REGIONALLY ENHANCED GLOBAL DATA ASSIMILATION (REG DA): AN EVALUATION OF THE LIMITED AREA MODEL PERFORMANCE

A Thesis

by

ADAM R. BRAINARD

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Istvan Szunyogh
Committee Members,	Christopher Nowotarski
	Ping Chang
Head of Department,	Ping Yang

May 2017

Major Subject: Atmospheric Sciences

ABSTRACT

Regionally Enhanced Global (REG) Data Assimilation (DA) is a method of global data assimilation in which high-resolution information from a single or multiple Limited Area Model (LAM) domains is blended with the global model information to create a regionally enhanced analysis of the global atmospheric state. This approach has been demonstrated to benefit both local and global model forecasts in idealized studies but has never been tested on operational numerical weather prediction models. This study investigates the limited area model forecast performance of an implementation of the REG DA approach on the operational 4D-Var data assimilation system, global model, and limited area model of the U.S. Navy. This implementation is called REG 4D-Var. The results of analysis-forecast experiments with the system show that the approach leads to small, but statistically significant overall forecast improvements and large and significant forecast improvements for Hurricane Sandy.

DEDICATION

I dedicate this thesis to my family. Life may be a state of constant change, but the love and support I feel from my family has been eternal.

ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Istvan Szunyogh for the opportunity to participate in this research endeavor while attaining a graduate education at Texas A&M. His support has been vital to my growth as an atmospheric scientist, and has allowed me to continue my development toward a very exciting career.

I would next like to acknowledge Dr. Gyorgyi Gyarmati and Dr. Szunyogh's research group: Michael Battlio, Fan Han, Max Gawryla, Carlee Loeser, and my mentor Dr. Michael A. Herrera. It is difficult to express the gratitude I owe Michael. His knowledge and experience were vital to my research endeavor. I could not have completed this work without his enthusiastic support and friendship.

I would finally like to thank my friends, both new and old. I arrived at Texas A&M without a personal connection to anyone or anything, but I immediately felt at home with my many wonderful peers at this institution. I quickly developed bonds with these peers inside and outside the classroom – friendships which have been invaluable to my time in graduate school – and I hope continue well into the future.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professors Istvan Szunyogh and Christopher Nowotarski of the Department of Atmospheric Sciences and Professor Ping Chang of the Department of Oceanography.

The study shown was conducted in collaboration with scientists from two of the NRL institutions. From NRL Monterey: Craig Bishop, Allen Zhao, Jason Nachamkin, and Teddy Holt, and from NRL DC: Dave Kuhl and Karl Hoppel.

Michael A Herrera performed the implementation of REG DA and produced the global analyses and boundary conditions used in the limited area model integrations (Herrera, 2016).

Funding Sources

Graduate study was supported by a teaching assistantship from Texas A&M University and the following grants:

- Office of Naval Research Grant N00014-12-1-0785
- Office of Naval Research Grant N00014-16-1-3011
- Naval Research Lab Award N00173-16-1-3011

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Over the past few decades, two fundamental types of operational numerical weather prediction models have evolved: global models, which cover the entire globe, and limited area models (LAMs), which cover only part of it. Given finite computational resources, there are limitations of the spatiotemporal resolution at which a model can be feasibly used. LAMs, with their more targeted area of interest, are typically operated at higher resolutions than their global counterparts, but inherently rely on information from a global model for the atmospheric state outside the LAM domain.

Historically, the interactions between the global and the concurrent LAM forecasts, have been limited to the global model providing boundary conditions to the higher resolution LAM forecasts. An opportunity exists, however, to introduce additional interactions between the global model and the limited area models through a data assimilation (DA) system that provides both the global model and the LAM with initial conditions.

Yoon et al. (2012) compared the traditional data assimilation strategy, which prepares the global initial condition and the LAM initial conditions independently, with an integrated data assimilation strategy, using an idealized Lorenz model of the atmosphere. The integrated approach led to lower root-mean-square errors (RMSE) for both the global analyses (18%) and forecasts (15%) and the LAM analyses (6%) and forecasts (5%). A similar idealized comparison study was performed by Kretschmer et al. (2015), who employed the Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al., 2007) to analyze a *composite state* that included all state vector components from a global model and multiple LAMs. The composite state provided both the global model and the LAMs with initial conditions. The technique led to significant analysis and forecast improvements for both the global model and the LAMs compared to the analyses and forecasts that were pro-

duced by using initial conditions obtained by the conventional independent data analysis approach.

Our research is motivated by the highly promising results of the idealized data assimilation studies. We carry out experiments with an operational global model and LAM. Our approach for the integration of the data assimilation process is similar to the composite state method of Kretschmer et al. (2015): short-term forecast states obtained by the global and limited area model are linearly combined to produce a blended estimate of the atmospheric state, which is then used to compute the predicted value of the observations in the data assimilation. The background estimate of the global state is updated by assimilating observations by the global data assimilation system, using the innovations (the difference between the observations and their predicted values) computed with the help of the blended state. The resulting global analysis provides both the global and limited area model with initial conditions after interpolation to match the global and limited area model representation of the atmospheric fields. We call this DA approach the Regionally Enhanced Global (REG) DA. The evaluation of the impact of REG DA on the global model forecast performance was the subject of an earlier study (Herrera, 2016). The focus of the present thesis is on the evaluation of the impact on the LAM forecast performance.

The thesis is organized as follows. Section 2 describes the general formulation of REG DA for a 4D-Var DA scheme, which we call REG 4D-Var, and our particular implementation on the operational models and data assimilation system of the U.S. Navy. Section 3 describes the design of the analysis-forecast experiment and introduces the verification scores used for the evaluation of the experiments. Section 4 discusses the results of the experiments, and section 5 provides the conclusions.

2. REG 4D-VAR

2.1 Formulation

2.1.1 4D-Var

REG DA is independent of the particular method of data assimilation used (e.g., 3D-Var, 4D-Var, or EnKF). Because we use the operational DA system of the U.S. Navy, which is 4D-Var, we introduce REG DA assuming the availability of a global 4D-Var system. We call this particular implementation of the REG DA concept REG 4D-Var.

Today's data assimilation systems use a sequential algorithm to obtain the analysis. The analysis $\mathbf{x}^a(t_0)$ at analysis time t_0 is obtained by

$$\mathbf{x}_g^a(t_0) = \mathbf{x}_g^b(t_0) + \delta\mathbf{x}_g^a(t_0), \tag{2.1}$$

where the $\mathbf{x}_g^b(t_0)$ background estimate of the state is a short term forecast valid at t_0, and $\delta\mathbf{x}_g^a(t_0)$ is the analysis increment. The 4D-Var analysis increment $\delta\mathbf{x}_g^a$ is the $\delta\mathbf{x}_g(t_0) = \delta\mathbf{x}_g^a$ minimizer of the cost function (e.g. Szunyogh, 2014)

$$\begin{aligned}\mathbf{J}\left[\delta\mathbf{x}_g(t_0)\right] = &\left[\delta\mathbf{x}_g(t_0)\right]^{\mathbf{T}}\left(\mathbf{P}^b\right)^{-1}\delta\mathbf{x}_g(t_0)+ \\ &\sum_{j=0}^{N}\left[\delta\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_g(t_j)\mathbf{M}_g(t_0,t_j)\delta\mathbf{x}_g(t_0)\right]^{\mathbf{T}}\times \\ &\mathbf{R}_{t_j}^{-1}\left[\delta\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_g(t_j)\mathbf{M}_g(t_0,t_j)\delta\mathbf{x}_g(t_0)\right],\end{aligned} \tag{2.2}$$

where $\mathbf{P}^b$ is the background error covariance matrix, $\mathbf{H}_g(t_j)$ is the linearization of the observation function $\mathbf{h}_g\left[\mathbf{x}_g(t_j)\right]$ about $\mathbf{x}_g^b(t_j)$, $\mathbf{M}_g(t_0,t_j)\delta\mathbf{x}_g(t_0)$ represents the tangent-linear model (TLM) integration, $\mathbf{R}$ is the observational error covariance matrix, and $t_j, j = 0, ..., N$, are the times at which observations are available within the assimilation time

window. In Eq. (2.2), the expression within the square brackets of the second term may also be written as

$$\mathbf{y}^{\mathrm{o}}(t_j) - \mathbf{h}_g\left[\mathbf{x}_g(t_j)\right] = \delta\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_g(t_j)\mathbf{M}_g(t_0, t_j)\delta\mathbf{x}_g(t_0), \tag{2.3}$$

where

$$\delta\mathbf{y}^{\mathrm{o}}(t_j) = \mathbf{y}^{\mathrm{o}}(t_j) - \mathbf{h}_g\left[\mathbf{x}_g^b(t_j)\right] \tag{2.4}$$

is the innovation, and because $\mathbf{H}_g(t_j)$ is the linearization of $\mathbf{h}_g\left[\mathbf{x}_g^b(t_j)\right]$ about $\mathbf{x}_g^b\,(t_j)$,

$$\mathbf{h}_g\left[\mathbf{x}_g(t_j)\right] \approx \mathbf{h}_g\left[\mathbf{x}_g^b(t_j)\right] + \mathbf{H}_g(t_j)\delta\mathbf{x}_g(t_j). \tag{2.5}$$

2.1.2 Regional Enhancement of the Background

To enhance the background estimate of the global state by incorporating higher resolution information from a LAM, the observation function $\mathbf{h}_g$, operating on the global state $\mathbf{x}_g$, is replaced by $\mathbf{h}_e$, which operates on the regionally enhanced, blended state $\mathbf{x}_e$, defined by

$$\mathbf{x}_e = (1-\alpha)\mathbf{L}(\mathbf{x}_g) + \alpha\mathbf{x}_\ell. \tag{2.6}$$

Here, $\mathbf{L}$ is the linear operator that maps the lower resolution global model state $\mathbf{x}_g$ to the higher resolution LAM representation of the state, $\mathbf{x}_\ell$ is the state vector of the LAM interpolated onto the global grid, and $\alpha \leq 1$ is the blending coefficient. At locations where no limited area model information is available, the related components of $\mathbf{x}_\ell$ are zero. In the cost function, the term described by Eq. (2.3) is replaced by

$$\mathbf{y}^{\mathrm{o}}(t_j) - \mathbf{h}_e\left[\mathbf{x}_e(t_j)\right] = \delta_e\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_e(t_j)\mathbf{M}_e(t_0, t_j)\delta\mathbf{x}_e, \tag{2.7}$$

where

$$\delta_e \mathbf{y}^{\mathrm{o}}(t_j) = \mathbf{y}^{\mathrm{o}}(t_j) - \mathbf{h}_e\left[\mathbf{x}_e^b(t_j)\right], \tag{2.8}$$

and $\mathbf{H}_e(t_j)$ is the linearization of $\mathbf{h}_e\left[\mathbf{x}_e(t_j)\right]$ about $\mathbf{x}_e^b(t_j)$, that is,

$$\mathbf{h}_e\left[\mathbf{x}_e(t_j)\right] \approx \mathbf{h}_e\left[\mathbf{x}_e^b(t_j)\right] + \mathbf{H}_e(t_j)\delta\mathbf{x}_e(t_j), \tag{2.9}$$

where

$$\begin{aligned} \delta\mathbf{x}_e(t_j) &= \mathbf{x}_e(t_j) - \mathbf{x}_e^b(t_j) \\ &= (1-\alpha)\mathbf{L}\left[\mathbf{x}_g(t_j) - \mathbf{x}_g^b(t_j)\right] + \alpha\left[\mathbf{x}_\ell(t_j) - \mathbf{x}_\ell^b(t_j)\right]. \end{aligned} \tag{2.10}$$

Introducing the notation

$$\delta\mathbf{x}_\ell(t_j) = \mathbf{x}_\ell(t_j) - \mathbf{x}_\ell^b(t_j), \tag{2.11}$$

Eq. (2.10) can be rewritten as

$$\delta\mathbf{x}_e(t_j) = (1-\alpha)\mathbf{L}\delta\mathbf{x}_g(t_j) + \alpha\delta\mathbf{x}_\ell(t_j). \tag{2.12}$$

After decomposing $\mathbf{M}_e(t_0, t_j)$ into $\mathbf{M}_g(t_0, t_j)$ and $\mathbf{M}_\ell(t_0, t_j)$, the latter being the linearization of the limited area dynamics about the nonlinear trajectory $\mathbf{x}_\ell^b(t_j)$, Eq. (2.12) can be written as

$$\delta\mathbf{x}_e(t_j) = (1-\alpha)\mathbf{L}\mathbf{M}_g(t_0, t_j)\delta\mathbf{x}_g(t_0) + \alpha\mathbf{M}_\ell(t_0, t_j)\delta\mathbf{x}_\ell(t_0). \tag{2.13}$$

Because the right-hand side of eq. (2.13) is a linear mapping of $\delta\mathbf{x}_e(t_0)$, Eq. (2.13) can also be written as

$$\delta\mathbf{x}_e(t_j) = \mathbf{M}_e(t_0, t_j)\delta\mathbf{x}_e(t_0), \tag{2.14}$$

where $\mathbf{M}_e(t_0, t_j)$ is the operator that represents the linear mapping of the regionally enhanced state. REG 4D-Var can therefore be implemented by replacing $\mathbf{M}_g(t_0, t_j)\delta\mathbf{x}_g$ by the right-hand side of Eq. (2.13), and $\mathbf{H}_g$ by $\mathbf{H}_e$ in the cost function. The cost function becomes

$$\begin{aligned} \mathrm{J}\left[\delta\mathbf{x}_g(t_0)\right] = & \left[\delta\mathbf{x}_g(t_0)\right]^{\mathbf{T}} (\mathbf{Pb})^{-1}\delta\mathbf{x}_g(t_0) + \\ & \sum_{j=0}^{N} \left[\delta_e\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_e(t_j)\mathbf{M}_e(t_0, t_j)\delta\mathbf{x}_e(t_0)\right]^{\mathbf{T}} \times \\ & \mathbf{R}_{t_j}^{-1}\left[\delta_e\mathbf{y}^{\mathrm{o}}(t_j) + \mathbf{H}_e(t_j)\mathbf{M}_e(t_0, t_j)\delta\mathbf{x}_e(t_0)\right], \end{aligned} \tag{2.15}$$

where the observation term of the cost function depends on the control variable $\delta\mathbf{x}_g$, as $\delta\mathbf{x}_e(t_0)$ is a function of $\delta\mathbf{x}_g$ (Eq. 2.13). Once $\delta\mathbf{x}_g^a$ has been obtained by the minimization of the cost function, the global analysis $\mathbf{x}_g^a$ can be calculated using Eq. (2.1).

