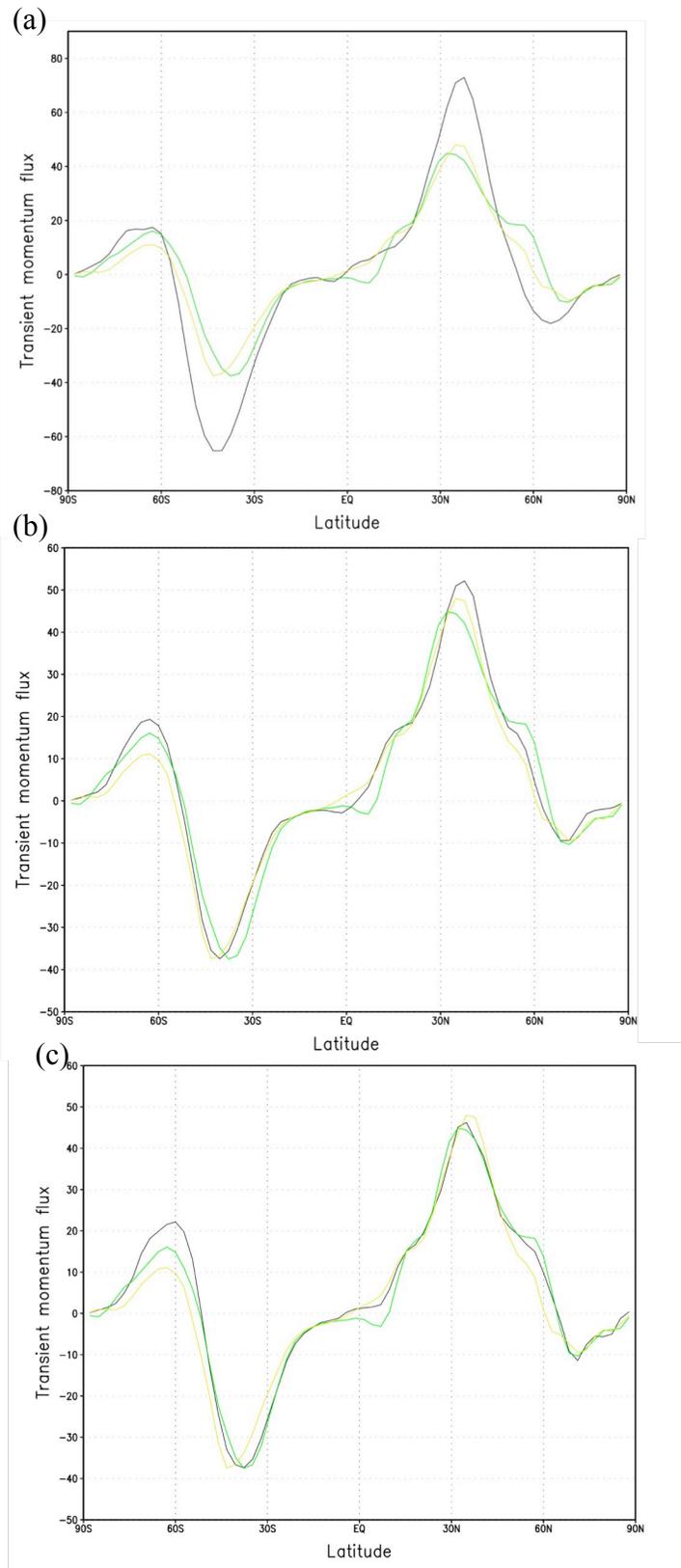


Figure 27: (a) shows the transient momentum fluxes (m^2/s^2) for the CAM run (black line), NCEP reanalysis (green line), and GPS assimilation (yellow line). (b) shows the transient momentum flux for NCEP reanalysis (green line) and the transient momentum flux for the GPS assimilation with the transient momentum flux calculated for individual ensemble members (black line) and from ensemble mean U and V (yellow line). (c) is the same as (b) but for the GPS+NCEP assimilation.

$[u'v']$ changes. The assimilation of GPS occultation observations effectively eliminates the transient momentum flux strength bias, and reduces biases in the latitudinal placement bias.

There is an additional question that is raised when using an ensemble analysis to determine transient quantities such as the momentum flux. In general, the ensemble mean is assumed to be the most likely state of the atmosphere at a given point in time for an ensemble analysis, as was done for the U and V fields used in constructing Figure 27a. However, the averaging process involves some amount of smoothing and reduces some of the sharpness in wave structures that contribute strongly to quantities such as the transient momentum flux. Therefore, is it more accurate to calculate the transient momentum flux for the individual members and take the ensemble mean of that quantity, as opposed to finding the ensemble mean zonal and meridional winds and determining the transient momentum flux from that mean? Figure 27b compares the results for the transient momentum fluxes obtained by each method for the GPS assimilation (Figure 27b), and the GPS+NCEP assimilation (Figure 27c). In each case the transient momentum fluxes were underestimated, particularly at high latitudes, by the fluxes when calculated using the ensemble mean U and V. The transient momentum fluxes increase in the extratropics and appear to improve overall when the transient momentum flux is obtained for each ensemble member individually and then averaged.

d. Covariance structure of linearized refractivity and the state vector results

When examining Figures 20-27 it is clear that the assimilation of GPS radio occultation measurements is capable of overcoming many of the strong biases that occur

in CAM. However, closer inspection reveals that the assimilation of these observations do not impact all regions of the atmosphere equally. For example, Figure 23 gave the differences in temperature between NCEP reanalysis and the GPS assimilation. The assimilation apparently greatly reduced the polar tropopause bias. However, biases in the mid-troposphere of the tropics were not impacted significantly. Similarly in Figure 27, the bias in transient meridional momentum flux was impacted significantly in the midlatitudes, but had a negligible impact on $[u'v']$ near the equator. This section

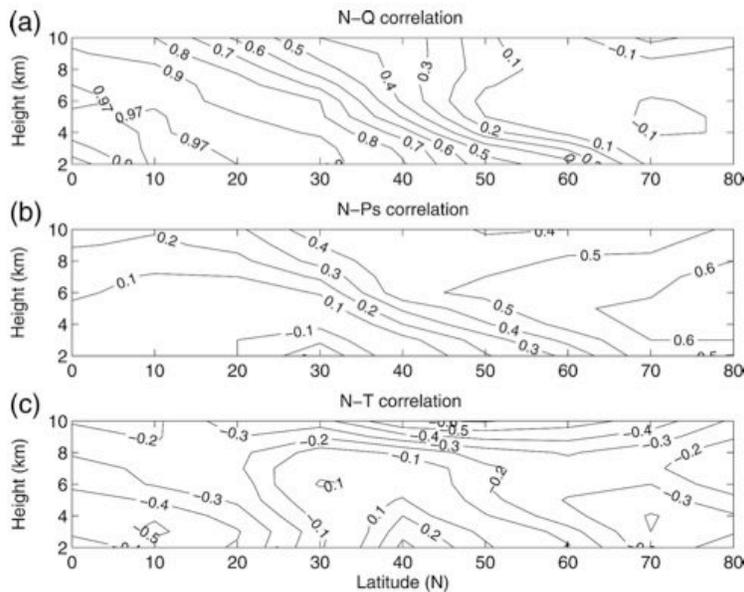


Figure 28: Northern Hemisphere time mean-zonal mean distribution of refractivity correlation with (a) specific humidity, (b) surface pressure, and (c) temperature for a December perfect model assimilation. From Liu et al. (2007), their Figure 3.

attempts to understand how assimilating GPS radio occultation measurements impacts the state vector, and what dependency that impact has on the physical region of assimilation.