2.1.3 The Limited Area Analysis

There are a number of potential approaches to generate a LAM analysis $\mathbf{x}_\ell^a$ based on the information provided by the global analysis increment $\delta\mathbf{x}_g^a$. For instance, it could be computed by

$$\mathbf{x}_\ell^a = \mathbf{x}_\ell^b + \mathbf{A}(\delta\mathbf{x}_g^a) \tag{2.16}$$

that is, by adding the global analysis increment after its interpolation to the LAM grid $(\mathbf{A})$ to the LAM shorter forecast. In our study, we test the simplest possible approach of interpolating the global analysis $\mathbf{x}_g^a$ onto the LAM grid, that is,

$$\mathbf{x}_\ell^a = \mathbf{A}\left(\mathbf{x}_g^a\right) \tag{2.17}$$

The difference between our approach and the standard approach of the operational numerical weather prediction centers for the generation of LAM initial conditions, in which the

limited area analysis is obtained by an interpolation of their global analysis to the LAM grid, is that our global analysis uses some LAM forecast information in addition to the global forecast information. We note that the general advantages of obtaining a LAM analysis by interpolation of a global analysis are that

1. the limited area analysis of the large scale flow is properly informed about the large scale flow outside of the lateral boundary conditions, and

2. the limited area analysis can benefit from the assimilation of the satellite radiance observations, which require global estimates of the flow dependent parameters of the observation bias correction terms.

2.2 The Models

In our study, we use the global model NAVGEM (Navy Global Environmental Model) (Hogan et al., 1991) and the LAM COAMPS (Coupled Ocean/Atmosphere Mesoscale Prediction System) (Hodur, 1997). Both of these models are used for operational numerical weather prediction at the United States Navy. NAVGEM is a spectral model, which is operationally run at horizontal resolution T425. COAMPS is a finite-difference model, which is run at varying resolutions in many limited area domains across the globe. Given the extensive modeling framework and finite computational resources available to us, it is imperative to run these models at resolutions that are coarser than those used in operations. We run NAVGEM at a spectral horizontal resolution of T119, and COAMPS at 32km, using the default operational parameterization schemes.

2.3 Implementation on the Operational Forecast System of the U.S. Navy

In our implementation of REG 4D-Var, NAVGEM provides the global state $\mathbf{x}_g$ and COAMPS the local state $\mathbf{x}_\ell$ in three unconnected LAM domains (Fig. B.1). The background $\mathbf{x}_g^b$ is a 6-hour NAVGEM forecast. The short-term model forecasts that are blended

at the times and locations of the observations in the 6-hour observation window centered at the analysis time are based on 9-hour integrations of NAVGEM and COAMPS.

The interpolation $\mathbf{L}(\mathbf{x}_g)$ in Eq. 2.6 is done in two steps. In the first step, a T319 spectral representation of each NAVGEM field is created by assigning a zero value to each spectral coefficient associated with a wave number between T119 and the T319. In the second step, the new T319 field is transformed onto a higher (about 42-km) resolution Gaussian grid using an inverse spherical harmonic transformation.

The COAMPS fields must also undergo interpolation onto the global high-resolution Gaussian grid before blending. The first step of the horizontal interpolation process is to use a 9-point smoothing function,

$$F_{i,j} = \sum_{\substack{i'=i-1 \\ j'=j-1}}^{\substack{i+1 \\ j+1}} a_{i',j'} f_{i',j'} \tag{2.18}$$

where $a_{i',j'}$ is a distance-weighted function, whose role is to reduce aliasing in the horizontal interpolation. A nearest neighbor interpolation completes the horizontal interpolation of the COAMPS fields to the high-resolution Gaussian grid. Vertical interpolation is also required: COAMPS fields, which lie on height-based sigma surfaces, are linearly mapped onto the pressure-based sigma surfaces of NAVGEM, using a hydrostatic surface pressure correction to account for differences between the NAVGEM and COAMPS topography. COAMPS fields are not interpolated to NAVGEM grid points that fall below ground in COAMPS. This approach does not lead to artificial discontinuities in the analyzed fields, because the interpolated COAMPS fields are used only in the computation of the innovations, and there are no observations at altitudes that fall below the ground in reality. (The higher resolution COAMPS orography is more similar to the true orography than the lower resolution NAVGEM orography.)

With both NAVGEM and COAMPS interpolated to the high-resolution Gaussian grid, the blending is applied according to Eq. (2.6). To ensure a smooth transition of the fields at the lateral boundaries of the LAM domains, the blending coefficient α is tapered to zero linearly over the 15 grid points closest to each lateral boundary. Figure (B.1) shows the values of the blending coefficient for the case when its value in the interior of the LAM domains is $\alpha = 1$. To avoid introducing artificial vertical discontinuities near the top of the COAMPS model atmosphere, which is significantly lower than the top of the NAVGEM model atmosphere, a vertical mask is applied, linearly decreasing the weight of the COAMPS fields in the blended state between 500 mb and 100 mb (above which no COAMPS model information is used).

The global and LAM analysis is then calculated as outlined in section 2.1 using NAVDAS-AR, the operational 4D-Var system of the U.S. Navy. Similar to all other operational 4D-Var systems (e.g. Szunyogh, 2014), NAVDAS-AR uses a dual resolution to search for the minimizer of the cost function. In particular, the operationally implemented system integrates the TLM at horizontal resolution T119, which is much lower than the T425 resolution of the full nonlinear model integrations. In our experiments, the TLM is used at the T119 resolution of the full nonlinear model integrations.

3. EXPERIMENT DESIGN

3.1 The Analysis-Forecast Experiments

The goal of this project is to assess the impact of REG 4D-Var on the COAMPS forecasts. Our hypothesis is that by providing a global analysis that is regionally enhanced by higher-resolution LAM information, the analyzed global model state will be pushed toward the true model attractor, which will lead to improved short-term forecasts. Because these short-term global model forecasts are also used in the preparation of the LAM initial conditions, we expect them to also lead to improved LAM forecasts. We will investigate whether the results of the analysis-forecast experiments confirm or reject this expectation.

We carry out analysis-forecast experiments in five different configurations of the data assimilation system. Each experiment consists of 124 six-hour data assimilation cycles – from 0000 UTC 1 October, 2012 through 1800 UTC 31 October, 2012 – and a 72-hour COAMPS forecast is started from each 0000 UTC and 1200 UTC analysis. COAMPS analysis and forecast error statistics are computed over the resulting sample of sixty-two forecasts.

The five configurations of the data assimilation system are as follows:

Experiment 1: Control. This experiment is designed to resemble the present standard data assimilation approach in which no limited area model information is used in the preparation of the global analysis. The linearized model integrations are carried out at a reduced resolution, T47.

Experiment 2: 'Blend Skip' Control. This experiment is designed to yield a more realistic comparison for our blended experiments. In this experiment, the TLM integrations are carried out at the same T119 resolution as in the experiments that use blending, but no blending is performed ($\alpha = 0$). The results of this experiment are also affected by the

same interpolation errors as the blending experiments.

Experiment 3: 30% blend. In this experiment, the blending coefficient is $\alpha = 0.3$; that is, the weight of the COAMPS field in the interior of the LAM domains is 30%.

Experiment 4: 50% blend. The same as Experiment 3, except that $\alpha = 0.5$.

Experiment 5: 100% blend. The same as Experiment 3 and 4, except that $\alpha = 1.0$.

For every model cycle and experiment, COAMPS information is output at standard pressure levels with 12-hour increments (from analysis time to forecast lead time 72 hr.). The air temperature, geopotential height, zonal wind, and meridional wind forecasts are assessed at each vertical level and forecast lead time.

The three LAM domains (Fig. B.1) are selected to provide a strong representation of the diverse surface boundary conditions, orography, and flow patterns of the globe: one domain is placed in the Atlantic hurricane basin to assess the benefits that REG 4D-Var may bring to tropical cyclone forecasting, another is placed in the exit region of the extratropical Pacific storm track, while the last one is over Europe, a region of complex orography.

3.2 Verification Techniques

3.2.1 ECMWF Analyses

The first verification data set used to assess COAMPS forecast accuracy is comprised of global European Center for Medium-Range Weather Forecasts (ECMWF) model analyses. In this data set, verification data is available for all variables, vertical levels, and times output by COAMPS, with a 0.5x0.5 degree resolution. Given the resolution discrepancy between the native COAMPS output and the available ECMWF analyses, all COAMPS fields undergo a cubic spline interpolation to the resolution of the ECMWF analyses before the computation of the error statistics. A mask is applied after interpolation to mitigate any artificial interpolation errors along the borders of each domain. Finally, all zonal

(u) and meridional (v) wind fields undergo a transformation from the COAMPS Lambert conformal projection to the true, earth-relative coordinates.

The square error is computed for each grid point of each COAMPS forecast, then two types of RMSE averages are calculated. The first type, given by

$$\epsilon_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{ti}^f - \mathbf{x}_{ti}^v)^2}, \tag{3.1}$$

calculates the mean over the full sample of forecasts (T=62) at each grid point (i) and forecast lead time (t). In Eq. (3.1), $\mathbf{x}_t^f$ represents a COAMPS forecast field, and $\mathbf{x}_t^v$ is the ECMWF verification data for that field. The second type of RMSE average, given by

$$\epsilon = \sqrt{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\mathbf{x}_{ti}^f - \mathbf{x}_{ti}^v)^2}, \tag{3.2}$$

calculates the mean over both the forecasts and the N grid points of a verification domain.

The systematic error (bias) is also calculated for the COAMPS forecasts. Similar to the RMSE, it is obtained by averaging either for each grid point:

$$\epsilon_i = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_t^f - \mathbf{x}_t^v) \tag{3.3}$$

or for the entire forecast domain:

$$\epsilon = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\mathbf{x}_{ti}^f - \mathbf{x}_{ti}^v). \tag{3.4}$$

Standard Deviation is the third error diagnostic computed, which is calculated at each point by subtracting the square of the bias from the square of the RMSE (the MSE), then taking the square root of the result. Temporal and spatiotemporal averages are computed

in the same way as the RMSE and the bias.

3.2.2 RAOB Data

The second verification data set used is comprised of in-situ radiosonde observations (RAOB) within the COAMPS domains. Spatiotemporally averaged RMSE, bias, and standard deviation values are calculated the same way as in the verification against the ECMWF analyses. COAMPS fields are bilinearly interpolated to each RAOB location.

3.2.3 Composite Domain

In addition to calculating verification statistics for each individual COAMPS domain, a composite domain is also created to calculate verification statistics that measure the overall LAM performance over the three LAM domains. In the verification against ECMWF analyses, this aggregation is performed by assigning equal weight to the data at each grid point in the three domains. In the verification against RAOB, each composite error value is computed as a weighted average of the individual domain errors by

$$\epsilon_{composite} = \frac{1}{T} \sum_{t=1}^{T} \sum_{d=1}^{3} (w_{dt} \epsilon_{dt}), \tag{3.5}$$

where w_{dt} is the portion of the number of RAOBs that falls into domain d at verification time t, and ϵ_{dt} is the related error measure.

3.2.4 Statistical Significance

Two-tailed t tests, as outlined in Wilks (2011), are performed to compare the time series of the error measures based on spatial averaging for the blended experiments and the blend skip experiment. The temporal correlations between the errors are accounted for by computing the effective sample size n' as

$$n' = T \cdot \frac{1 - \rho}{1 + \rho}, \tag{3.6}$$

where ρ is the lag-1 autocorrelation coefficient. A test statistic (z) is then computed by

$$z = ((\epsilon_{blend} - \epsilon_{bskip}) - \mu_{blend-bskip})/(\frac{s_{blend-bskip}}{\sqrt{n'}}), \tag{3.7}$$

where $\epsilon_{blend} - \epsilon_{bskip}$ are the pair of RMSE for a blended experiment and blend skip experiment, respectively, and μ is the hypothesized population mean ($\mu = 0$ when testing against a null hypothesis), and $s_{blend-bskip}$ is the standard deviation of the difference $\epsilon_{blend} - \epsilon_{bskip}$ for the sample of forecasts. Finally, the test statistic is converted to a confidence interval Φ by,

$$\Phi = 1 - 2w(z), \tag{3.8}$$

where $w(z)$ is found from a table of left-tail cumulative probabilities for the standard Gaussian distribution [Table B.1 in (Wilks, 2011)]. The two-multiplier is added to account for the two-tailed probability. A blend experiment error is determined to be significantly different from the blend skip experiment error when $\Phi \geq 0.95$, which corresponds to a confidence interval of 95%.

4. RESULTS

The results will be presented in three sections. The first section will assess the performance of the blend experiments for four commonly scrutinized atmospheric fields: 500 mb geopotential height, 250 mb wind, 850 mb temperature, and 1000 mb wind. The second section will examine the vertical distribution of blending impacts, and the third section will examine the forecast performance of REG 4D-Var for active tropical cyclones.

4.1 Impact on Four Selected Fields

4.1.1 500 mb Geopotential Height

We first show results for the 500 mb geopotential height. Table (A.1) gives a comprehensive summary of model performance for this field. RMSE averaged over the composite domain is shown for all experiments and forecast lead times, with green and orange shading used to mark a statistically significant improvement and degradation, respectively, relative to the blend skip control analysis-forecast experiment. The most notable results shown in the table are the following. First, the original control experiment – executed at the operational TLM resolution – performs significantly worse than all other experiments. This result will also hold for the other investigated fields. The magnitude of improvement in the blend skip experiment compared to the analyses and forecasts of this control experiment is generally higher than the improvements in the blended experiments compared to the blend skip experiment. Second, the 50% blend yields the best model performance among all experiments. Statistically significant improvements are found in the 50% blend experiment for the 12 to 60 hour forecast lead times, when the ECMWF analyses are used for verification, and for the 24 and 60 hour forecast lead times, when the RAOB data are used for verification. Finally, errors for the blended experiments are initially higher than those of the blend skip experiment, but quickly become smaller as these forecasts progress.

This is true even when the analysis of a blended experiment is significantly less accurate than that of the blend skip experiment; for instance, for the 100% blend experiment when the verification is done against the ECMWF analyses. This trend is also evident in Figure (B.2), which is a graphical illustration of Table (A.1). In this figure, the results for the poorly performing control experiment are not shown, while the averaged RMSE for the 30, 50, and 100% experiments is plotted relative to the RMSE of the blend skip experiment. The left panel shows the verification results against the ECMWF analyses, while the right panel shows the verification results against RAOB. As previously discussed for Table (A.1), the 50% experiment shows the most consistent improvements among the blended experiments. In addition, the advantage of the blended experiments gradually increases as the lead time increases, though the verification results against RAOB show a drop in the improvement after 60 hours. Both verification approaches indicate a sharp reduction of the analysis accuracy when blending is increased to 100%.