Equation 4.3

describes how the state of the atmosphere determines

the refractivity at a given point. According to this equation, the three main components that define refractivity at a given point are temperature, pressure or atmospheric mass, and water vapor. Since these elements of the state vector have a direct relation to refractivity we expect them (or closely related quantities) to be highly correlated with refractivity in regions where their term in equation 4.1 dominates. Furthermore, since

these components of the state vector can be expected to be strongly correlated with refractivity in some regions, this implies that the assimilation of refractivity observations in those regions would cause large analysis increments updates to the state vector as described in equation 2.23. Therefore, by determining the spatial correlation structure of temperature, pressure, and water vapor with refractivity we can predict what state variables will be most influenced by the assimilation of GPS radio occultation measurements in various regions of the globe. Liu et al. (2007a) used this procedure in a perfect model setting to determine how observations would impact the state vector for a given December. Their results for the correlation structure of refractivity with components of the state vector for the Northern Hemisphere are given in figure 28. At high latitudes in the midtroposphere, particularly near the pole, they found that refractivity was significantly positively correlated with surface pressure. Meanwhile, refractivity was correlated with temperature in the upper troposphere in mid-high latitudes, and strongly correlated with water vapor (specific humidity) in the tropics.

While it is clear from equation 4.3 that refractivity should be strongly correlated with water vapor in regions where the atmosphere is moist such as the tropics, it is not as obvious how refractivity should correlate with temperature and atmospheric mass. To understand these correlation structures, we can separate the components of equation 4.3 by linearizing it about the ensemble mean state of $p_o, T_o,$ and q_o . It becomes

$$L(N) = N(p_o, T_o, q_o) + \left. \frac{\partial N}{\partial T} \right|_{p_o, T_o, q_o} (T - T_o) + \left. \frac{\partial N}{\partial P} \right|_{p_o, T_o, q_o} (P - P_o) + \left. \frac{\partial N}{\partial q} \right|_{p_o, T_o, q_o} (q - q_o) + O(2) \dots \quad (4.4)$$

This results in the linearized version of refractivity that can be written as

$$L(N) = \left(-l \frac{P_o}{T_o^2} - 2m \frac{P_o q_o}{T_o^3} \right) T + \left(m \frac{q_o}{T_o^2} + l \frac{1}{T_o} \right) P + \left(m \frac{P_o}{T_o^2} \right) q + const, \quad (4.5)$$

where l and m are unitless quantities with values $l = 77.6$ and $m = 5.997 \cdot 10^5$. This can be simplified as

$$L(N) = aT + bp + cq + const, \quad (4.6)$$

where a, b , and c represent the terms in parentheses in equation 4.5. Equation 4.3 has now become a simple formula that describes how refractivity is linearly influenced by local changes in T , p , and q . Finally, multivariate linear regression can be used to determine the covariances and correlations of N with T , p , and q . For the covariances we have

$$\begin{bmatrix} \langle N, T \rangle \\ \langle N, p \rangle \\ \langle N, q \rangle \end{bmatrix} = \begin{bmatrix} \langle T, T \rangle & \langle T, p \rangle & \langle T, q \rangle \\ \langle p, T \rangle & \langle p, p \rangle & \langle p, q \rangle \\ \langle q, T \rangle & \langle q, p \rangle & \langle q, q \rangle \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \quad (4.7)$$

Where $\langle \dots \rangle$ denotes a covariance, and a, b , and c are given in equation 4.6. In addition, plots of the fields described by a , b , and c are given in Figure 29. The covariances of the elements of the state vector between themselves are calculated from the GPS assimilation ensemble at each time step for each grid point. The correlation of N with q is

$$corr(N, q) = \frac{a \langle q, T \rangle}{\sqrt{[\langle q, q \rangle \langle N, N \rangle]}} + \frac{b \langle q, P \rangle}{\sqrt{[\langle q, q \rangle \langle N, N \rangle]}} + \frac{c \langle q, q \rangle}{\sqrt{[\langle q, q \rangle \langle N, N \rangle]}}, \quad (4.8)$$

and similar equations exist for the correlations of N with p and T . Equations 4.7 and 4.8 can be calculated at each time step for every grid point to get an estimate of the various covariances and correlations between refractivity and components of the state vector at that point.

Figure 30 shows the zonally averaged mean correlations of N with T, p , and q using statistics from the last 10 days of the GPS assimilation run for January 2007. The fields of T , p , and q were found in observation space by interpolating their value in state space to the location of each GPS occultation whose latitude, longitude, and height

are known. These fields were then updated by the assimilation, and the prior and post values of T , p , and q were recorded for each ensemble member. From these values, post and prior refractivity was calculated using equation 4.3. For this section, prior values are used exclusively since it is the properties of the prior ensemble that determines the ensemble update. These values were then binned and interpolated to a grid.

Since the GPS occultation

observations are globally distributed and numerous, bin size was chosen to maximize resolution while ensuring a significant number of observations would be collected in the bin over the 10 days. In the end the atmosphere was binned with 64 latitude bins, 128 longitude bins, and 18 height bins going from the surface up to 14km. Figure 31 gives the corresponding figure using our linearized equations 4.7 and 4.8. Comparing Figures 30 and 31, we see that the linearization method reproduces the observed correlations and covariances accurately. Also, comparing our results with Figure 28 it is clear that many of the December correlation structures found in Liu et al. (2007a) are reproduced in our analysis. At this point we now have a concise system of equations that we can use to

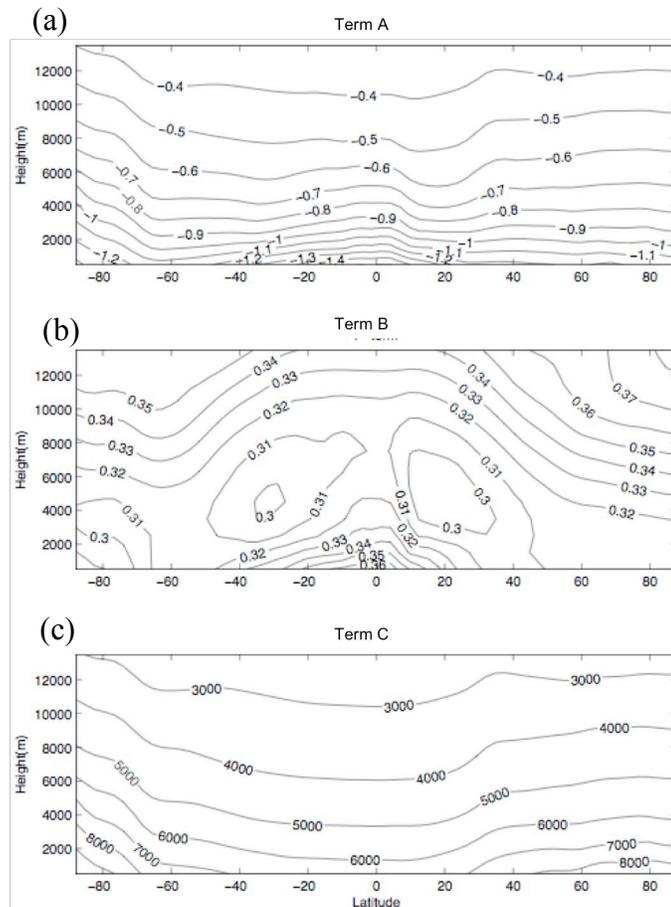


Figure 29: Zonal mean of (a) term a, (b) term b, and (c) term c from equation 4.6.

extract the components that determine the correlation of refractivity with the state vector. By doing this, we can consider them separately in order to understand the correlation structures of figures 28, 30, and 31.

We will first look at the correlation structure of refractivity and specific humidity. From Figure 31c, refractivity and specific humidity have a strong positive correlation in the tropics and lower troposphere. If we deconstruct equation 4.8, we find that there are three components that determine the correlation of refractivity and specific humidity.

Figure 32 shows the time-mean zonal average for each of the components. Here we find

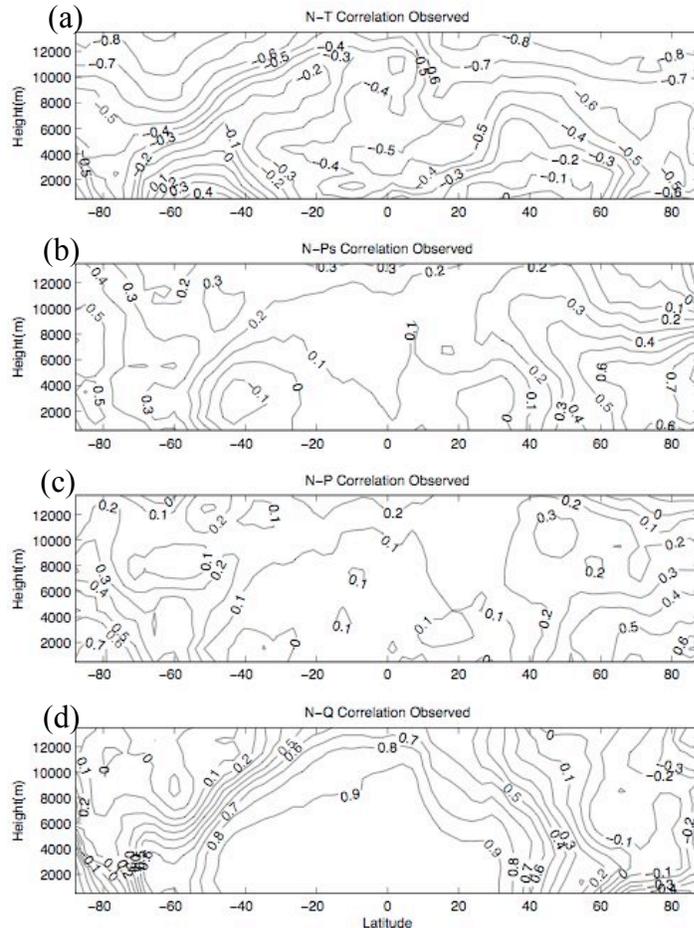


Figure 30: Observed zonal mean correlations between refractivity and (a) temperature, (b) surface pressure, (c) pressure, and (d) specific humidity for the January 2007 GPS assimilation.

that the variance of specific humidity is the dominant term in determining the correlation of refractivity and specific humidity. This term gives the strong positive response in the lower troposphere and tropics. In the upper troposphere the covariance terms of $\langle q, T \rangle$ and $\langle q, p \rangle$ tend to have equal and opposite impacts on the correlation of N and q. This indicates that variations in refractivity are highly sensitive to changes in water vapor in the tropics and lower troposphere, and

that little of this sensitivity due to changes in water vapor is from local dependencies on temperature and pressure variations. We can further deconstruct the third term of equation 4.8 by comparing the structures defined by the numerator and denominator. Figure 33c,d shows the multiplicative factor that the terms in the denominator represent. Figure 34 gives the corresponding plot for the numerator. Comparing figures 31c, 32, 33, and 34 we see that the correlation structure of refractivity with specific humidity is determined mostly by the strong variance of specific humidity in the tropics. In the extratropics, and in particular the upper atmosphere near the poles, the covariance of q with T and p becomes important.

We can use the linearized framework to further understand the relationship between refractivity, specific humidity, and the other state variables by using equation 4.7 to write

$$\langle q, N \rangle = a \langle q, T \rangle + b \langle q, p \rangle + c \langle q, q \rangle. \quad (4.9)$$

From equation 2.23 we can calculate what we will call the linearization analysis increment for q as

$$\Delta q = \frac{\langle q, N \rangle}{\langle N, N \rangle} \Delta N. \quad (4.10)$$

For this calculation we use the observed refractivity variance ($\langle N, N \rangle$) and analysis increment (ΔN), and $\langle q, N \rangle$ is calculated using equation 4.9. From here we can calculate a linearization posterior specific humidity field by

$$q_{post} = q_{prior} + \Delta q. \quad (4.11)$$

This new posterior field can now be compared with the observed posterior field for specific humidity. Now by repeating these calculations while setting either term 1, 2, or 3 in equation 4.9 to zero, we can clearly see the impact that the various covariance and

variances of equation 4.9 have on updates to q . Figure 35 shows the RMSE profiles between the actual posterior specific humidity field and our linearized version with various terms set to zero. As expected, Figure 35a shows that in the tropics the analysis increment is dominated by the variability of water vapor. Also, Figure 35b shows that in the midlatitudes the covariability of specific humidity with temperature and pressure plays a larger role, particularly at upper levels of the atmosphere.

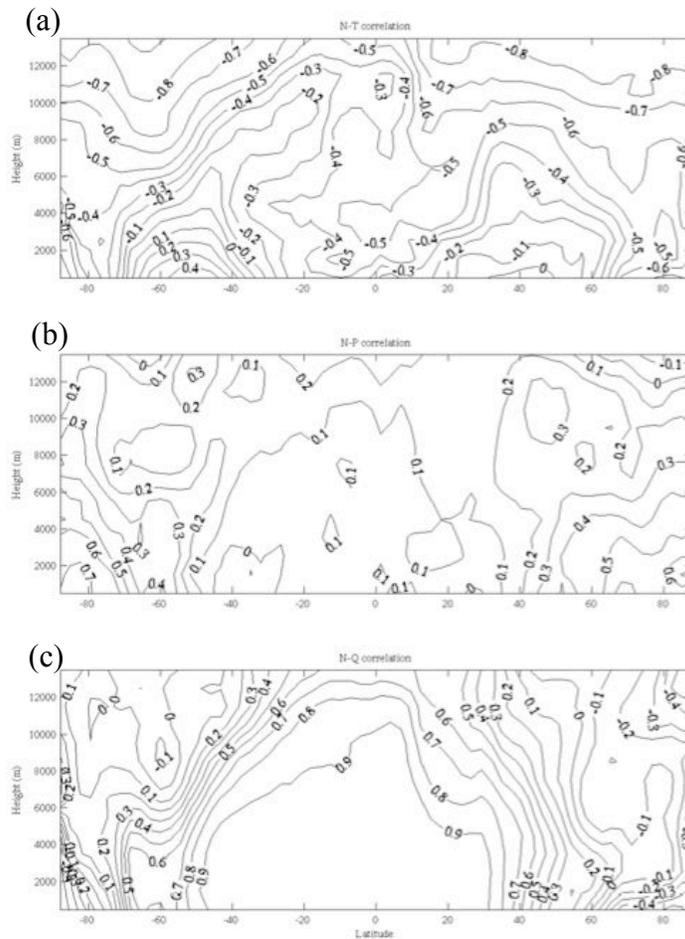


Figure 31: Zonal mean correlations from linearization between refractivity and (a) temperature, (b) pressure, and (c) specific humidity for the January 2007 GPS assimilation.

We can complete similar exercises for the correlation structure of refractivity with pressure and temperature as well. From Figure 31b it appears as if the correlation between refractivity and pressure is relatively weak throughout the atmosphere. The deconstruction of the terms of equation 4.8 with q replaced by p is given in Figure 36. From this figure we see that there is very little contribution from the covariance of pressure with specific humidity. Meanwhile, there are very clear structures both for the pressure variance term, and the $\langle p, T \rangle$ term. However, when summing these structures

there is almost a perfect cancellation between the two terms. In particular there are large cancellations between 40-60° and 6000-8000m in both hemispheres. This result seems to have important consequences for the potential impact of refractivity observations the

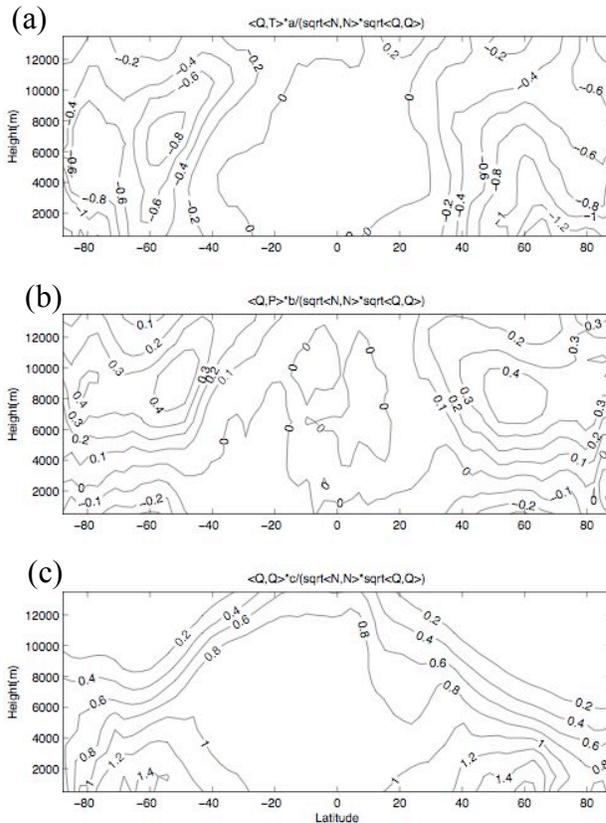


Figure 32: The time and zonal mean of (a) term 1 of correlation equation for specific humidity (eqn. 4.8), (b) term 2, and (c) term 3 for January 2007.