Figure (B.3) is a map of the time-averaged RMSE relative to the time-averaged RMSE of the blend skip experiment. The analysis (top), and the 24 h (middle) and 48 h (bottom) lead time results are shown, for the verification against the ECMWF analyses. At analysis time, when the analyses of the 50% blend experiment have an overall higher error than the analyses of the blend skip experiment, considerable degradations occur across the Canadian Pacific coast and the northwestern edge of the CONUS domain. These error degradations almost completely disappear after the first 24 hours of model integration (see middle panel). RMSE reduction is found across the Pacific storm track along the southern coast of Alaska, as well as in the Atlantic hurricane basin and across much of the European domain. At a 48 hour lead time, the improvements become even stronger, covering broad areas across all three domains with only isolated pockets of weak degradations. Figure (B.4) is the same in format as Figure (B.3), but it shows results for the 100% rather than the 50% blend experiment. At analysis time (top), the patterns of im-

provements and degradations are very similar to those in the 50% blend experiment, but their magnitude is even larger: strong degradation is observed along the Pacific coast and western CONUS domain, and the improvements across Alaska and the Atlantic hurricane basin are also accentuated. In addition, the area covered by degradations is larger than the area of improvements.

4.1.2 250 mb Wind

Table (A.2) shows a summary of the RMSE for the composite 250 mb wind field, with the format of Table (A.1). Wind is broken down into u and v components for the ECMWF verification, while total wind magnitude is used for the RAOB verification. Similar to the 500 mb geopotential height field, RMSE for the blended experiments is initially higher than that for the blend skip experiment, especially for the 100% blend experiment, but becomes smaller with increasing forecast time. Another similarity to the results for the 500 mb geopotential height is that the 50% blend experiment is the best performing experiment. This experiment produced statistically significant improvement relative to the blend skip experiment for four of the seven investigated lead times for the RAOB verification. It also produced the lowest overall RMSE. Table (A.3) presents the same results as Table (A.2), but in a different format: it shows the forecast error reduction for the different experiments relative to the blend skip experiment, $\Delta\%$, as a percentage of the error for the blend skip experiment, that is,

$$\Delta\% = \frac{\epsilon_{blendskip} - \epsilon_{exp}}{\epsilon_{blendskip}} * 100 \tag{4.1}$$

The relatively large negative values for the control experiment indicate that the analyses and forecasts of the control experiment are much less accurate than those of the blend skip experiment. That is, running the TLM at resolution T119 rather than T47 greatly reduces the limited area analyses and forecast errors for the 250 mb wind. Compared to this large

error reduction, the magnitude of the forecast error reductions brought about by blending is small.

In comparison to the results for the 500 mb geopotential height field, the spatial distribution of the patterns of forecast improvements and degradations is much less structured. For example, Figure (B.5), which displays the spatial distribution of the forecast improvements and degradations for the 50% blend experiment at the 72 h lead time relative to the blend skip experiment, shows spotty areas of strong improvement, along with isolated areas of considerable degradation. While the spatiotemporally averaged RMSE improvement of the 250 mb zonal wind forecasts of the 50% blend experiment are statistically significant, this improvement comes from improvements at the smaller scales as opposed to the broad cohesive regions of improvement observed for the 500 mb geopotential height. This result is not unexpected, as the spatial variability of the 250 mb wind field is higher than the spatial variability of the 500 mb geopotential height field. (Recall that under the assumption of geostrophic balance, the two horizontal components of the wind vector are proportional to the first spatial derivatives of the geopotential height.)

4.1.3 850 mb Temperature

A summary of the verification results for the 850 mb temperature is shown in Table (A.4). In stark contrast to the results for both the 500 mb geopotential height and the 250 mb wind, the greatest improvements relative to the blend skip experiment are found at analysis time. In addition, these improvements diminish rather than increase with increasing forecast lead time. This behavior is found for the verification results against both the ECMWF analyses and the RAOB data. Statistically significant improvement is found only at analysis and the shortest forecast lead times. (The longest lasting statistically significant improvement is for 36 hours for the 50% blend experiment.) To explore the origins of this behavior, we further investigate the error statistics for the individual COAMPS domains

(Table A.5 and B.6). While there is a clear and broad positive impact of blending at analysis time, especially in the Northeast Pacific domain, the positive forecast impact quickly turns into a generally neutral impact in all model domains. This process is the fastest in the European domain and the slowest in the Northeast Pacific domain. Interestingly, the positive forecast impact in the latter domain lasts the longest for the 100% blend experiment, the experiment whose forecast performance was the poorest among the blended experiments for the other forecast variables investigated so far. The profound qualitative differences between the forecast effects of REG 4D-Var for the temperature, and for the geopotential height and the wind, suggest that they are the results of two different mechanisms that are both affected by the blending approach.

4.1.4 1000 mb Wind

The behavior of the verification statistics for the 1000 mb wind (Table A.6, and Figures B.7, B.8, and B.9) is very similar to that for 850 mb temperature: the large initial advantage of the blended experiments over the blend skip experiment quickly dissipates as forecast time increases, despite the initial improvements being spread more broadly across all three domains. Forecast improvements in the 100% blend experiment decreased in a shorter time than in the 30% blend and 50% blend experiment.

4.2 Vertical Structure of the Forecast Impact

Of the four atmospheric state variables for which we have discussed the verification results, two (850 mb temperature and 1000 mb wind) are lower tropospheric variables, one (500 mb height) is a mid-tropospheric variable, and one (250 mb wind) is an upper tropospheric variable. For the two lower tropospheric variables, forecast improvements due to blending were observed at analysis and the short forecast times. For the other two variables, improvements were observed at later times. As it turns out, this behavior of the verification statistics is part of a general trend: improvements occur at the low atmo-

spheric levels at the analysis and early lead times, propagating to the higher atmospheric levels as lead time increases. We observed this phenomena for all variables for which forecast skill was assessed, but it was most pronounced for temperature and wind. Tables (A.7) and (A.8) provide illustration of this behavior for air temperature. At analysis time (Table A.7), the vast majority of the statistically significant improvements is found at the low levels, while some statistically significant degradation, especially for the verification against ECMWF analyses, is found at the upper levels. In contrast, at the 60 h forecast lead time (Table A.8) fewer statistically significant results are found, but they are all significant improvements at 200 mb. The signal is even stronger for the composite average wind fields at the analysis time (Table A.9). Broad statistically significant improvements are found for the 30% and 50% blend experiment from 1000 mb through 700 mb, while significant degradations occur immediately above that layer, especially for the verification against the ECMWF analyses; by the 60 h forecast lead time (Table A.10), all significant improvement shift to the higher atmospheric levels, with no significant improvement below 500 mb.

The smaller analysis improvements above 500 mb are not completely unexpected, because as explained earlier, the weight of the COAMPS blended fields used in REG DA decreases linearly from 500 mb to the top of the COAMPS model atmosphere. Yet, it is surprising to see the very small, but statistically significant analysis degradations above 500 mb (Tables A.7 and A.9). To help understand where and why this analysis degradation occurs, Figure (B.10) shows a decomposition of the 200 mb mean square temperature analysis error differences for the 100% blend experiment into two components: square bias and variance. This decomposition can help distinguish between the contributions of the systemic and transient errors to the MSE, as $MSE = Variance + Bias^2$. Examining the spatial distribution of MSE (top), broad areas of error degradation are found across the Northeast Pacific domain, with a particularly high degradation over the complex terrain

along the Pacific Rim. The MSE changes are quite neutral across the CONUS domain, with a band of small improvement stretching over the Rocky Mountains, a result also seen above mountainous European terrain. Degradations in the European domain are concentrated near bodies of water, including the Eastern Atlantic Ocean, and the Mediterranean, Black, and Caspian Seas. Improvements and degradations in the squared bias closely match the pattern and magnitude of MSE. The distribution of the variance fluctuations, on the other hand, does not have quite as clear of a relation to the MSE distribution as the square bias distribution. While the values are broadly negative, indicating degradations, and certainly contribute to error degradation in areas such as the Alaskan Panhandle, the connection between the two distributions is not striking. The results of the decomposition suggest that the differences between the accuracy of the analysis fields are predominantly due to differences in the bias. Interestingly, blending reduces the bias over topography, while blending increases the bias over the waters of the Northeast Pacific and European domains. A further examination of the bias (Figure B.11) suggests that the increase of bias above water is due to an increase of a warm bias present in both experiments, while the bias reduction over topography is due to reducing a warm bias in the blend skip.

Figure (B.12) shows the same error decomposition as Figure (B.10) but for the 60 h forecast lead time (note that the scale in the color bar is doubled compared to the range in the analysis decomposition). The MSE difference distribution (top) now has little relation to physical geography. The square bias differences (middle) are much smaller in magnitude compared to the MSE differences, and are no longer predisposed to occur over bodies of water. The changes in the MSE are clearly driven by changes in the variance of the error. Thus the results suggest that the early degradations due to blending are replaced by improvements at the later forecast times, as reductions in the magnitude of the transient errors start to dominate over the changes in the systematic errors.

4.3 Forecasts Effects in the Atlantic Hurricane Basin

One aspect of model performance that is very important to the operational community is the performance predicting tropical cyclones. To assess this performance, our attention will focus on the CONUS domain, where hurricanes Sandy and Rafael passed through the western Atlantic during the time-frame of our experiments.

4.3.1 1000 mb Geopotential Height

Figure (B.13) shows the 1000 mb geopotential height RMSE at analysis time. The top-left panel shows the RMSE for the blend skip control, while the remaining three plots show the RMSE relative to the blend skip RMSE for the 30, 50, and 100% blend experiment, respectively. In examining the first plot, two significant sources of error are apparent. The first, over the high terrain of the Rocky Mountains, is likely due to differences between the ways the post-processing of the COAMPS and ECMWF forecast systems extrapolate the analysis fields to obtain the 1000 mb geopotential height field when it lies beneath the surface of the earth. The second, and more important area, is in the Atlantic basin. Two separate error tracks can clearly be seen, where hurricane intensity and/or location analyses differed from the ECMWF analyses. The three panels that show the error for the blended experiments relative to the error for the blend skip experiment indicate stark analysis error reductions along the path of Hurricane Sandy. This error reduction is most pronounced in the area where Sandy made landfall, near Delaware Bay. In that area, strong analysis improvement (up to 25 meters in the time averaged RMSE map) is observed for all three blended experiments. Interestingly, a higher blending weight leads to a larger reduction of the spatiotemporally averaged RMSE, with the 100% blend showing the most robust improvement in the analysis of Hurricane Sandy. However, the magnitude of the degradations elsewhere also increases as the blending weight increases, which has a potentially negative effect on forecast accuracy at later lead times.

Figure (B.14) shows the forecast improvements at the 12 h lead time in the format of Figure (B.13). Broadly similar patterns are seen in the blend skip RMSE as at analysis time, with an overall slightly increasing RMSE. The magnitude of the improvements due to blending decreases considerably, though strong areas of improvement remain along the path of Hurricane Sandy. The 30% blend experiment shows the largest improvements and the smallest degradations, but the 50% blend experiment is the only experiment for which the spatiotemporal RMSE average (not shown) shows statistically significant improvements over the forecasts of the blend skip experiment. The forecasts of the 100% blend experiment are improved the most for Hurricane Rafael (northeast corner of the domain), but are considerably degraded in southern Texas and near the mid-Atlantic coast. Figure (B.15) shows results for 48 h forecast lead time in the format of Figures (B.13 and B.14). While RMSE for the blend skip experiment is considerably higher than for the shorter lead times, the hurricane error tracks are still visible. The region of strongest improvement is further to the south than before, indicating that the two-day forecasts from the blended analyses more accurately predict Hurricane Sandy's evolution northeastward away from the Bahamas before its turn to the northwest. The 100% blend continues to have the strongest improvement along Hurricane Sandy's path, but suffers from stronger regions of degradation for Hurricane Rafael, as well as broader degradations along the eastern seaboard. No experiment has a statistically significant improvement after averaging across across both time and space.

Figure (B.16) shows the verification results for forecast lead time 72 h. The magnitude of the blend skip RMSE for the blend skip experiment is higher than before, but the error patterns remain the same, with higher errors in the northwestern domain and around the tropical cyclone tracks in the Atlantic. RMSE reduction due to blending is now apparent for all three blending experiments along Hurricane Sandy's path. As was the case at the shorter lead times, the 100% blend yields the strongest, most pronounced area of

improvement for Hurricane Sandy, but has the lowest accuracy for Hurricane Rafael along the northeast corner of the domain, and also may have predicted a landfall near South Carolina that did not verify. The 50% blend, although anomalously falling as the worst experiment of the three blends, shows strong improvement in geopotential height error along Hurricane Sandy's path, including a region off the Northeast coast. Finally, the 30% blend leads to the least degradation across the CONUS, while it retains a strong improvement in the 3-day forecast of Hurricane Sandy.

4.3.2 1000 mb Wind

The near surface wind is another important field to analyze when examining tropical cyclone performance. Figure (B.17) shows RMSE for the blend skip control experiment (left) and the blend skip relative RMSE for the 50% blend (right) at the analysis time (top), at the 36 h forecast lead time (middle), and at the 72 h forecast lead time (bottom). At analysis time, the magnitude of the RMSE is substantial for the blend skip control experiment in areas of elevated topography, but as for the geopotential height, this is not a cause for operational concern, as the 1000 mb surface is beneath the ground. A smaller but more intriguing area of elevated RMSE appears along the paths of Hurricanes Sandy and Rafael. For the 50% blend experiment, areas of broad improvement are found across the Atlantic Ocean, in particular the region off the eastern Florida coast where the RMSE is high for the blend skip experiment. Smaller magnitude improvements are also observed over the continental U.S., although isolated small degradations also occur.

At the 36 h forecast lead time, higher RMSE remains across the Great Lakes, and along the two tropical cyclone storm tracks in the Atlantic. Interestingly, the forecasts of the blended experiment show very inconsequential improvement across the domain, with an overall small, non-significant degradation occurring when RMSE is spatiotemporally averaged.

At the longest forecast lead time (bottom), RMSE increases considerably, particularly over water where the wind magnitude is higher. A clear pattern of higher RMSE along the paths of Sandy and Rafael remain evident. Highly pronounced RMSE improvement due to blending is found along Hurricane Sandy's track northeast of the Bahamas – indicating that the blended experiment predicts the tropical system with greater accuracy than the blend skip experiment. Blending has little effect on the accuracy of the wind forecasts for Hurricane Rafael. Outside the Western Atlantic, the changes in RMSE are very minor, and the spatiotemporal average of the RMSE reduction is not significant.

5. CONCLUSIONS

REG 4D-Var, and REG DA in general, is an operationally feasible method for the integration of the global and limited area data assimilation processes. We implemented the approach on the operational numerical weather prediction system of the U.S. Navy, and carried out a series of analysis-forecast experiments to find a near optimal configuration of the resulting REG 4D-Var system. The assessment of the quality of the global model forecasts from these experiments was the subject of an earlier study (Herrera, 2016). The focus of the present thesis was on the investigation of the effect of REG 4D-Var on the LAM forecast performance.

Our results suggest that REG 4D-Var has a positive impact on the limited area forecast accuracy. While the spatiotemporally averaged improvements are generally small, less than 1%, they are statistically significant for a several forecast variables. Isolated statistically significant short-lived error degradations were also observed, predominantly in the upper troposphere at the analysis time. In addition, large, statistically significant analysis improvements were found for Hurricane Sandy. The magnitude of these improvements decreased between lead times 12 h and 36 h, but increased again from the 48 h lead time.