Moreover, Figure 38 shows the observed correlation of temperature with pressure across the ensemble. This shows that the region of strong cancellation is a local maximum in their correlation.

The region of large cancellation is also the same region where the variance of the first term of the refractivity becomes larger than the second term (Figure 21e). This indicates that in this region, any variations in the prior

state vector, since variations in the atmospheric state are only “correctable” to the extent that they cause substantial variations in atmospheric refractivity through the forward operator (equations 2.24-2.27). By going through the same process for p , as we did for q in equations 4.9-4.11 we see in Figure 37 that increments for p in the tropics are dominated by its covariance with q . However, in the midlatitudes both the $\langle q, q \rangle$ and $\langle q, T \rangle$ terms make significant impacts particularly in the vertical levels where large cancellations

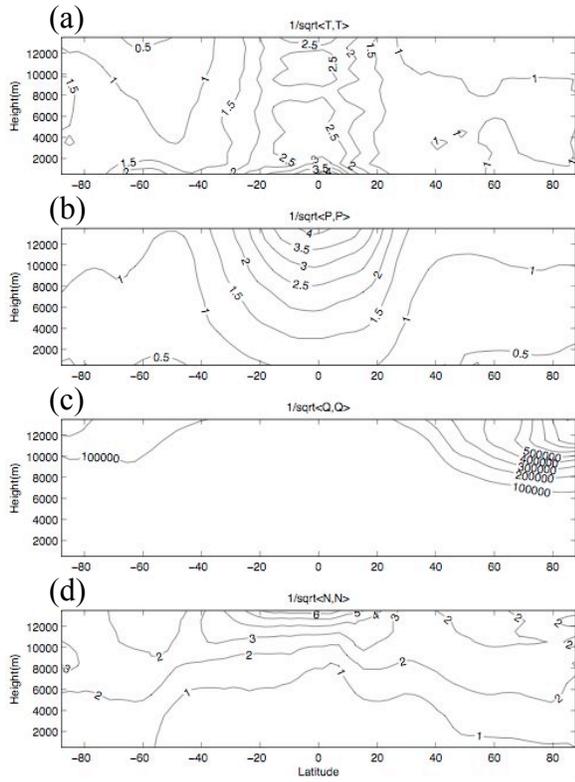


Figure 33: These panels represent the multiplicative factor that the variances in the denominator of the correlation equations (eqn. 4.8) for T,q, and p represent. (a) is the inverse of the square root of temperature variance, (b) is the same for pressure variance, (c) specific humidity variance, and (d) refractivity variance.

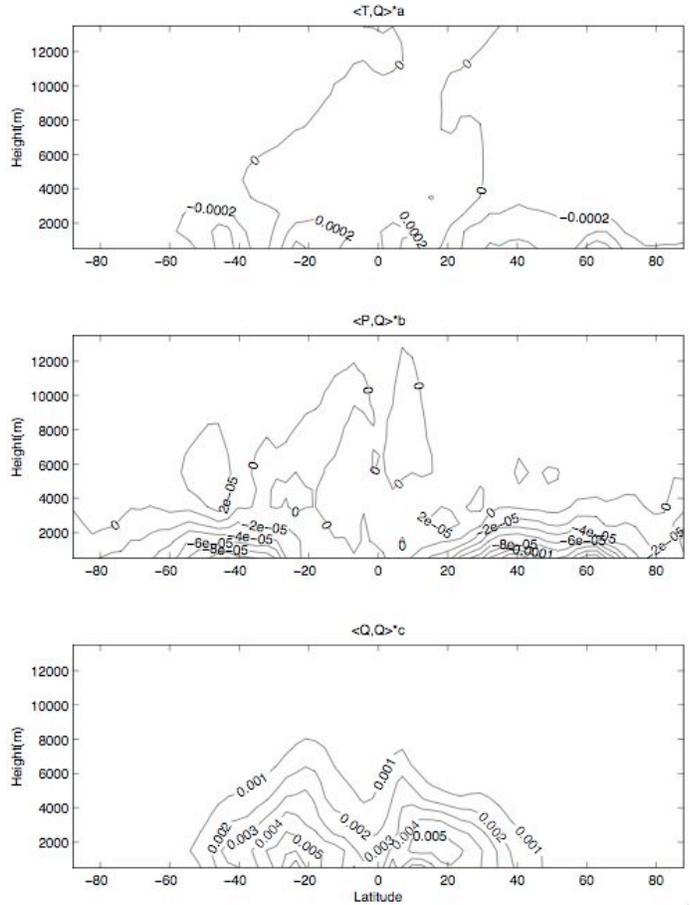


Figure 34: Numerator of term1 of correlation equation 4.8 for (a) specific humidity, (b) numerator of term2, and (c) numerator of term3.