REFERENCES

Herrera, M. A., 2016: Techniques for the improvement of numerical weather prediction: Investigating the dynamics of forecast uncertainty and the implementation of regionally enhanced global data assimilation. Ph.D. thesis, Texas A&M University.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **125 (7)**, 1414–1430.

Hogan, T., T. Rosmond, and R. Gelaro, 1991: The NOGAPS forecast model: A technical description. **15**, 218.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D*, **230**, 112 – 126.

Kretschmer, M., B. R. Hunt, E. Ott, C. H. Bishop, S. Rainwater, and I. Szunyogh, 2015: A composite state method for ensemble data assimilation with multiple limited-area models. *Tellus A*, **67**, 26 495.

Szunyogh, I., 2014: *Applicable Atmospheric Dynamics*. 1st ed., World Scientific.

Wilks, D., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed., Elsevier Academic Press.

Yoon, Y., B. Hunt, E. Ott, and I. Szunyogh, 2012: Simultaneous global and limited-area ensemble data assimilation using joint states. *Tellus A*, **64**, 18 407.

APPENDIX A

TABLES

Composite 500 mb Geopotential Height RMSE (m)	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control vs ECMWF	**9.841**	**12.077**	**15.504**	**20.184**	**25.579**	31.620	**38.378**
Control vs RAOB	**11.253**	**13.327**	**16.138**	**19.933**	**23.947**	27.980	**31.734**
Blend Skip vs ECMWF	9.270	11.397	14.947	19.475	24.900	30.773	36.994
Blend Skip vs RAOB	10.896	12.753	15.598	19.404	23.272	27.433	30.912
30% Blend vs ECMWF	9.287	**11.355**	**14.855**	**19.362**	**24.737**	30.610	**36.764**
30% Blend vs RAOB	10.933	12.718	**15.520**	19.327	23.190	27.296	30.889
50% Blend vs ECMWF	9.343	**11.324**	**14.815**	**19.328**	**24.689**	**30.584**	36.793
50% Blend vs RAOB	10.953	12.679	**15.483**	19.302	23.131	**27.219**	30.864
100% Blend vs ECMWF	**9.657**	11.384	14.844	19.394	24.747	30.643	36.864
100% Blend vs RAOB	11.096	12.655	**15.463**	19.301	23.167	27.301	30.931

Table A.1: Spatiotemporally averaged RMSE (m) of the 500 mb geopotential height forecasts for all five experiments and all forecast lead times. Bold text and shading indicates the difference between the experiment and the blend skip error is statistically significant. Green shading indicates a statistically significant error improvement, while orange shading indicates a statistically significant error degradation.

Composite 250 mb Wind RMSE (ms^{-1})	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control U-Wind vs ECMWF	**3.091**	**3.961**	**4.902**	**5.832**	**6.795**	**7.768**	**8.855**
Control V-Wind vs ECMWF	**2.832**	**3.786**	**4.821**	**5.850**	**6.925**	**7.956**	**9.170**
Control Wind vs RAOB	**3.538**	**4.232**	**5.042**	**5.898**	**6.771**	7.609	**8.544**
Blend Skip U-Wind vs ECMWF	2.868	3.766	4.703	5.639	6.574	7.603	8.640
Blend Skip V-Wind vs ECMWF	2.551	3.550	4.604	5.593	6.621	7.660	8.864
Blend Skip Wind vs RAOB	3.189	4.008	4.882	5.717	6.560	7.474	8.334
30% Blend U-Wind vs ECMWF	2.868	3.766	4.704	5.639	6.562	7.587	**8.604**
30% Blend V-Wind vs ECMWF	2.552	3.541	4.597	5.583	**6.579**	**7.627**	8.833
30% Blend Wind vs RAOB	3.182	3.995	4.868	5.710	6.525	**7.421**	8.303
50% Blend U-Wind vs ECMWF	2.872	3.771	4.705	5.641	6.561	7.578	**8.589**
50% Blend V-Wind vs ECMWF	2.556	3.544	4.598	5.584	**6.576**	7.622	8.840
50% Blend Wind vs RAOB	3.179	3.992	**4.856**	5.710	**6.516**	**7.407**	**8.274**
100% Blend U-Wind vs ECMWF	**2.886**	**3.788**	4.715	5.644	6.553	7.573	**8.580**
100% Blend V-Wind vs ECMWF	**2.572**	3.560	4.605	5.587	6.568	7.622	8.828
100% Blend Wind vs RAOB	3.176	3.995	**4.842**	5.718	6.523	**7.404**	8.289

Table A.2: Same as Table (A.1) but for 250 mb wind. RMSE units are ms^{-1}.

Composite 250mb Wind % Relative RMSE	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control U-Wind vs ECMWF	**-7.77%**	**-5.17%**	**-4.22%**	**-3.43%**	**-3.37%**	**-2.17%**	**-2.48%**
Control V-Wind vs ECMWF	**-11.04%**	**-6.64%**	**-4.72%**	**-4.60%**	**-4.60%**	**-3.87%**	**-3.45%**
Control Wind vs RAOB	**-10.96%**	**-5.61%**	**-3.29%**	**-3.17%**	**-3.21%**	-1.82%	**-2.52%**
30% Blend U-Wind vs ECMWF	-0.02%	-0.01%	-0.02%	0.00%	0.18%	0.21%	**0.41%**
30% Blend V-Wind vs ECMWF	-0.05%	0.25%	0.13%	0.18%	**0.63%**	**0.43%**	0.35%
30% Blend Wind vs RAOB	0.21%	0.31%	0.28%	0.12%	0.54%	**0.70%**	0.38%
50% Blend U-Wind vs ECMWF	-0.14%	-0.14%	-0.04%	-0.03%	0.19%	0.32%	**0.59%**
50% Blend V-Wind vs ECMWF	-0.21%	0.16%	0.12%	0.16%	**0.67%**	0.50%	0.27%
50% Blend Wind vs RAOB	0.31%	0.39%	**0.53%**	0.12%	**0.68%**	**0.89%**	**0.73%**
100% Blend U-Wind vs ECMWF	**-0.65%**	**-0.57%**	-0.25%	-0.09%	0.32%	0.39%	**0.70%**
100% Blend V-Wind vs ECMWF	**-0.82%**	-0.28%	-0.04%	0.10%	0.79%	0.49%	0.40%
100% Blend Wind vs RAOB	0.39%	0.30%	**0.81%**	-0.02%	0.56%	**0.93%**	0.55%

Table A.3: Same as Table (A.2) but values shown are the % change of RMSE between the given experiment and the blend skip experiment. Positive values indicate improvement over the blend skip, negative values indicate degradation.

Composite 850 mb Temperature RMSE (K)	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control vs ECMWF	1.328	1.486	1.675	1.895	2.109	2.316	2.529
Control vs RAOB	1.386	1.521	1.721	1.973	2.210	2.462	2.698
Blend Skip vs ECMWF	1.277	1.437	1.630	1.853	2.073	2.282	2.485
Blend Skip vs RAOB	1.307	1.481	1.688	1.933	2.173	2.413	2.640
30% Blend vs ECMWF	1.248	1.425	1.625	1.850	2.071	2.279	2.481
30% Blend vs RAOB	1.289	1.474	1.684	1.936	2.174	2.412	2.633
50% Blend vs ECMWF	1.235	1.421	1.623	1.847	2.069	2.279	2.480
50% Blend vs RAOB	1.279	1.470	1.682	1.937	2.174	2.410	2.630
100% Blend vs ECMWF	1.233	1.426	1.629	1.850	2.069	2.281	2.480
100% Blend vs RAOB	1.279	1.472	1.687	1.943	2.177	2.418	2.641

Table A.4: Same as Table (A.1) but for 850 mb temperature. RMSE units are K.

CONUS 850 mb Temperature RMSE (K)	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control vs ECMWF	1.198	1.499	1.680	1.838	1.988	2.139	2.276
Control vs RAOB	1.460	1.636	1.827	2.035	2.232	2.430	2.594
Blend Skip vs ECMWF	1.127	1.449	1.638	1.802	1.948	2.106	2.242
Blend Skip vs RAOB	1.365	1.587	1.798	1.997	2.180	2.372	2.544
30% Blend vs ECMWF	1.109	1.441	1.638	1.801	1.946	2.097	2.225
30% Blend vs RAOB	1.336	1.576	1.787	1.998	2.179	2.370	2.520
50% Blend vs ECMWF	1.103	1.438	1.642	1.803	1.947	2.100	2.230
50% Blend vs RAOB	1.324	1.569	1.788	2.000	2.181	2.373	2.519
100% Blend vs ECMWF	1.125	1.451	1.653	1.809	1.953	2.102	2.236
100% Blend vs RAOB	1.331	1.573	1.787	2.001	2.186	2.388	2.525

Europe 850 mb Temperature RMSE (K)	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control vs ECMWF	1.141	1.272	1.530	1.829	2.105	2.368	2.624
Control vs RAOB	1.275	1.352	1.566	1.875	2.169	2.475	2.773
Blend Skip vs ECMWF	1.116	1.238	1.498	1.792	2.075	2.331	2.572
Blend Skip vs RAOB	1.213	1.320	1.531	1.831	2.135	2.425	2.703
30% Blend vs ECMWF	1.106	1.237	1.498	1.793	2.076	2.332	2.574
30% Blend vs RAOB	1.209	1.321	1.533	1.835	2.137	2.426	2.706
50% Blend vs ECMWF	1.099	1.238	1.497	1.790	2.074	2.331	2.577
50% Blend vs RAOB	1.203	1.320	1.531	1.838	2.140	2.421	2.704
100% Blend vs ECMWF	1.106	1.261	1.517	1.803	2.081	2.340	2.579
100% Blend vs RAOB	1.198	1.329	1.545	1.851	2.144	2.430	2.719

NE Pacific 850 mb Temperature RMSE (K)	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control vs ECMWF	1.596	1.698	1.829	2.010	2.197	2.375	2.587
Control vs RAOB	1.666	1.940	2.109	2.227	2.326	2.494	2.656
Blend Skip vs ECMWF	1.532	1.634	1.768	1.958	2.156	2.344	2.546
Blend Skip vs RAOB	1.562	1.897	2.075	2.209	2.318	2.474	2.625
30% Blend vs ECMWF	1.480	1.611	1.755	1.950	2.150	2.342	2.541
30% Blend vs RAOB	1.513	1.864	2.058	2.204	2.317	2.470	2.624
50% Blend vs ECMWF	1.456	1.599	1.747	1.943	2.147	2.338	2.533
50% Blend vs RAOB	1.489	1.850	2.051	2.196	2.308	2.460	2.608
100% Blend vs ECMWF	1.435	1.583	1.736	1.934	2.133	2.334	2.527
100% Blend vs RAOB	1.494	1.817	2.039	2.183	2.299	2.451	2.616

Table A.5: Same as Table (A.4) but spatiotemporal RMSE averages are taken over each individual COAMPS domain.

Composite 1000 mb Wind RMSE (ms^{-1})	Analysis	Tau 12	Tau 24	Tau 36	Tau 48	Tau 60	Tau 72
Control U-Wind vs ECMWF	**2.574**	**2.217**	**2.449**	**2.773**	**3.057**	**3.335**	**3.619**
Control V-Wind vs ECMWF	**2.559**	**2.125**	**2.373**	**2.660**	**2.938**	**3.227**	**3.539**
Control Wind vs RAOB	**2.750**	2.348	2.415	2.534	2.640	2.776	2.865
Blend Skip U-Wind vs ECMWF	2.546	2.139	2.393	2.707	2.997	3.258	3.546
Blend Skip V-Wind vs ECMWF	2.528	2.050	2.309	2.590	2.879	3.172	3.476
Blend Skip Wind vs RAOB	2.662	2.334	2.396	2.522	2.608	2.748	2.844
30% Blend U-Wind vs ECMWF	**2.452**	**2.124**	**2.388**	2.706	2.996	3.251	3.534
30% Blend V-Wind vs ECMWF	**2.426**	**2.038**	2.305	2.588	2.873	3.169	3.471
30% Blend Wind vs RAOB	**2.558**	2.333	2.397	2.532	**2.623**	2.739	2.839
50% Blend U-Wind vs ECMWF	**2.412**	**2.119**	**2.386**	2.705	3.000	3.249	3.527
50% Blend V-Wind vs ECMWF	**2.384**	**2.033**	2.306	2.589	2.880	3.174	3.473
50% Blend Wind vs RAOB	**2.519**	2.330	2.399	2.529	2.623	2.748	2.857
100% Blend U-Wind vs ECMWF	**2.451**	**2.121**	2.392	2.714	3.009	3.259	3.528
100% Blend V-Wind vs ECMWF	**2.401**	**2.041**	2.316	2.600	2.882	3.182	3.476
100% Blend Wind vs RAOB	**2.521**	2.338	2.409	**2.551**	2.631	2.756	2.848

Table A.6: Same as Table (A.1) but for 1000 mb wind. RMSE units are ms^{-1}.

Composite 0hr Temperature RMSE (K)	Control	Blend Skip	30% Blend	50% Blend	100% Blend
200 mb vs ECMWF	**1.072**	1.036	**1.044**	**1.051**	**1.076**
200 mb vs RAOB	**1.233**	1.182	1.182	1.183	**1.190**
250 mb vs ECMWF	**0.896**	0.809	**0.812**	**0.816**	**0.834**
250 mb vs RAOB	**1.104**	1.028	**1.025**	**1.024**	1.028
300 mb vs ECMWF	**0.780**	0.707	0.707	0.708	**0.719**
300 mb vs RAOB	**0.972**	0.907	**0.905**	**0.904**	0.904
500 mb vs ECMWF	**0.793**	0.729	0.727	0.728	**0.743**
500 mb vs RAOB	**0.871**	0.800	0.799	0.800	**0.808**
700 mb vs ECMWF	**0.876**	0.839	**0.832**	**0.830**	0.845
700 mb vs RAOB	**0.999**	0.921	**0.914**	**0.909**	0.915
850 mb vs ECMWF	**1.328**	1.277	**1.248**	**1.235**	**1.233**
850 mb vs RAOB	**1.386**	1.307	**1.289**	**1.279**	**1.279**
925 mb vs ECMWF	1.498	1.481	1.462	**1.438**	**1.414**
925 mb vs RAOB	**1.563**	1.533	1.524	1.498	**1.470**
1000 mb vs ECMWF	1.599	1.742	1.657	**1.597**	**1.583**
1000 mb vs RAOB	1.953	2.168	2.106	2.022	**1.926**

Table A.7: Spatiotemporally averaged RMSE (K) of air temperature at the analysis time for all experiments and isobaric levels. Bold text and shading are used to indicate the value is statistically significant relative to the blend skip experiment. Green shading indicates a statistically significant improvement, while orange shading indicates a statistically significant degradation.