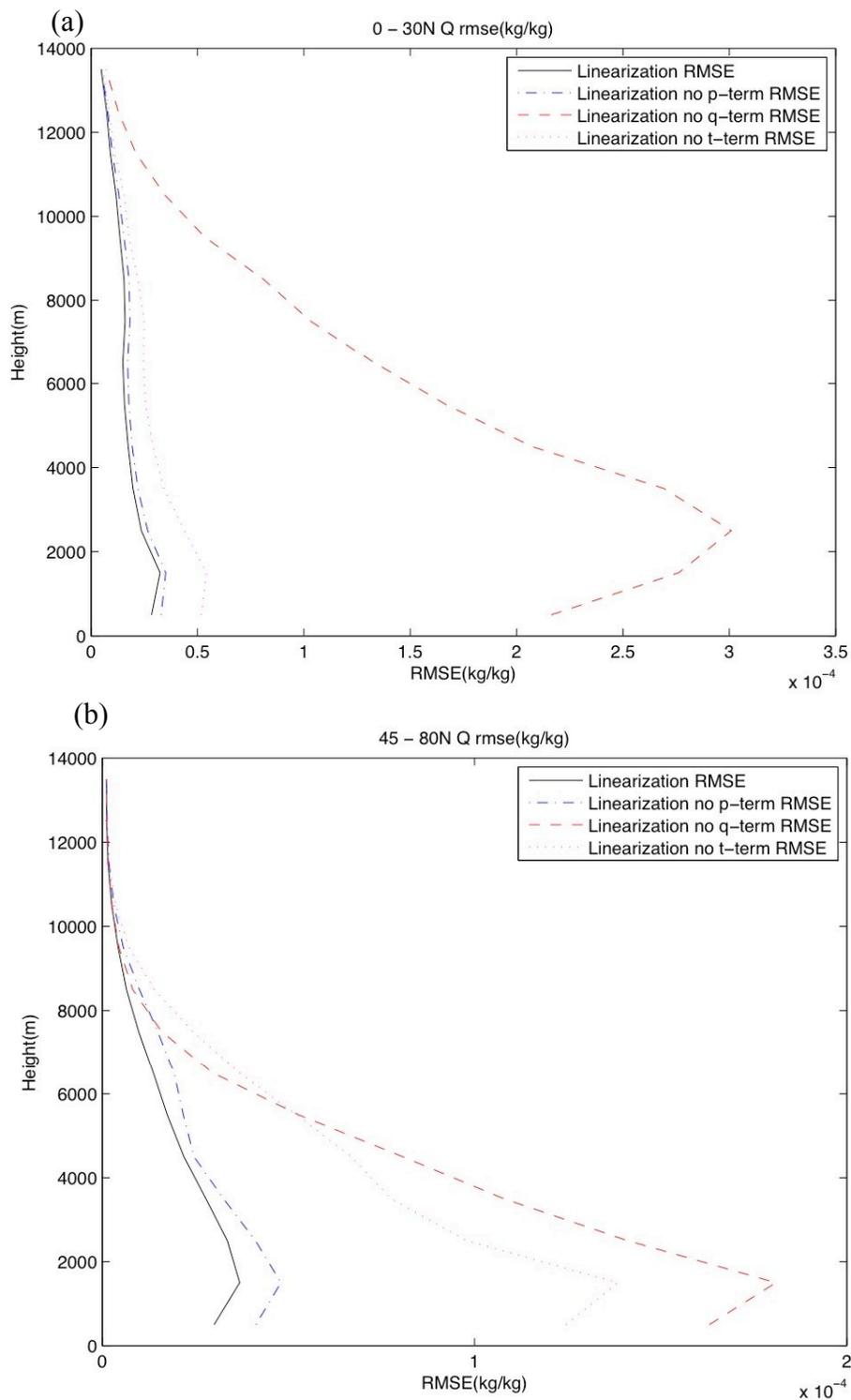


Figure 35: RMSE of specific humidity between actual posterior and linearized posterior in kg/kg (black line). Also shown are the RMSE with terms from eqn. 4.9 removed. Blue is the p-term removed, red q-term, and maroon the T-term. (a) represents the tropics, and (b) the midlatitudes.

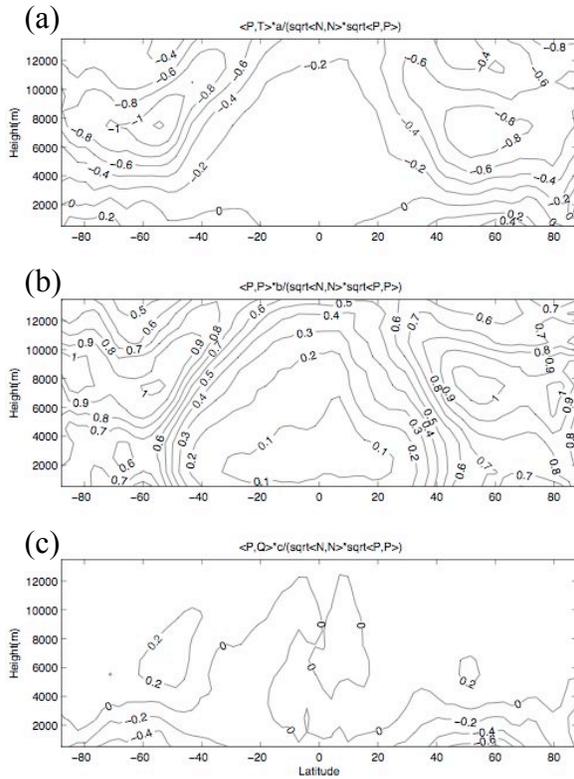


Figure 36: The time and zonal mean of (a) term 1 of correlation equation for pressure (eqn. 4.8 with q replaced with p), (b) term 2, and (c) term 3 for January 2007.

ensemble's estimation of refractivity through the forward operator are strongly influenced by the first term in equation 4.3, which is proportional to p/T . If N is proportional to p/T then changes in N are small if p and T are strongly positively correlated. Since T and p both decrease with height in the troposphere, we should expect cancellation if variations in p and T are happening because of vertical displacements.

We can inspect how the variance of refractivity in the ensemble is related

to pressure and temperature using the linearized refractivity equation 4.5. If we calculate a linearized prior refractivity with the water vapor term of equation 4.5 neglected, we find the variance of refractivity becomes that of Figure 39a. Now with water vapor removed the variance of refractivity can be written as

$$\langle N, N \rangle = b^2 \langle p, N \rangle + a^2 \langle T, N \rangle + 2ab \langle p, T \rangle. \quad (4.12)$$

Where a and b are coefficients given by equation 4.5. The variance of refractivity calculated from equation 4.12 is shown in Figure 39b. If p and T are independent as would be expected in a nonadiabatic relationship, this equation becomes

$$\langle N, N \rangle = b^2 \langle p, N \rangle + a^2 \langle T, N \rangle. \quad (4.13)$$

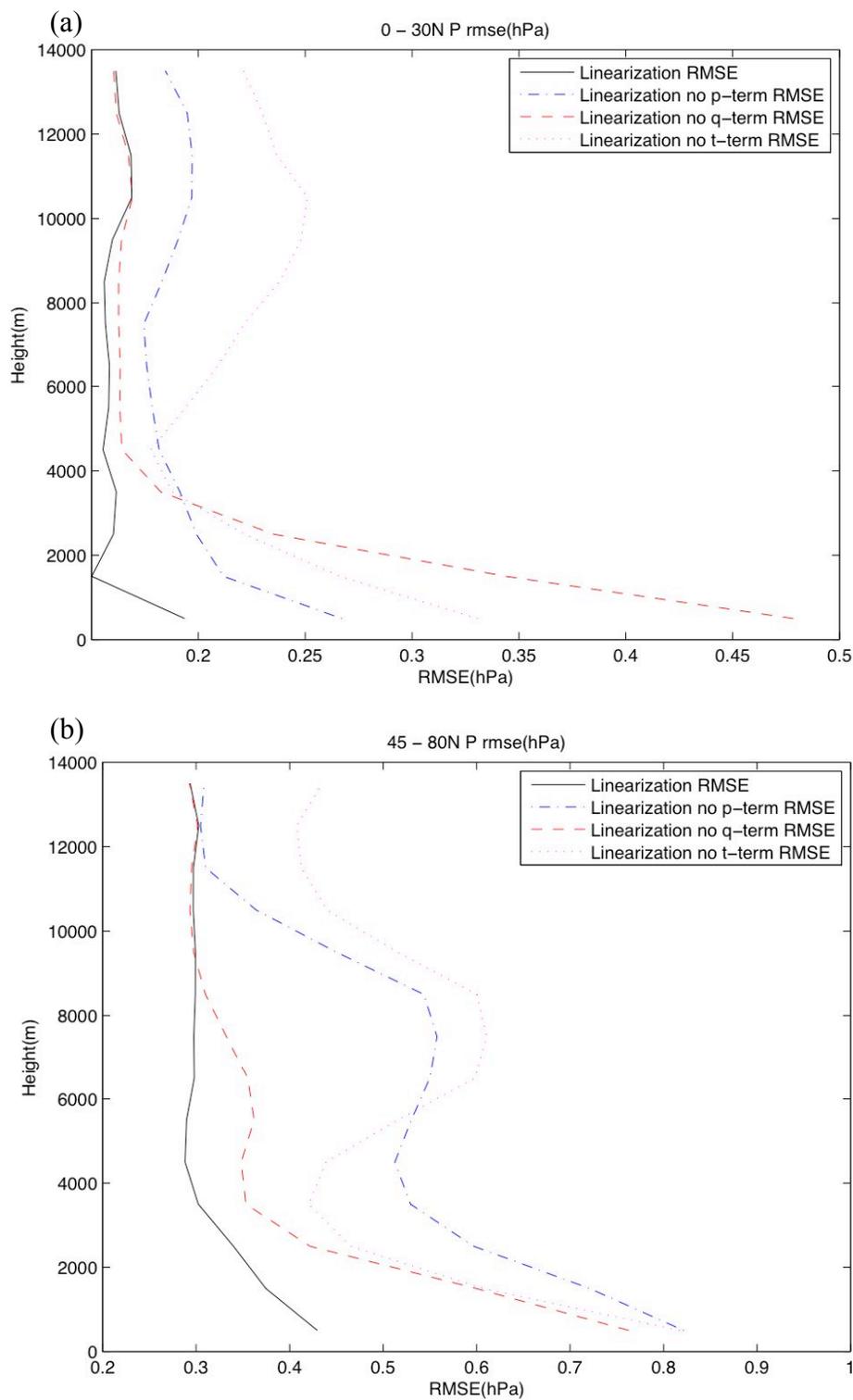


Figure 37: RMSE of pressure between actual posterior and linearized posterior in hPa (black line). Also shown are the RMSE with terms equivalent to eqn. 4.9 for pressure removed. Blue is the p-term removed, red q-term, and maroon the T-term. (a) represents the tropics, and (b) the midlatitudes.