Composite 60hr Temperature RMSE (K)	Control	Blend Skip	30% Blend	50% Blend	100% Blend
200mb vs ECMWF	2.428	2.383	2.373	2.370	2.369
200mb vs RAOB	2.469	2.423	2.412	2.409	2.406
250mb vs ECMWF	2.206	2.182	2.174	2.175	2.176
250mb vs RAOB	2.157	2.124	2.122	2.122	2.122
300mb vs ECMWF	1.915	1.864	1.860	1.858	1.858
300mb vs RAOB	1.835	1.790	1.782	1.782	1.789
500mb vs ECMWF	1.924	1.867	1.862	1.862	1.864
500mb vs RAOB	1.857	1.788	1.782	1.781	1.786
700mb vs ECMWF	1.843	1.800	1.796	1.794	1.802
700mb vs RAOB	1.931	1.888	1.881	1.877	1.887
850mb vs ECMWF	2.316	2.282	2.279	2.279	2.281
850mb vs RAOB	2.462	2.413	2.412	2.410	2.418
925mb vs ECMWF	2.516	2.490	2.486	2.486	2.485
925mb vs RAOB	2.740	2.723	2.715	2.716	2.716
1000mb vs ECMWF	2.704	2.682	2.680	2.679	2.677
1000mb vs RAOB	2.798	2.808	2.802	2.795	2.792

Table A.8: Same as Table (A.7) but for a 60-hour forecast lead time.

Composite 0hr Wind RMSE (ms^{-1})	Control	Blend Skip	30% Blend	50% Blend	100% Blend
200 mb U-Wind vs ECMWF	2.620	2.479	2.479	2.481	2.488
200 mb V-Wind vs ECMWF	2.507	2.352	2.351	2.351	2.357
200 mb Wind vs RAOB	3.206	3.006	3.005	3.006	3.006
250 mb U-Wind vs ECMWF	3.091	2.868	2.868	2.872	2.886
250 mb V-Wind vs ECMWF	2.832	2.551	2.552	2.556	2.572
250 mb Wind vs RAOB	3.538	3.189	3.182	3.179	3.176
300 mb U-Wind vs ECMWF	3.441	3.174	3.182	3.189	3.216
300 mb V-Wind vs ECMWF	3.126	2.775	2.781	2.786	2.809
300 mb Wind vs RAOB	3.794	3.368	3.365	3.364	3.369
500 mb U-Wind vs ECMWF	2.691	2.493	2.497	2.503	2.542
500 mb V-Wind vs ECMWF	2.501	2.267	2.265	2.269	2.303
500 mb Wind vs RAOB	2.909	2.608	2.608	2.611	2.629
700 mb U-Wind vs ECMWF	2.358	2.214	2.205	2.203	2.232
700 mb V-Wind vs ECMWF	2.245	2.089	2.073	2.068	2.091
700 mb Wind vs RAOB	2.876	2.617	2.605	2.601	2.602
850 mb U-Wind vs ECMWF	2.411	2.319	2.294	2.281	2.287
850 mb V-Wind vs ECMWF	2.274	2.143	2.113	2.101	2.113
850 mb Wind vs RAOB	3.158	2.911	2.882	2.869	2.855
925 mb U-Wind vs ECMWF	2.481	2.388	2.336	2.320	2.370
925 mb V-Wind vs ECMWF	2.367	2.249	2.197	2.183	2.247
925 mb Wind vs RAOB	3.077	2.791	2.704	2.667	2.652
1000 mb U-Wind vs ECMWF	2.574	2.546	2.452	2.412	2.451
1000 mb V-Wind vs ECMWF	2.559	2.528	2.426	2.384	2.401
1000 mb Wind vs RAOB	2.750	2.662	2.558	2.519	2.521

Table A.9: Same as Table (A.7) but for wind. RMSE units are ms^{-1}.

Composite 60hr Wind RMSE (ms^{-1})	Control	Blend Skip	30% Blend	50% Blend	100% Blend
200 mb U-Wind vs ECMWF	**6.337**	6.212	**6.191**	6.196	6.194
200 mb V-Wind vs ECMWF	**6.649**	6.442	**6.410**	6.422	6.423
200 mb Wind vs RAOB	6.315	6.245	**6.212**	**6.211**	**6.190**
250 mb U-Wind vs ECMWF	**7.768**	7.603	7.587	7.578	7.573
250 mb V-Wind vs ECMWF	**7.956**	7.660	**7.627**	7.622	7.622
250 mb Wind vs RAOB	7.609	7.474	**7.421**	**7.407**	**7.404**
300 mb U-Wind vs ECMWF	**8.107**	7.933	**7.902**	**7.884**	7.892
300 mb V-Wind vs ECMWF	**8.123**	7.823	**7.784**	**7.775**	7.776
300 mb Wind vs RAOB	7.736	7.585	7.554	**7.538**	7.570
500 mb U-Wind vs ECMWF	**5.440**	5.303	**5.279**	5.277	5.279
500 mb V-Wind vs ECMWF	**5.364**	5.175	**5.156**	5.159	5.168
500 mb Wind vs RAOB	**5.251**	5.112	5.089	5.091	5.101
700 mb U-Wind vs ECMWF	**4.336**	4.231	4.217	4.215	4.220
700 mb V-Wind vs ECMWF	**4.274**	4.196	4.183	4.189	4.198
700 mb Wind vs RAOB	**4.380**	4.306	4.293	4.308	4.305
850 mb U-Wind vs ECMWF	**4.248**	4.141	4.130	4.127	4.132
850 mb V-Wind vs ECMWF	**4.180**	4.114	4.103	4.112	4.122
850 mb Wind vs RAOB	4.341	4.268	4.255	4.258	4.241
925 mb U-Wind vs ECMWF	**4.306**	4.188	4.181	4.173	4.183
925 mb V-Wind vs ECMWF	**4.242**	4.151	4.142	4.153	4.160
925 mb Wind vs RAOB	4.144	4.098	4.083	4.077	4.063
1000 mb U-Wind vs ECMWF	**3.335**	3.258	3.251	3.249	3.259
1000 mb V-Wind vs ECMWF	**3.227**	3.172	3.169	3.174	3.182
1000 mb Wind vs RAOB	2.776	2.748	2.739	2.748	2.756

Table A.10: Same as Table (A.7) but for wind at a 60-hour forecast lead time. RMSE units are ms^{-1}.

APPENDIX B

FIGURES

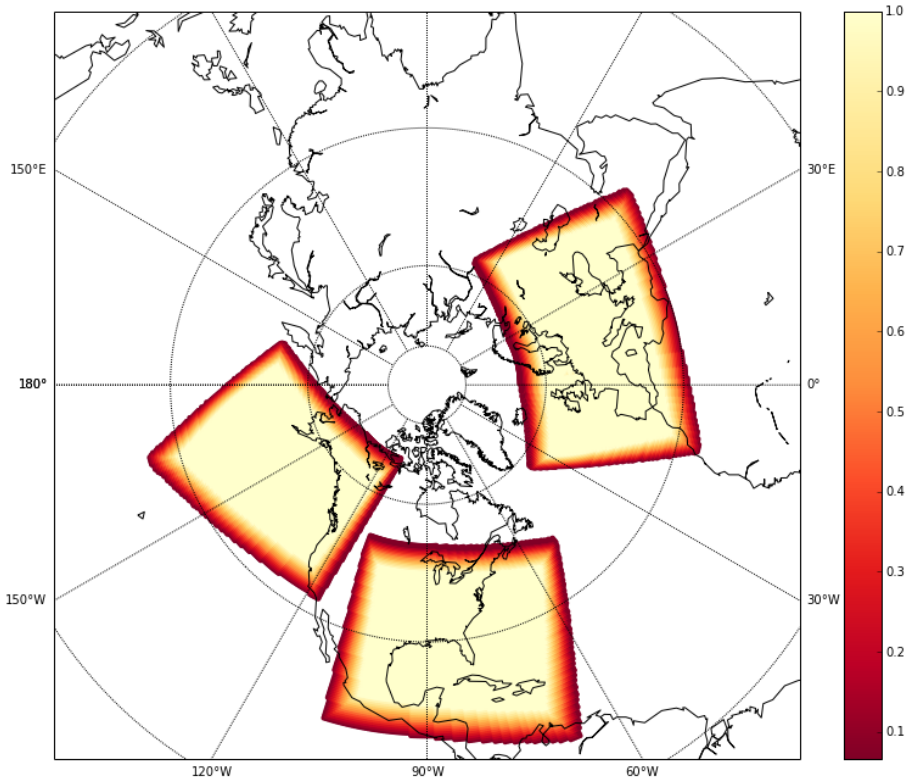

Figure B.1: Illustration of the three COAMPS domains. Color shades show the blending coefficients for the 100% blend experiment.

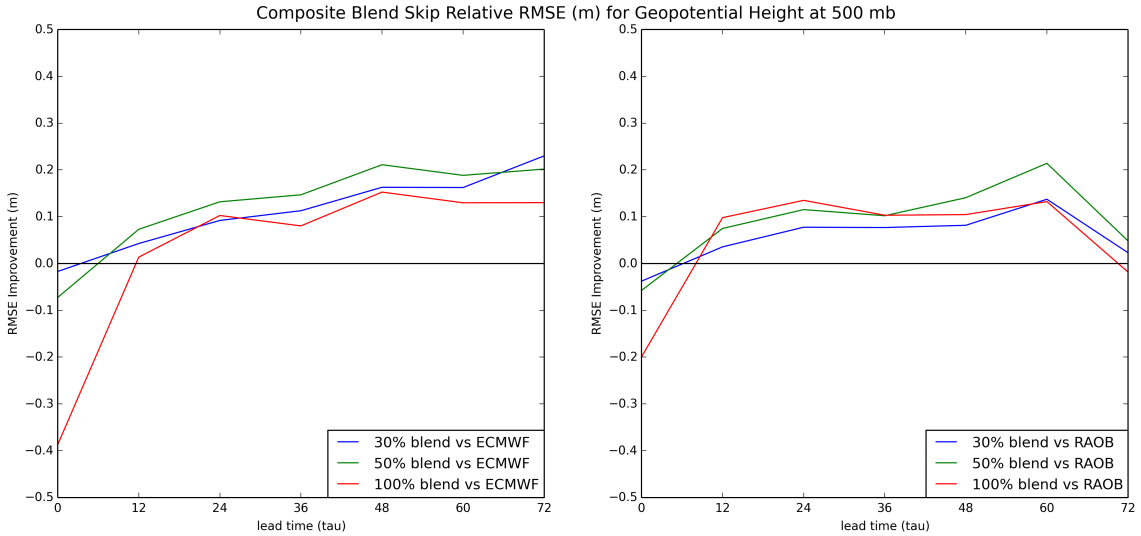

Figure B.2: Spatiotemporally averaged RMSE (m) of the 30, 50, and 100% blends relative to the blend skip for 500 mb geopotential height. Positive values indicate the blend has a lower RMSE than the blend skip (i.e. error improvement).

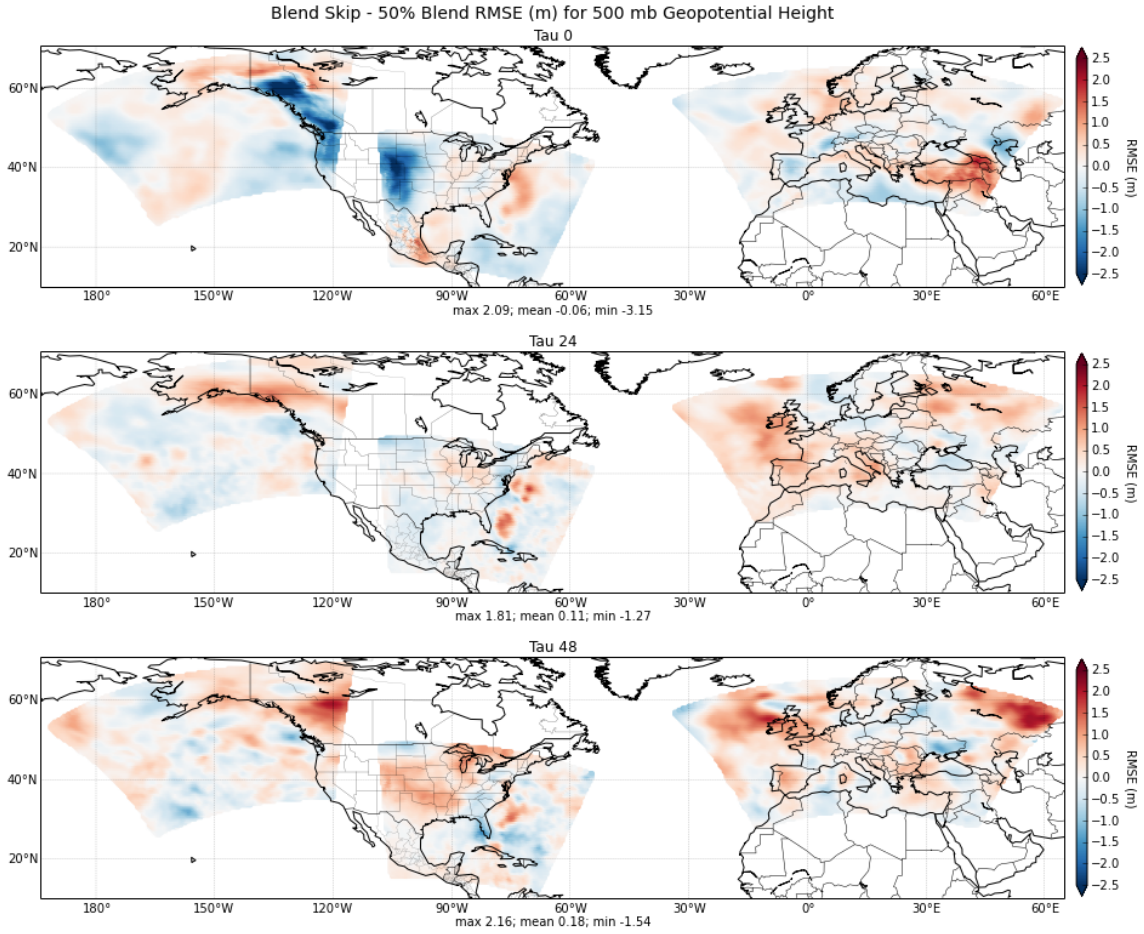

Figure B.3: Temporally averaged RMSE (m) for the 50% blend relative to the blend skip at all interpolated COAMPS grid points against the ECMWF analysis verification. The 500 mb geopotential height field is shown for the analysis time (top), 24 h lead time (middle), and 48 h lead time (bottom). Positive values (red) indicate improvement for the 50% blend relative to the blend skip, negative values (blue) indicate degradation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