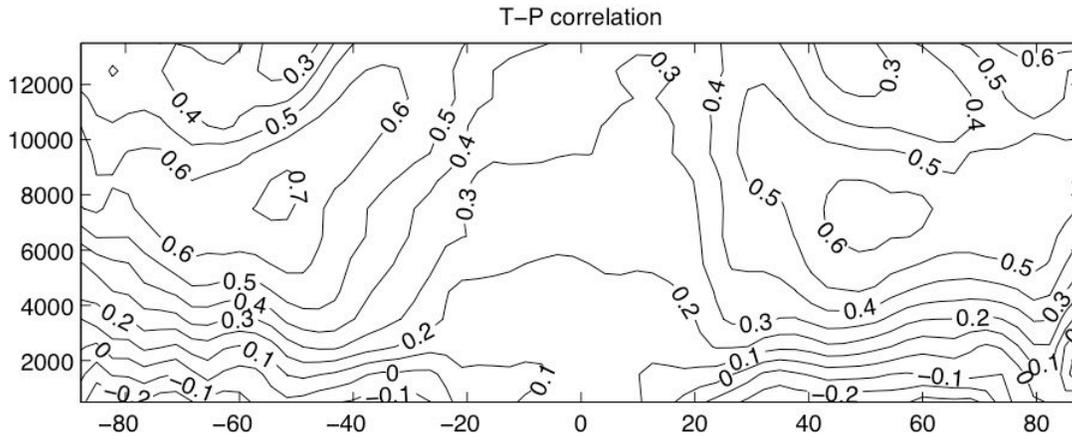


Figure 38: Correlation across prior ensemble of temperature with pressure.

The variance of refractivity with p and T independent is shown in Figure 39c, and the difference between the dependent variance and independent variance is shown in Figure 39d. Figure 39d implies that if T and p are allowed to covary there is locally less variance of refractivity in the regions of strong cancellation. This means that in these regions, the covariance of T and p across the ensemble is reducing the efficiency by which GPS refractivity observations can correct the ensemble.

However, there is one major caveat to this discussion. In general, people are not concerned with updates to the pressure field of a model since it is neither a prognostic or diagnostic variable. However, the previous discussion seems to indicate that cancellation may also strongly impact ability of refractivity to correct variations in geopotential height, since geopotential height on pressure surfaces and pressure on geometric surfaces are so closely related. However, Figure 40 shows the correlation between refractivity and geopotential height on pressure surfaces. This shows that similar to the correlation between refractivity and pressure there is little correlation in the tropics. However, unlike pressure, geopotential height retains a relatively large correlation in the regions of large cancellation for pressure. This indicates that refractivity observations are capable

of correcting variations in geopotential height. The discrepancy between these two results requires more research.

Finally, inspection of figure 31a shows that, in general, refractivity is negatively correlated with temperature. Also, the higher in the atmosphere the stronger this negative correlation becomes. One feature that deserves closer inspection is found below 2000m

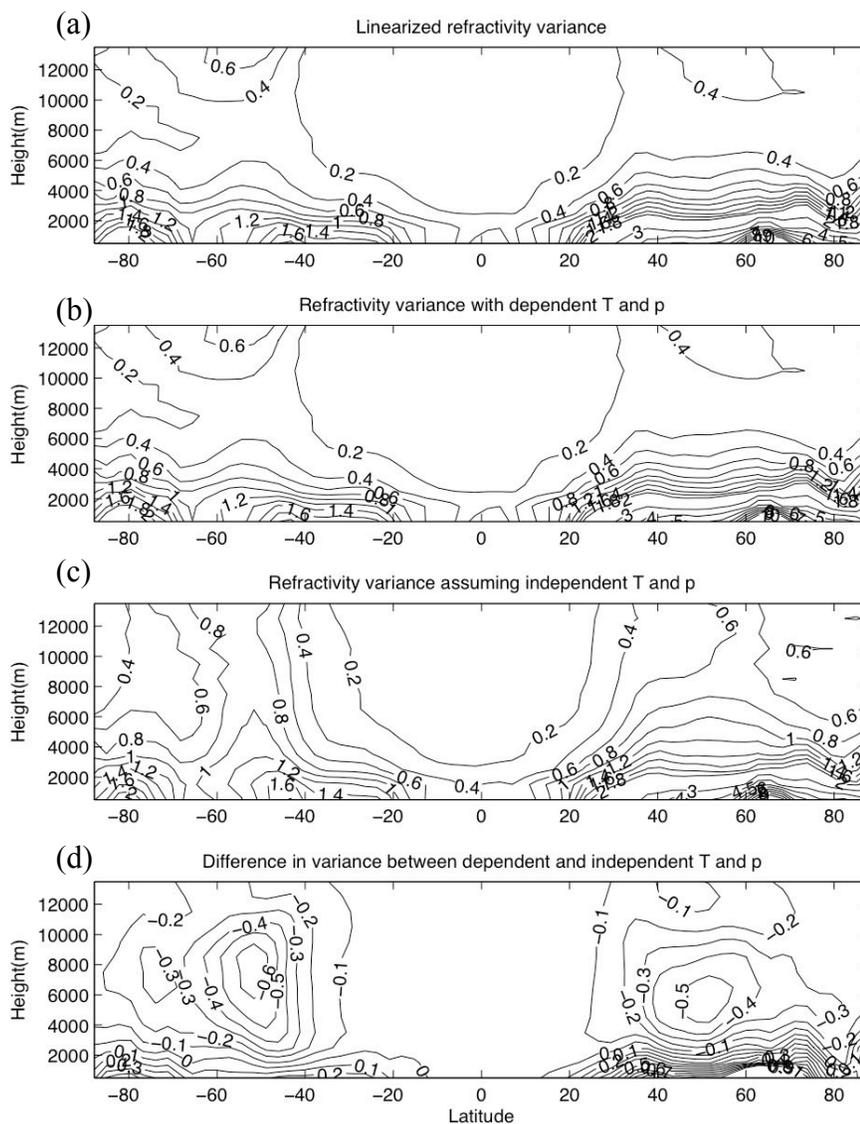


Figure 39: (a) is the linearized variance of refractivity neglecting specific humidity contributions. (b) is the refractivity variance from equation 4.12. (c) is the refractivity variance from equation 4.13 assuming independent T and p. (d) is the difference between (b) and (c).

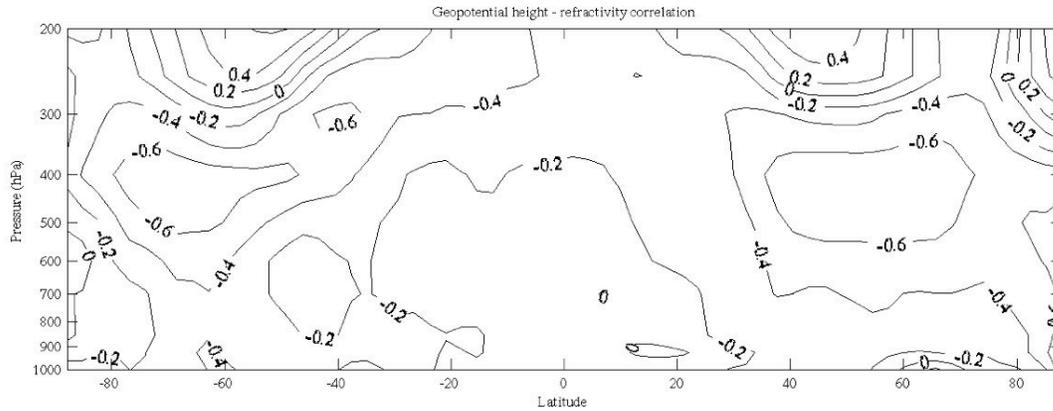


Figure 40: (a) Correlation of geopotential height with refractivity in pressure coordinates. 200hPa roughly approximates 12km in previous height coordinate plots.

near 50°S. Here the correlation between refractivity and temperature actually becomes positive, the opposite of what would be expected from density considerations. Also, from Figure 31a it seems that this positive anomaly extends upward and to the south causing a sharp gradient in the N-T correlation structure. Figure 41 shows the deconstruction of equation 4.8 with temperature replacing specific humidity. Here we see that the region where the correlation becomes positive, is precisely the region where the $\langle T, q \rangle$ term dominates. Also, the influence of the water vapor term extends upwards and to the south in the same structure found in figure 31a. This structure combines with cancellation from $\langle T, P \rangle$ to reduce the correlation of refractivity with temperature in an isolated region, and create a strong gradient in the correlation field above it. Figures 33a,d and 42 show the respective numerator and denominator deconstruction. Inspection of Figure 42 confirms that the strong covariance of T with q near 50°S is responsible for the sign reversal.

e. Conclusions for the Assimilation of GPS Observations in CAM

These experiments begin to show the influence that GPS radio occultation observations can have when assimilated with an atmospheric climate model. By

comparing a case where we assimilate only COSMIC data into CAM versus assimilating multiple data sources or no data sources it is apparent that assimilating GPS observations with the EAKF and CAM does improve CAM's representation of the atmosphere on a given day. In addition, when adding GPS occultation measurements to existing measurements it seems that GPS measurements do add information that was previously missing. However, it is also evident that assimilating GPS radio occultation observations alone does not completely constrain the model. When we compare the GPS run to other runs that assimilate observations from radiosondes, other satellites, etc... there is clearly a lot of information in those observations that GPS observations alone cannot reproduce. For climate studies, the same can be said about major biases. When we assimilated GPS observations we were able to reduce all the biases we looked at. However, the assimilation of these observations did not constrain the model to such a degree as to overcome these biases completely. In addition, we highlighted considerations for the calculating transient momentum fluxes using ensemble data. Here it appeared that calculating transient momentum fluxes with ensemble mean wind fields degrades the calculation versus calculating the transient momentum fluxes for the individual ensemble members, and taking the ensemble mean of this quantity.

When looking for ways to improve the power of GPS observations to constrain CAM there are several promising possibilities for future work. One possible way to pull more data from the observations is to include data from the vector path of the GPS observation in the forward operator. This data is available with the refractivity observations, but its assimilation becomes more computationally expensive. However, methods that do assimilate this information have been implemented in several systems

already, with promising results (Liu et al. 2007b). Also, as more GPS satellites continue

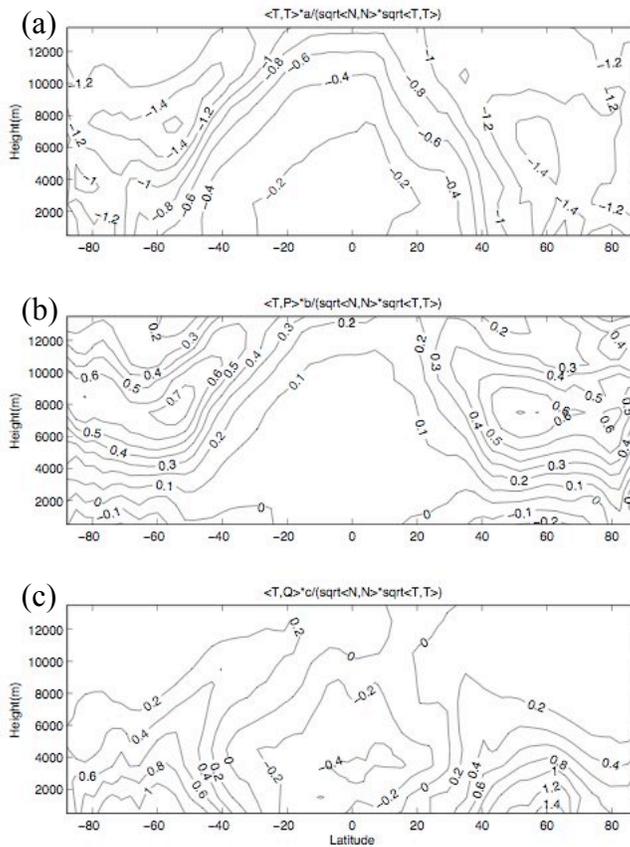


Figure 41: The time and zonal mean of (a) term 1 of correlation equation for specific humidity (eqn. 4.8 with q replaced by T), (b) term 2, and (c) term 3 for January 2007.

to launch the number of observations will continue to increase, further improving their coverage. Finally, more months of assimilation with varying ensemble sizes and varying model resolutions would be useful to see how the ability of assimilated GPS observations reduce model biases depends on these parameters.

The fact that refractivity is dependent on multiple pieces of the state vector is a fundamental property of these observations is

self-evident. Since refractivity is essentially a function of density in the lower atmosphere, the assimilation of refractivity observations will influence the various components of the state vector that determine the density of the atmosphere at a given point. The proportion of density that is due to temperature or pressure or water vapor changes depending on where in the atmosphere one looks. This makes understanding the relationship between refractivity and components of the state vector essential to predicting how the assimilation of a GPS radio occultation measurement will impact the modeled state of the atmosphere. By linearizing the forward operator we were able to

separate the components that strongly influence refractivity. Once separated, we could determine what components of the state vector were strongly influenced by refractivity observations. One result that came out of this is that the relationship between the state

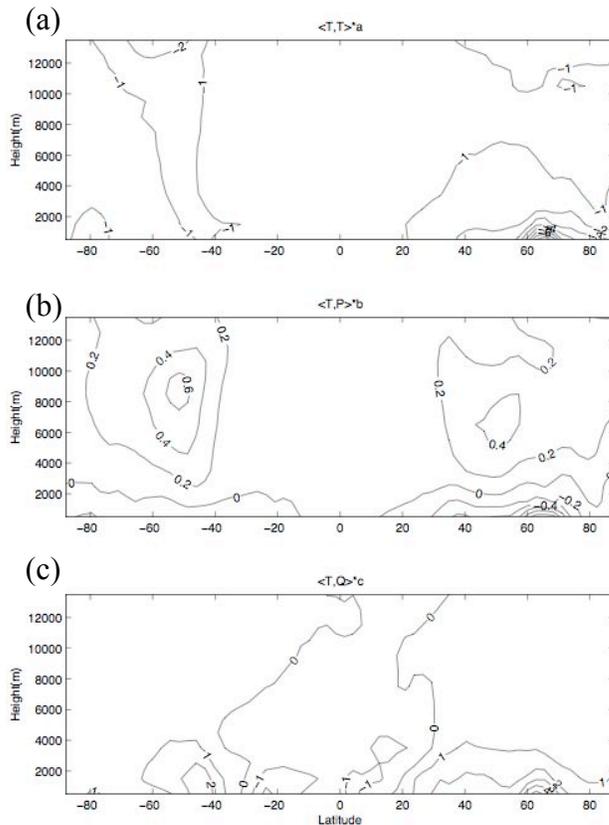


Figure 42: (a) Numerator of term 1 of correlation equation for temperature (eqn. 4.8 with q replaced by T), (b) numerator of term 2, and (c) numerator of term 3.

vector and refractivity in the tropics appears to be dominated by water vapor. In this region it is not clear how much the assimilation of GPS observations will impact any variable besides those that are water vapor related. Possible indications of this may be seen in figures 23 and 27 where the assimilation of GPS observations appeared to do little to reduce tropical biases in temperature and transient momentum fluxes.

More work is needed to determine if GPS observations are capable of reducing tropical biases for fields not

influenced by water vapor.

Finally, we found that large cancellations in the prior estimate the correlation of pressure with refractivity occur in the midlatitudes at 6000-8000m due to covariance between temperature and pressure. This has important implications for the power of GPS observations to adjust the atmospheric state in these regions. Adiabatic variations in the

ensemble members appear to be partially responsible for this cancellation. However, calculations of the correlation of refractivity with geopotential height did not mirror this result, so whether this result is consequential requires more analysis.

5. Conclusion

a. Summary and Future Work

This research was based on taking a first step toward the overarching goal of determining how data assimilation can be used to improve our understanding of climate models, and what value various observations and data assimilation systems can add to future work. With the large number of possible directions to go to achieve these goals, and the always evolving and expanding set of observations available for assimilation, we approached this problem by considering a few simple examples.