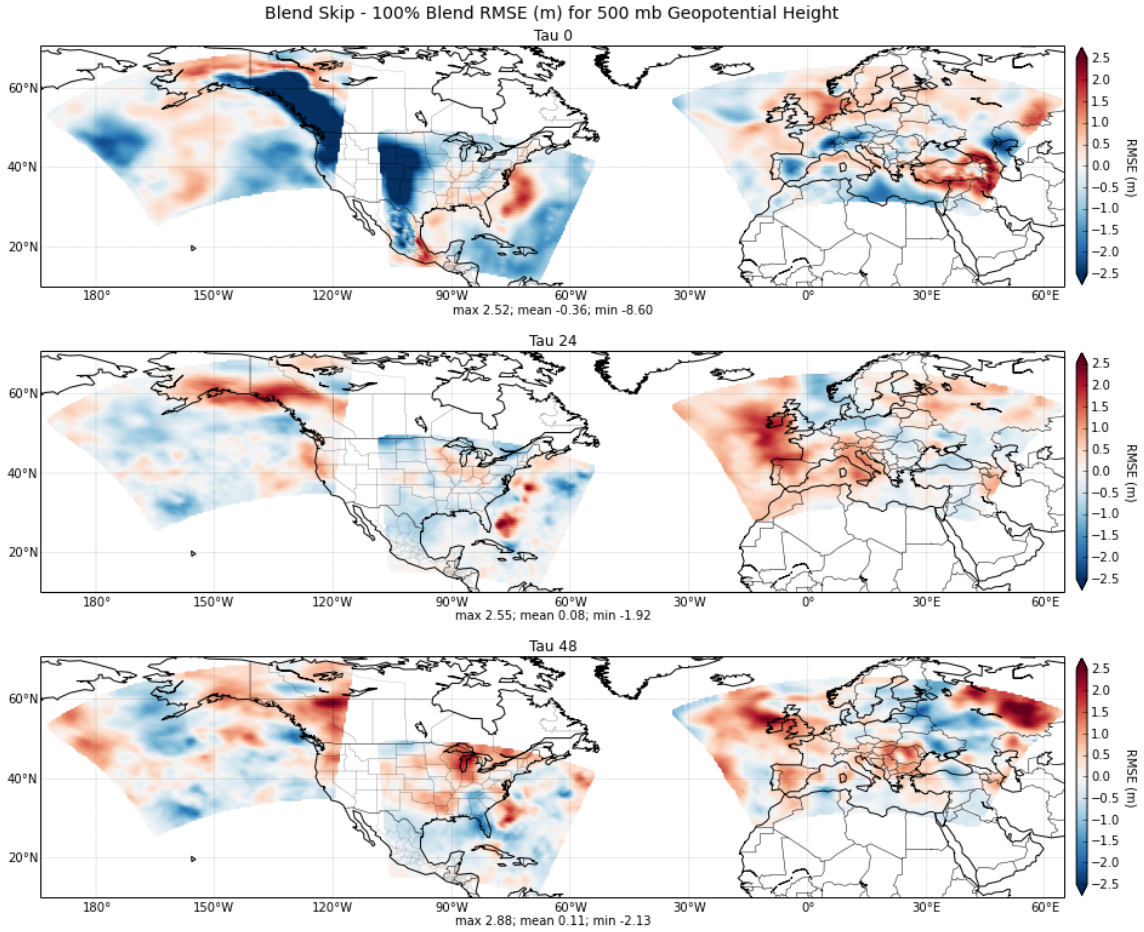

Figure B.4: Same as Figure (B.3) but for the 100% blend relative to the blend skip.

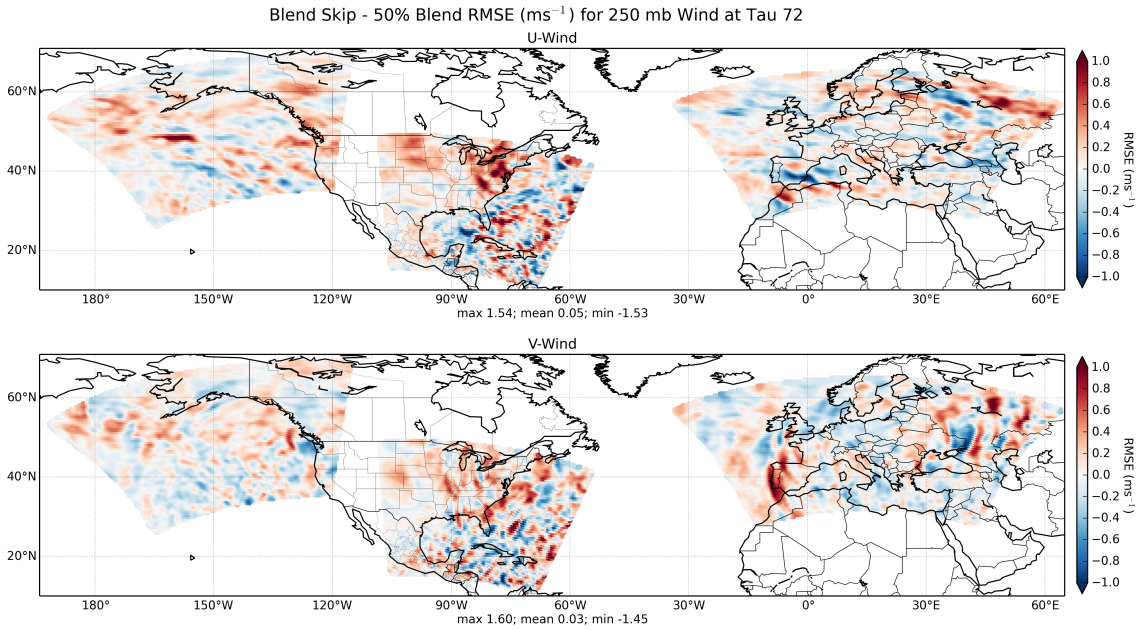

Figure B.5: Temporally averaged RMSE (ms^{-1}) for the 50% blend relative to the blend skip at all interpolated COAMPS grid points against ECMWF analysis verification. 250 mb zonal wind (top) and meridional wind (bottom) are shown at a 72 h forecast lead time. Positive values (red) indicate improvement for the 50% experiment relative to the blend skip, negative values (blue) indicate degradation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

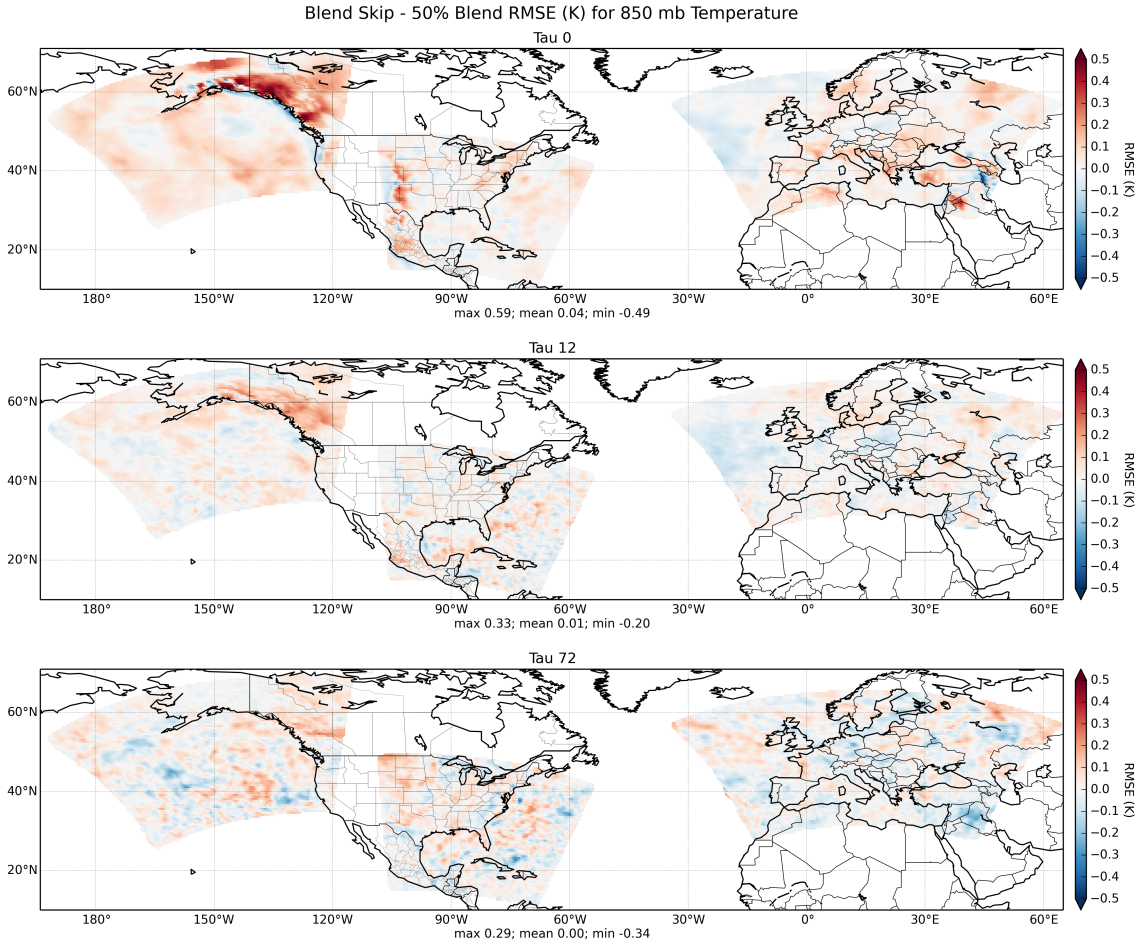

Figure B.6: Temporally averaged RMSE (K) for the 50% blend relative to the blend skip at all interpolated COAMPS grid points against ECMWF analysis verification. The 850mb temperature field is shown at the analysis time (top), 36 h lead time (middle), and 72 h lead time (bottom). Positive values (red) indicate improvement for the 50% experiment relative to the blend skip, negative values (blue) indicate degradation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

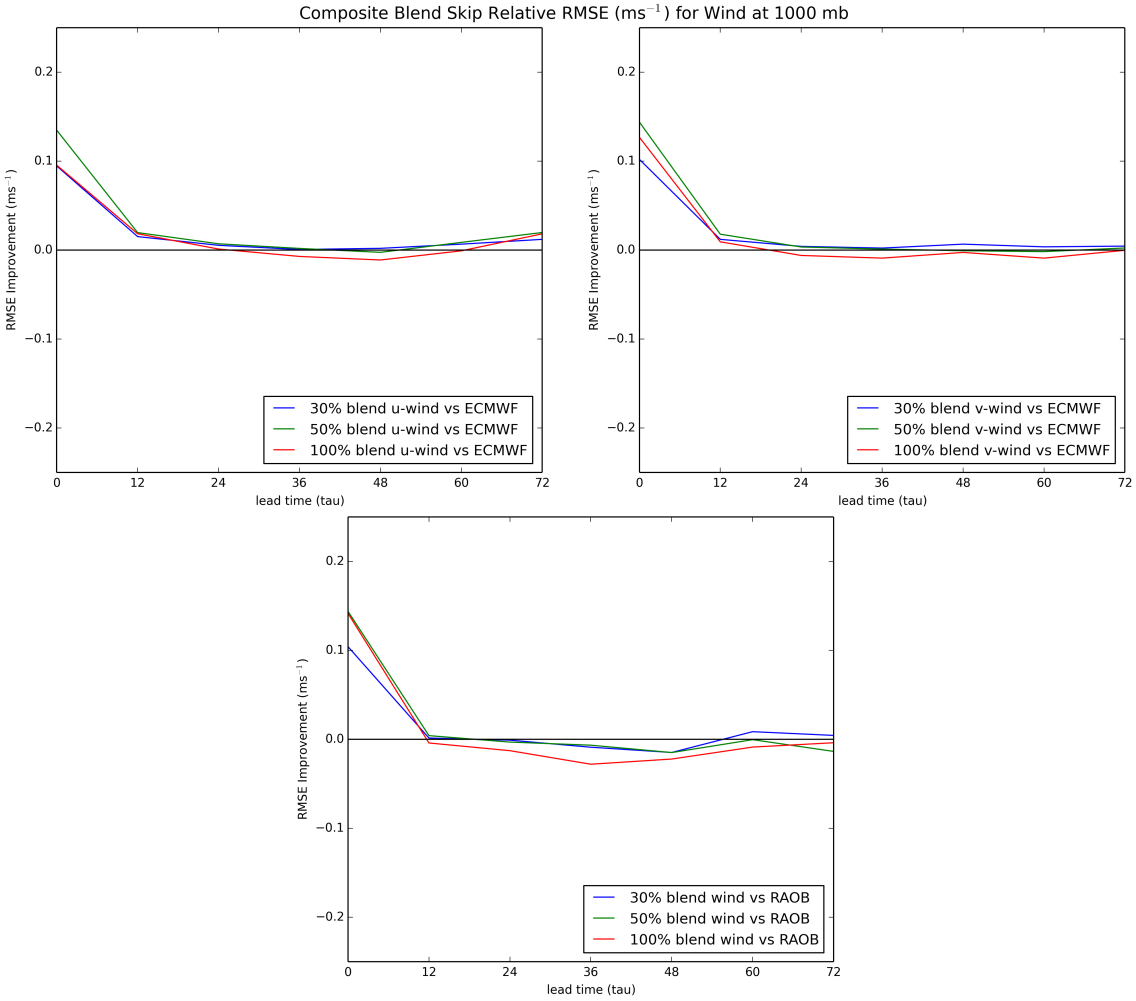

Figure B.7: Spatiotemporally averaged RMSE (ms^{-1}) of the 30, 50, and 100% experiments relative to the blend skip experiment for 1000 mb wind. RMSE differences for zonal wind against the ECMWF are shown at the top-left, meridional wind against the ECMWF at the top-right, and wind magnitude against RAOB in the bottom-center. Positive values indicate the blend improved relative to the blend skip, negative values indicate degredation.

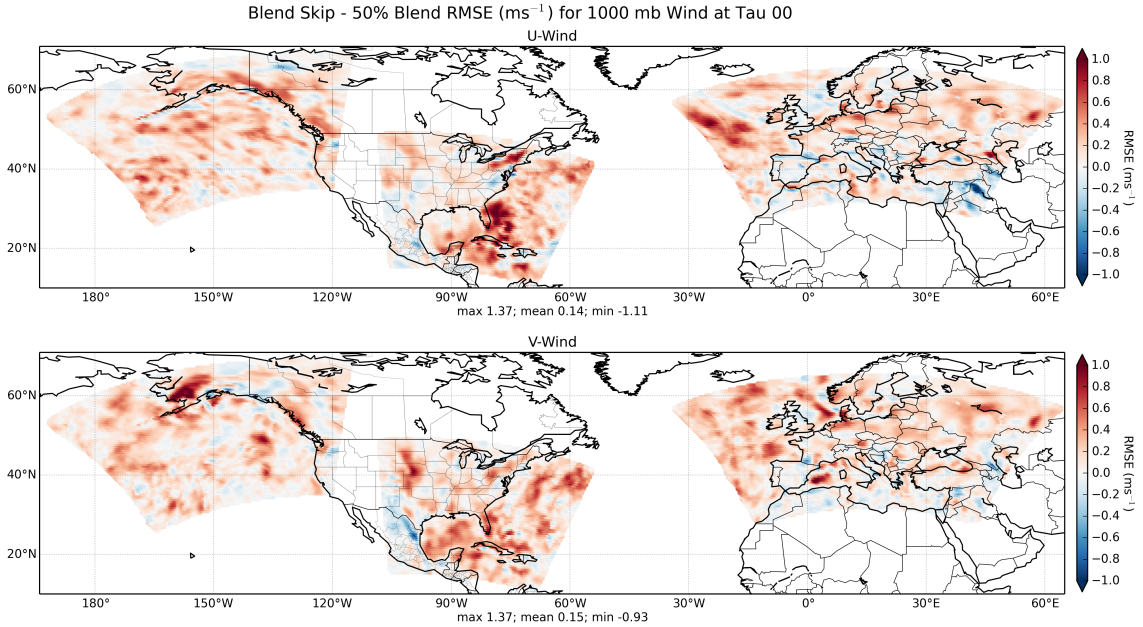

Figure B.8: Same as Figure (B.5) but for 1000 mb at analysis time.

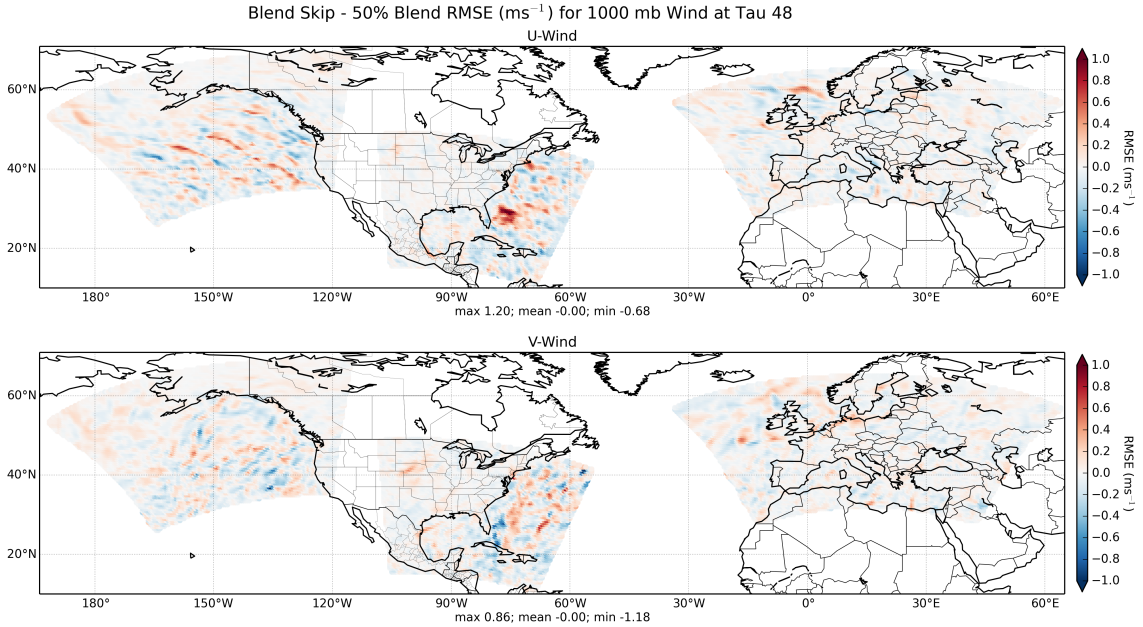

Figure B.9: Same as Figure (B.5) but for 1000 mb at forecast lead time 48.

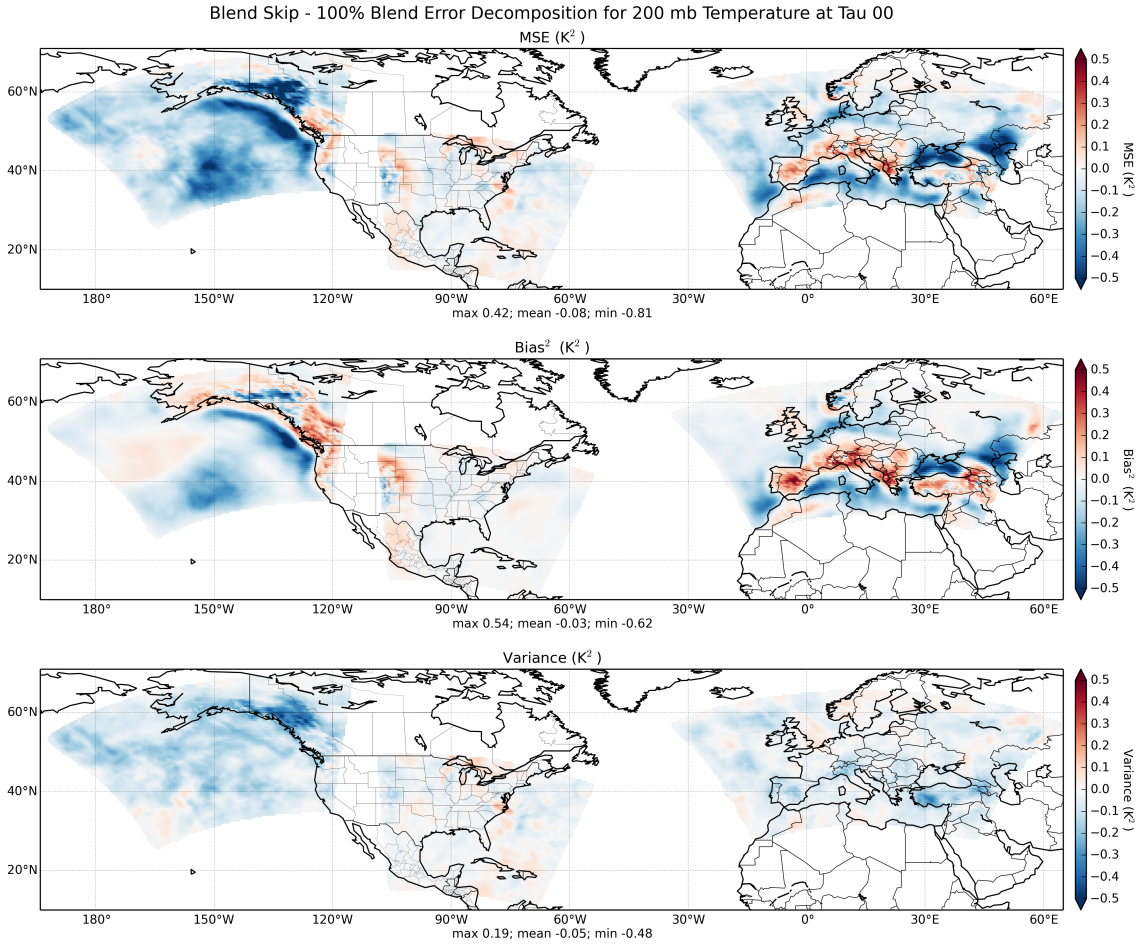

Figure B.10: Temporally averaged error difference decomposition of the 200 mb temperature analysis field into in mean square error (MSE, top), bias squared (middle), and variance (bottom) at all interpolated COAMPS grid points against ECMWF analysis verification. All units are in K^2. Positive values (red) indicate improvement for the 100% experiment relative to the blend skip, negative values (blue) indicate degradation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

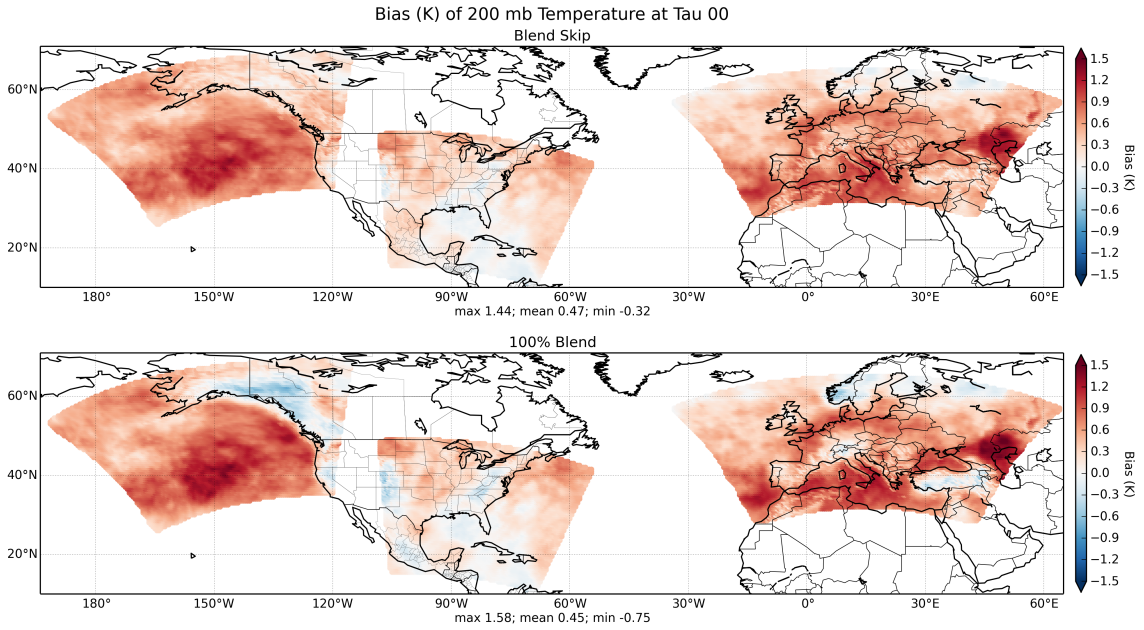

Figure B.11: Temporally averaged bias for the blend skip control (top) and 100% blend (bottom) at all interpolated COAMPS grid points against ECMWF analysis verification. All units are K. Positive values (red) indicate a warm bias, negative values (blue) indicate a cold bias. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

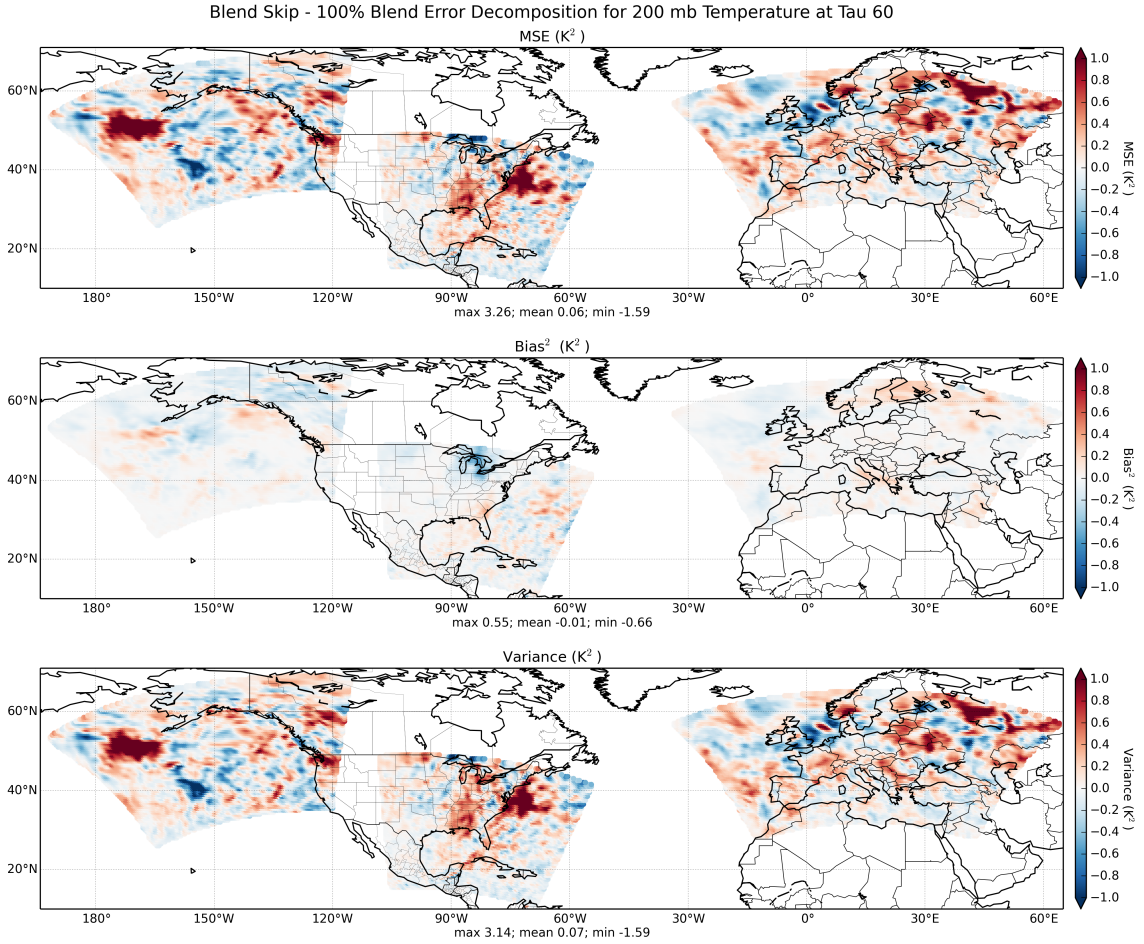

Figure B.12: Same as Figure (B.10) but for a 60 h forecast lead time. Note the scale in the color bar is expanded from that of the analysis figure.