In Chapter 3, we used a brute force method of selectively assimilating data only in the region of known climate bias in CAM. We found that by reducing this bias we were able to begin to understand the sensitivity of major climate features such as the zonal jet and stormtracks to this bias. In doing so, we found that our work mirrored many of the results found in previous studies such as that of Lorenz and DeWeaver (2007). This sets the stage for us to use more advanced ensemble assimilation techniques such as sensitivity analysis as described by Torn and Hakim (2008) to more elegantly describe the impacts of this bias. Also, with many climate features including the zonal jet being effectively tuned by parameterization, the use of data assimilation for parameter estimation could also be an important future step.

Chapter 4 took a close look at the assimilation of a single observation type, the GPS radio occultation observation. This observations type was of particular interest for several reasons. First, this observations type is a relatively recent addition to the global observation net. Also, GPS radio occultation observations have exciting implications for

climate research due to its long-term stable nature, global coverage, and ability to ignore clouds. By considering the effectiveness of this observation to reduce CAM model bias, we found that these observations were valuable. However they were not effective to the point where they could be used exclusively in the EAKF to reproduce a decent facsimile of the actual atmosphere. However, the ability of GPS observation assimilation to reproduce a close representation of the actual atmosphere may be improved in future work. Some possibilities for improvement include integrating the application of a non-local forward operator that takes into account additional path information for GPS observations, and varying ensemble sizes that may improve the first guess required to separate out the connection between refractivity and the components of the state vector.

There are many possibilities for using data assimilation to go beyond its traditional use as a tool for creating weather forecasts and reanalysis products, and become an instrument for investigating dynamical relationships and diagnosing the impacts of model biases by creating a system capable of combining the advantages of both observed and modeled data. However, in order for data assimilation to become commonly used in that capacity, many more questions must be answered. Taken as a whole, the experiments contained in this research represent investigations into a small selection of the questions that need to be answered for the EnKF to reach its full potential.

References

- Anderson, J. L., and S.L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.
- Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.
- Annan, J.D., D. J. Lunt, J. C. Hargreaves, and P. J. Valdes, 2005: Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Non. Proc. in Geo.*, **12**, 363-371.
- Ancell, B., and G. J. Hakim 2007: Comparing adjoint and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117-4134.
- Chen, G., I.M. Held, and W.A. Robinson, 2007: Sensitivity of the Latitude of the Surface Westerlies to Surface Friction. *J. Atmos. Sci.*, **64**, 2899–2915.
- Collins, W.D., C.M. Bitz, M.L. Blackmon, G.B. Bonan, C.S. Bretherton, J.A. Carton, P. Chang, S.C. Doney, J.J. Hack, T.B. Henderson, J.T. Kiehl, W.G. Large, D.S. McKenna, B.D. Santer, and R.D. Smith, 2006: The Community Climate System Model Version 3 (CCSM3). *J. Climate*, **19**, 2122–2143.
- Covey, C., K. M. AchutaRao, U. Cubasch, P.D. Jones, S.J. Lambert, M. E. Mann, T. J. Philips, and K. D. Taylor, 2003: An overview of results from the Coupled Model Intercomparison Project (CMIP), *Global Planet. Change*, **37**, 103-133.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 1043-1062.
- Evensen, G., 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342-1354.
- Frierson, D. M. W., I. M. Held, and P. Zurita-Gator, 2006: A gray-radiation aquaplanet moist GCM. Part I: Static stability and eddy scales. *J. Atmos. Sci.*, **63**, 2548-2566.

- Gaspari G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Hack, J.J., J.M. Caron, G. Danabasoglu, K.W. Oleson, C. Bitz, and J.E. Truesdale, 2006: CCSM–CAM3 Climate Simulation Sensitivity to Changes in Horizontal Resolution. *J. Climate*, **19**, 2267–2289.
- Healy, S. B., and J.-N. Thepaut, 2006: Assimilation experiments with CHAMP GPS radio occultation measurements. *Quart. J. Roy. Meteor. Soc.*, **132**, 605–623.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteor. Soc. Japan*, **75**, 181–189.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Kalnay, E., B. Hunt, E. Ott, and I. Szunyogh, 2006: Chapter 7 of: *Predictability of Weather and Climate*. Cambridge University Press, 701 pp.
- Kalnay, E., H. Li, T. Miyoshi, S. Yang, and J. Ballabrera-Poy, 2007: 4-D-Var or ensemble Kalman filter? *Tellus A*, **59 (5)**, 758–773 doi:10.1111/j.1600-0870.2007.00261.x
- Kursinski, E. R., G. A. Hajj, K. R. Hardy, J. T. Shofield, and R. Linfield, 1997: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. *J. Geophys. Res.*, **102**, 429–465.
- Kushner, P.J., I.M. Held, and T.L. Delworth, 2001: Southern Hemisphere Atmospheric Circulation Response to Global Warming. *J. Climate*, **14**, 2238–2249.

- Liu, H., J.L. Anderson, Y.H. Kuo, and K. Raeder, 2007: Importance of Forecast Error Multivariate Correlations in Idealized Assimilations of GPS Radio Occultation Data with the Ensemble Adjustment Filter. *Mon. Wea. Rev.*, **135**, 173–185.
- Liu, H., J. Anderson, Y.H. Kuo, C. Snyder, and A. Caya, 2008: Evaluation of a Nonlocal Quasi-Phase Observation Operator in Assimilation of CHAMP Radio Occultation Refractivity with WRF. *Mon. Wea. Rev.*, **136**, 242–256.
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation system. *Mon. Wea. Rev.*, **109**, 701–721
- Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *J. of Atmos. Sci.* **20**, 130-141.
- Lorenz, D. J., and E. T. DeWeaver, 2007: Tropopause height and zonal wind response to global warming in the IPCC scenario integrations. *J. Geophys. Res.*, **112**, doi:10.1029/2006JD008087.
- Morss, R. E., and D. S. Battisti, 2004: Evaluating observing requirements for ENSO prediction: Experiments with an intermediate coupled model. *J. Climate*, **17**, 3057-3073.
- Persson, A., and F. Grazzini, 2005: Users Guide to ECMWF forecast products. Version 4.0. Available from <http://ecmwf.int/products/forecasts/guide/index.html>.
- Polvani, L. M., and P. J. Kushner, 2002: Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophys. Res. Lett.*, **29**, 114, doi:10.1029/2001GL014284.
- Robinson, W. A., 1997: Dissipation dependence of the jet latitude. *J. Clim.*, **10**, 176-182.
- Sokolovskiy, S., Y. H. Kuo, and W. Wang, 2005: Assessing the accuracy of a linearized observation operator for assimilation of radio occultation data: Case simulations with a high-resolution weather model. *Mon. Wea. Rev.*, **133**, 2200-2212.
- Son, S. W., and Lee, 2005: The response of westerly jets to thermal driving in a primitive equation model. *J. Atmos. Sci.*, **62**, 3741-3757.
- Talagrand, O., 1997: Assimilation of Observations, an Introduction. *J. Meteor. Soc. Japan*, **75(1B)**, 191-209.
- Thompson, D. W. J., M. P. Baldwin, and S. Solomon, 2005: Stratosphere-troposphere coupling in the Southern Hemisphere. *J. Atmos. Sci.*, **62**, 708-715.
- 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.

- Torn, R. D. and G. J. Hakim, 2008, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663-677.
- Ware, R. H., and Coauthors, 1996: GPS sounding of the atmosphere from low Earth orbit: Preliminary results. *Bull. Amer. Meteor. Soc.*, **77**, 19-40.
- Williams, G. P., 2006: Circulation sensitivity to tropopause height. *J. Atmos. Sci.*, **63**, 1954-1961.
- Yeager, S. G., W. G. Large, J. J. Hack, and C. A. Shields, 2006: The Low Resolution CCSM3. *J. Clim.*, **19**, 2545-2566.
- Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.*, **32**, L18701, doi:10.1029/2005GL023684.
- Zhang, S., and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus A*, **55**, 126-147.