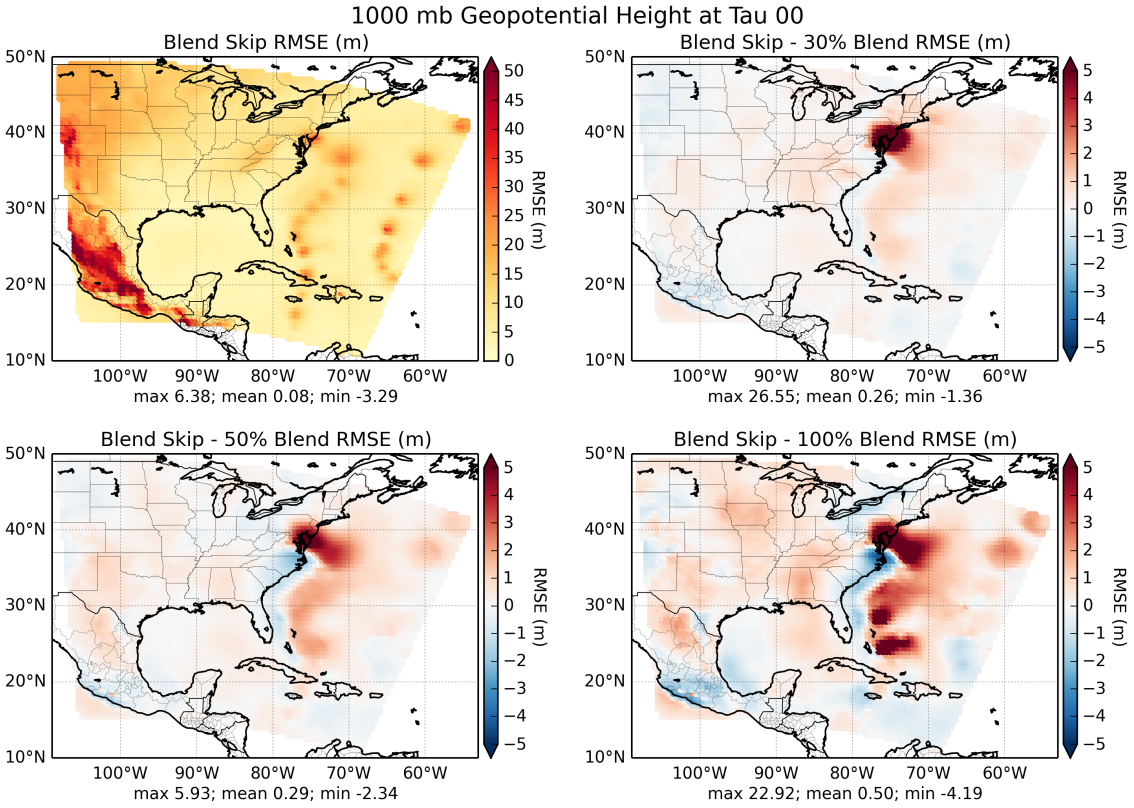

Figure B.13: Temporally averaged RMSE (m) of the 1000 mb Geopotential Height at analysis on the CONUS domain. RMSE of the blend skip experiment is shown in the top-left, while RMSE relative to the blend skip is shown for the 30% (top-right), 50% (bottom-left), and 100% (bottom-right) experiments. For the blend skip relative RMSE plots, positive values (red) indicate improvement for the blend experiment relative to the blend skip, negative values (blue) indicate degradation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.

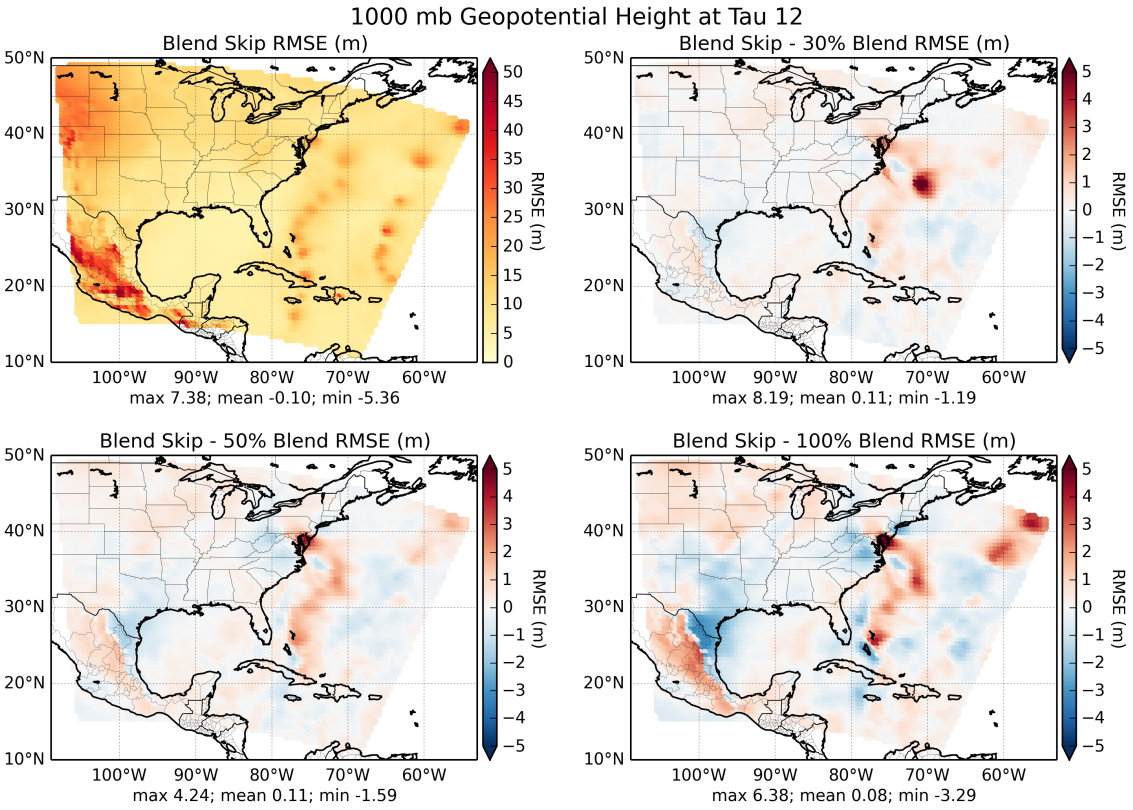

Figure B.14: Same as Figure (B.13) but for a 12 h forecast lead time.

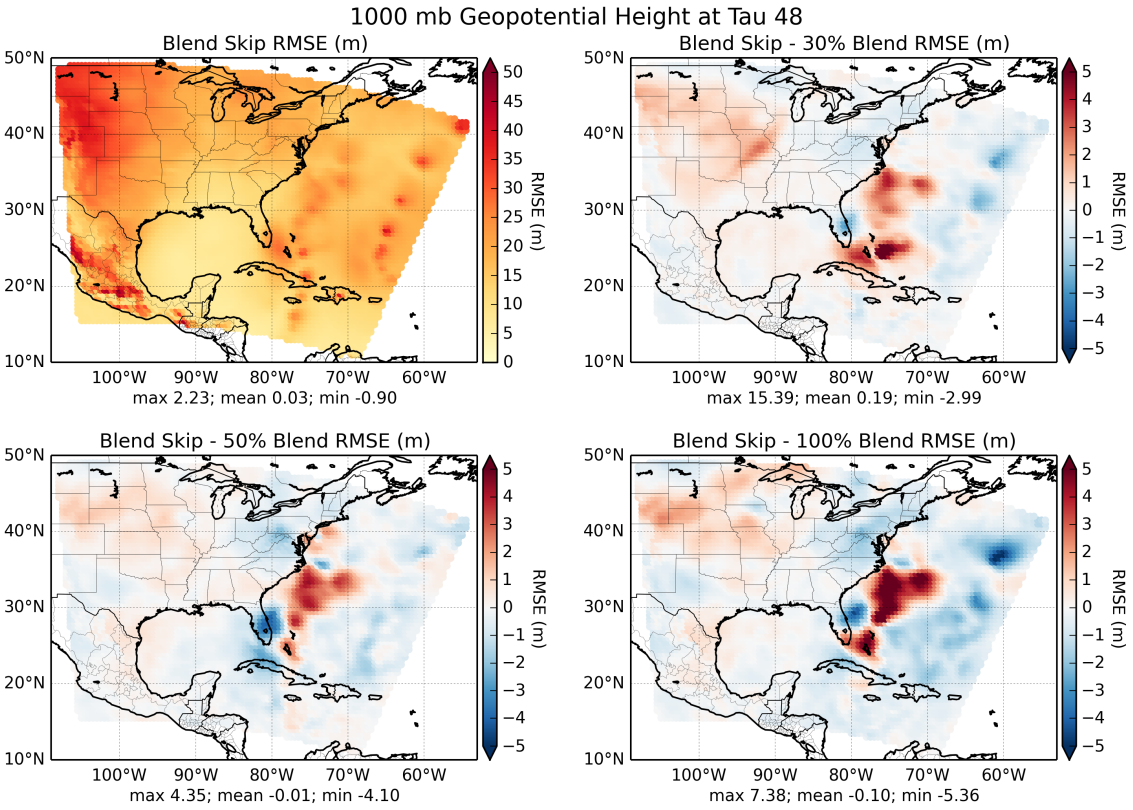

Figure B.15: Same as Figure (B.13) but for a 48 h forecast lead time.

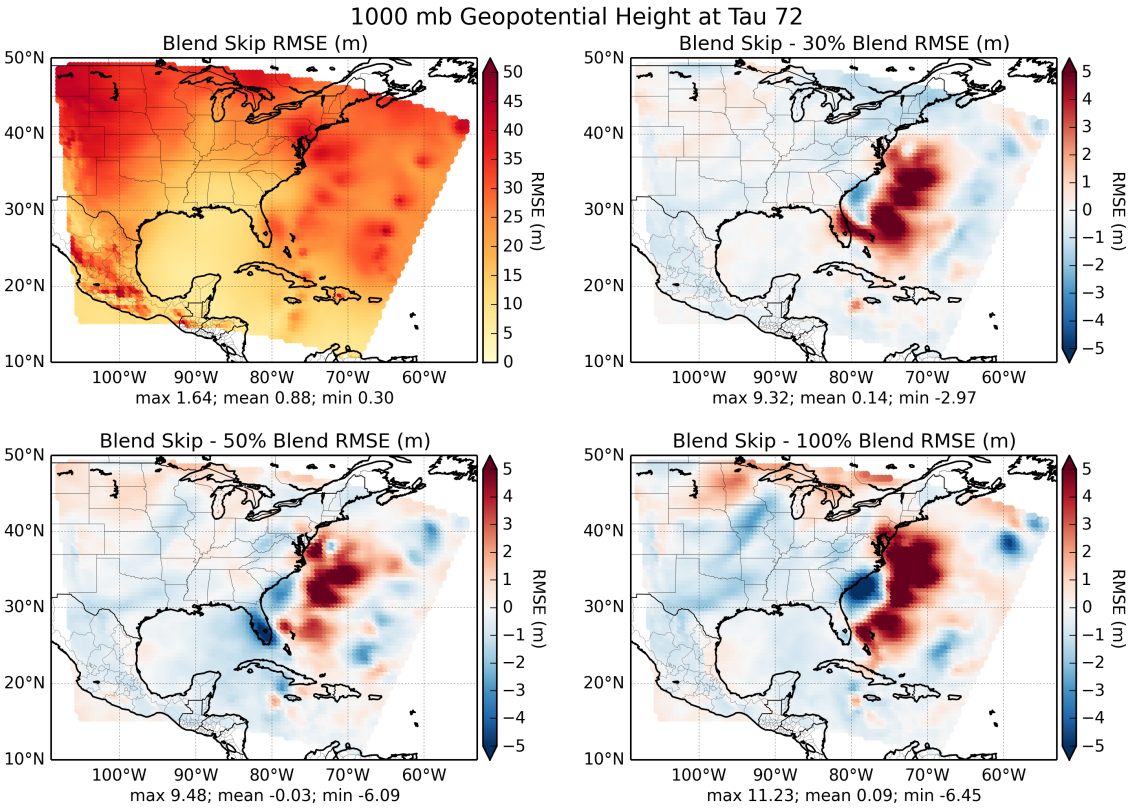

Figure B.16: Same as Figure (B.13) but for a 72 h forecast lead time.

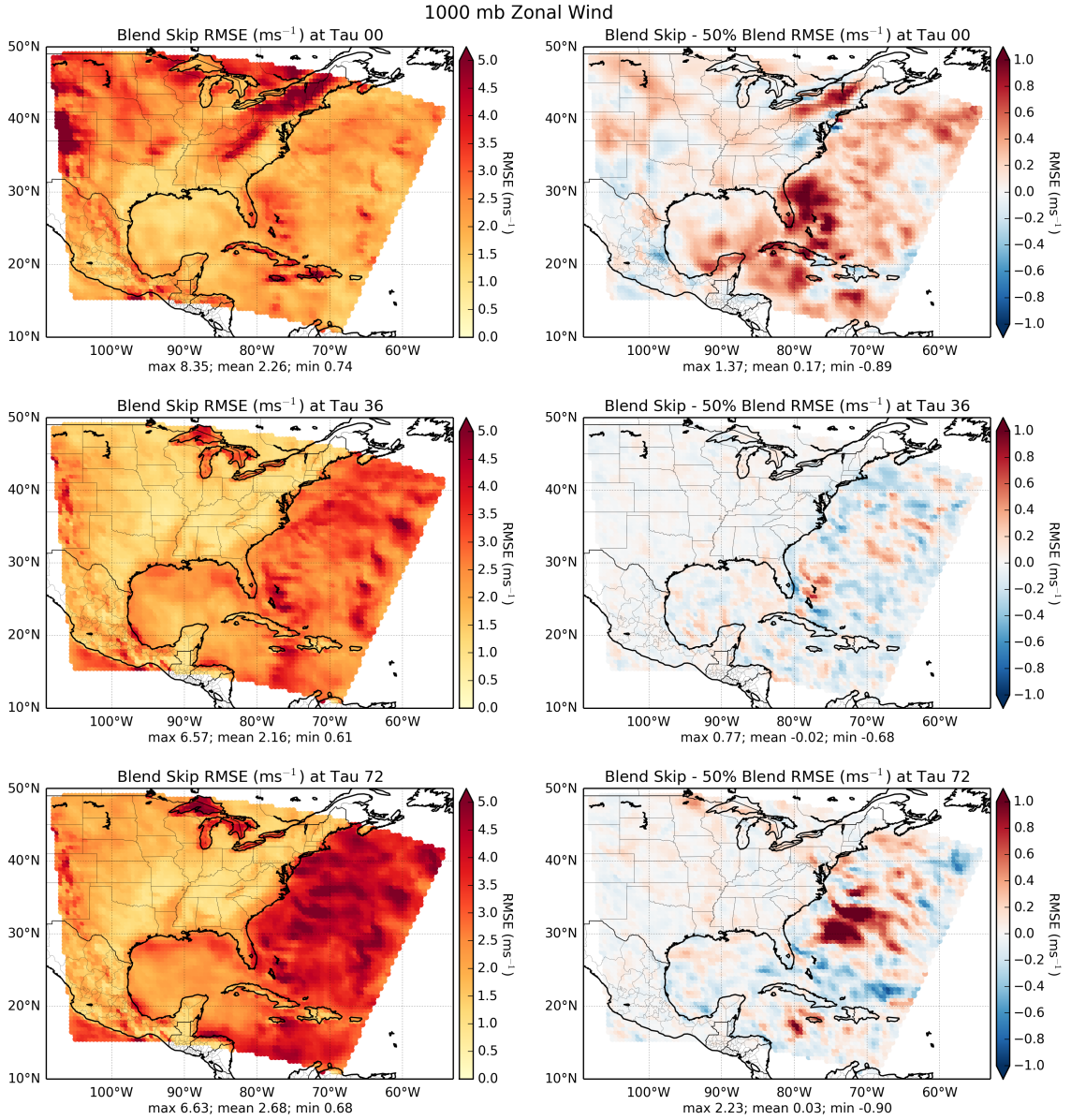

Figure B.17: Temporally averaged RMSE (ms^{-1}) of the 1000 mb zonal wind at analysis time on the CONUS domain. RMSE of the blend skip experiment is shown on the left, while RMSE of the 50% blend relative to the blend skip is shown on the right. For blend skip relative RMSE plots (right), positive values (red) indicate improvement for the blend experiment relative to the blend skip, while negative values (blue) indicate degredation. Maximum, mean, and minimum plot values are plotted in the caption beneath each plot.